Atlantic States Marine Fisheries Commission

Red Drum Stock Assessment Subcommittee Call

Draft Agenda

January 9, 2023

1:00 p.m. - 3:00 p.m.

The times listed are approximate; the order in which these items will be taken is subject to change; other items may be added as necessary.

Webinar: https://meet.goto.com/153155053

If you can't connect to the webinar with computer audio, you can dial in at:
Phone: +1 (646) 749-3129
Access Code: 153-155-053

1.	Welcome (*E. Simpson*)	1:00 p.m.
2.	Discuss Simulation Assessment Recommendations	1:05 p.m.
3.	Discuss Data Tasks and Data Request	2:00 p.m.
4.	Adjourn	3:00 p.m.

<u>**2024 Red Drum Benchmark Assessment Tasks**</u>

Simulation Follow-Up Tasks (see reviewer report for details)

1. Demonstrate that the Stock Synthesis estimation model for the southern stock can reproduce the dynamics of the operating model when given data without observation error.
2. Repeat the Traffic Light Analysis grid search using only the 'burn in' and pre-2023 periods to see if the reference points identified were similar to the ones identified in the presented assessment.
3. Evaluate robustness of simulation results to number of operating model iterations.
4. Conduct further exploration of how bias in growth parameters could influence the estimation model results.
5. Conduct sensitivity analyses to explore how changes in the selectivity curves influence estimation model predictions when given data without observation error.
6. Conduct sensitivity analyses to evaluate how the size and number of discarded fish could influence the assessment trends and reference points.
7. Explore the effect of start year on estimation model results.

Data Tasks

Addressed task during simulation assessment

Not done during simulation assessment

Discuss during SAS call to finalize data template

Agenda item to cover during Data Workshop

Possible data needs/tasks for exploration of alternative model structures

1. Total Catch
    a. Gear-specific commercial landings (see simulation assessment Table 8 for gear categorizations): Mike, but will be Anna-Mai for this assessment
    b. Commercial discards: Lee, but Cara has taken over as NC TC rep
    c. Query and calibrate MRIP state-specific recreational harvest and discards: Jeff
    d. Southeast Region Headboat Survey (*considered not useful in SEDAR 18): Make data request to Beaufort
2. Catch-at-length and catch-at-age
    a. Need fleet- (FL, GA, SC, northern commercial GNBS, northern commercial other, northern rec) and disposition(harvest and discard where applicable)-specific catch-at-length in even 2cm bins (change to inches and total length to align with management units and type??) for complete population size range for SS models
        i. Rec harvest length comps available by MRIP query: Jeff
            - Include supplemental program length data as was done in past assessments?
        ii. Rec discard length comps from app data and tagging data (see item 4 below): Angela and Thom

- How do we proceed with app-collected length data?

iii. Commercial length comps: TC members with data (need VA data by gear/fleet and expanded to total catch)
- Which sources should be used and how should they be combined?

b. Need slot-sized catch-at-length summed across fleets (FL, SC, northern stock) for TLA: Can be derived from data sets above

c. Need fleet- and disposition-specific catch-at-age for complete population age range for SS models

i. Commercial landings and discards: TC members with data (need NC data expanded to full age range, need VA data by gear/fleet and expanded to total catch)

ii. Recreational harvest and discards (where applicable)

d. Need fleet- and disposition-specific catch-at-age for ages 1-7+ for SCA: Can be derived from data sets above

e. Fleet-specific length distribution by age for conditional age-at-length data: Can be derived from ALK data for catch-at-age

3. Indices of Abundance

a. Stock-specific MRIP CPUE: Chris S.

i. MRIP CPUE length and age compositions for southern SS model

b. Survey indices

i. Need aggregate index (all sizes and ages): TC members with survey data

ii. Need length composition in even 2cm (change to inches and total length to align with management units and type??) bins for complete population age range for SS models: TC members with survey data

iii. Need slot-sized indices for TLA: Can be derived from aggregate index and length composition data sets, but would lack associated variance measure

iv. Need age composition for complete population age range for SS models: TC members with survey data

v. Need age-specific indices for SCA and TLA (can be derived from aggregate index and age composition data sets, but would lack associated variance measure): TC members with survey data (should our request be for just aggregate indices or age-specific as well?)

vi. Standardize indices? Just aggregate and then use to derive other alternatives?

vii. Index correlation analysis (simulation assessment reviewer recommendation)

viii. Index synthesis (simulation assessment reviewer recommendation): Potential methods include Conn 2010, VAST (Thorson 2019), hierarchical modeling, and dynamic factor analysis

4. Tag-Recapture Data (what format is needed to support all potential uses??-see data template for current format requested)

a. Summarize in bins for recreational discard length composition proxy?

b. Growth analyses using tag-recapture data (simulation assessment reviewer recommendation): See potential methods in Fabens 1965, Welsh et al. 2003, and Kirkwood 1983

c. Summarize for SS model tag-recapture sub-models?

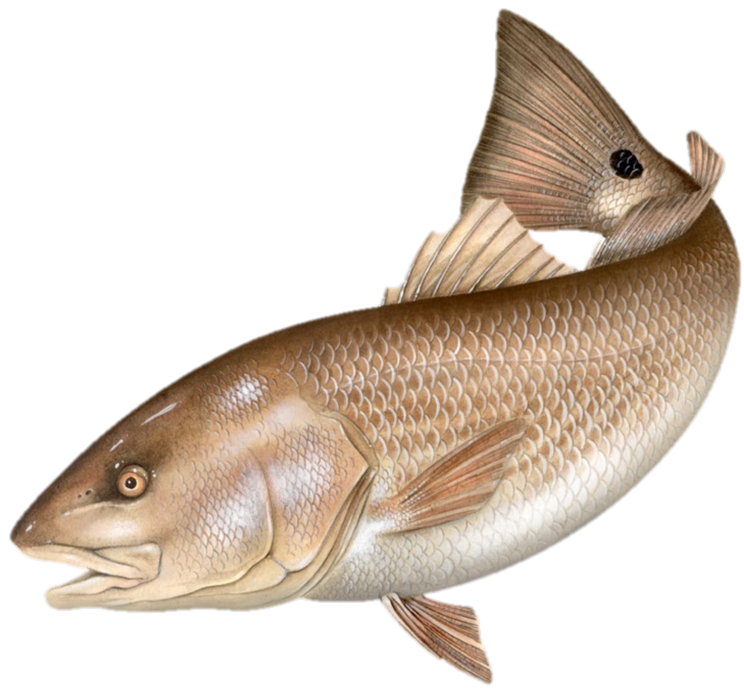
5. Biosampling Data
    a. Any special considerations for these data?
    b. Growth analyses using age-length data (simulation assessment reviewer recommendation): See potential methods in Lester 2004 and Schuller et al. 2014
6. Other
    a. Literature review: Thom
    b. Identify environmental data that could be used as recruitment covariates based on findings by Goldberg et al. 2021

Modeling Tasks

1. SS models as preferred models
2. TLAs as supplementary analysis and potential tool for between assessment monitoring
3. SCA models for continuity runs
4. Tag-recapture model to estimate mortality (simulation assessment reviewer recommendation, but not clear if this was intended as part of SS models or as independent analyses): See Wert 2017 for newer tag-based estimates of mortality

Atlantic States Marine Fisheries Commission

Red Drum Simulation Assessment and Peer Review Report



Accepted for Management Use
by the Sciaenids Management Board
May 2, 2022

Sustainable and Cooperative Management of Atlantic Coastal Fisheries

ACKNOWLEDGEMENTS

The Atlantic States Marine Fisheries Commission (ASMFC) thanks all of the individuals who contributed to development of the 2022 Red Drum Simulation Stock Assessment. The ASMFC specifically thanks members of the Red Drum Technical Committee (TC) and Stock Assessment Subcommittee (SAS) who developed the consensus stock assessment report, the Review Panel for reviewing the assessment and developing the peer review report, and ASMFC staff, Jeff Kipp, Savannah Lewis, and Pat Campfield, for coordinating the assessment and peer review.

Prepared by the
Red Drum Stock Assessment Subcommittee
Joseph (Joey) C. Ballenger, PhD, SCDNR, Chair
Jared Flowers, GA DNR
Angela Giuliano, MD DNR
Jeff Kipp, ASMFC
Lee Paramore, NC DEQ
Chris Swanson, FL FWC
Thom Teears, NC DMF

Red Drum Technical Committee
Lee Paramore, NC DEQ, Chair
Joseph (Joey) C. Ballenger, PhD, SCDNR
Robert Bourdon, MD DNR
Michael Greco, DE DFW
Chris Kalinowsky, GA DNR
Roger Pugliese, SAFMC
Harry Rickabaugh, MD DNR
Ethan Simpson, VMRC
Alissa Wilson, NJ DEP

The TC and SAS acknowledge the following individuals for their support during the assessment: Katie Drew (ASMFC) for providing technical support and analytical contributions, Mike Rinaldi (ACCSP) for validating and providing commercial landings data from partner agencies, Kelli Johnson (NOAA Fisheries) for providing guidance and technical support for the ss3sim simulation software, Kathryn Doering (NOAA Fisheries) and Rick Methot (NOAA Fisheries) for guidance on Stock Synthesis modeling, the Assessment Science Committee, particularly the subcommittee of Amy Schueller (NOAA Fisheries), Matt Cieri (ME DMR), and Alexei Sharov (MD DNR), for support developing a road map to future stock assessments of red drum, and Tina Berger (ASMFC) for formatting and publishing this report.

A publication of the Atlantic States Marine Fisheries Commission pursuant to National Oceanic and Atmospheric Administration Award No. NA20NMF4740012.

PREFACE

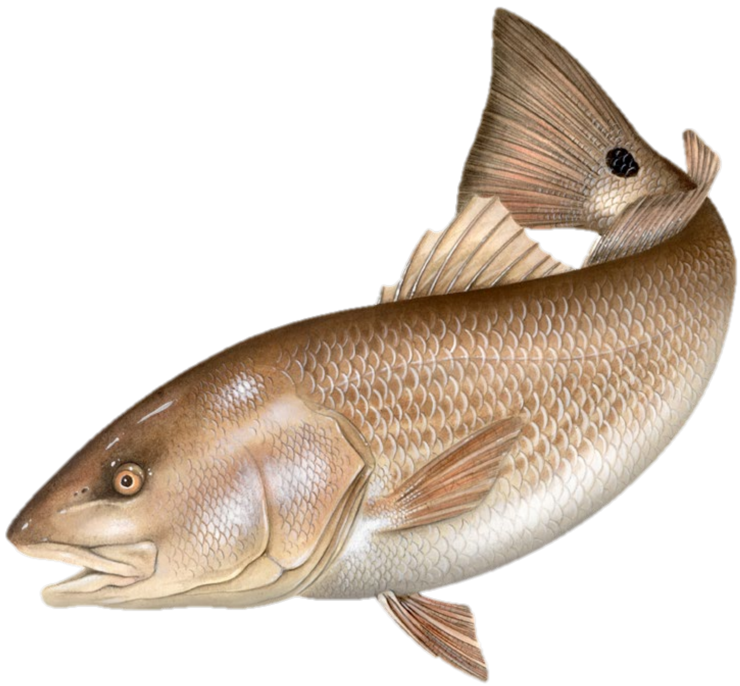
The Red Drum Simulation Assessment and Peer Review Report is divided into two sections:

Section A – 2022 Red Drum Simulation Stock Assessment Peer Review Report
PDF pages X – XX
This section provides a summary of the Red Drum Simulation Stock Assessment results supported by the Peer Review Panel. The Terms of Reference section provides a detailed evaluation of how each assessment Term of Reference was addressed by the Red Drum SAS and TC and provides recommendations from the Panel for future areas of research to improve red drum assessments.

Section B – 2022 Red Drum Simulation Stock Assessment Report
PDF pages XX – XXX
This section covers the Red Drum Simulation Assessment developed by the Red Drum TC and SAS and describes background information on red drum, data used, analyses for the assessment, and recommendations for modeling in future red drum stock assessments submitted to the Peer Review Panel.

Atlantic States Marine Fisheries Commission

2022 Red Drum Simulation Assessment Peer Review Report



Conducted on
March 28-30, 2022

Prepared by the
Red Drum Simulation Assessment Review Panel

Dr. Amy Schueller (Chair), National Marine Fisheries Service, Beaufort, NC
Dr. Michael Allen, University of Florida, Cedar Key, FL
Dr. Jie Cao, North Carolina State University, Morehead City, NC
Dr. Daniel Hennen, National Marine Fisheries Service, Woods Hole, MA

SUMMARY

Red drum *Sciaenops ocellatus* is a popular recreational fish along the Atlantic and Gulf coasts of the United States. Red drum exhibit ontogenous movement dynamics whereby young of year and sub-adults spend their time in estuarine environments and adults migrate further offshore. Recreationally caught fish can be harvested if fish are within a slot length limit. Data collected from the adult population is sparse and mainly consists of information from various long line surveys. The lack of available adult abundance information results in stock assessments that have been unable to accurately estimate stock status. The purpose of this review is to evaluate and identify stock assessment methods most robust to the types of data available for red drum. This was accomplished by simulating data using an operating model and then fitting simulated data using various stock assessment models. The three models considered and compared in the simulation assessment include the Traffic Light Approach (TLA), SCA (a statistical catch-at-age model developed in ADMB and used historically for red drum), and Stock Synthesis (SS; a statistical catch-at-age model developed in the SS program).

The Review Panel (RP) recommends the use of the SS program to assess both the northern and southern stocks of red drum, with the use of the TLA as an accessory tool between assessments. The SCA model was not able to reproduce the outcomes from the operating model when fitting to near-perfect data. The SS model is ready to use for the northern stock, while the model for the southern stock requires more exploration before use in stock assessment. In particular, some results were unexpected and unexplained. The unexpected results are detailed below and require further attention.

The Review Panel (RP) appreciates all of the hard work by the Red Drum Stock Assessment Subcommittee (SAS) and Technical Committee to create a comprehensive simulation assessment. The Panel also thanks the Director of Fisheries Science for organizing the meeting, providing materials to the Review Panel in a timely fashion, and additional support throughout the review. A Review Workshop was conducted in Raleigh, North Carolina, during the week of March 28, 2022. Workshop discussions were professional and constructive, and overall the simulation assessment passes review.

The following report provides an evaluation of the simulation work and recommendations from the Panel, with detailed comments for each Term of Reference.

TERMS OF REFERENCE

1. Evaluate the thoroughness of data collection, data treatment, data presentation, and characterization of data uncertainty.

The Review Panel believes the Stock Assessment Subcommitee did an excellent job of summarizing and analyzing a large number of complex data sets that went into the assessment models. The simulation assessment is thorough in its description of the data sources and how they are used in the three different models. Uncertainty is well characterized overall, although we note a few cases where the models are biased via not making accurate predictions when given perfect data (e.g., SS model for southern stock). We suggest attempting a few adjustments to remove bias and improve the utility of the models.

We believe the authors should consider alternate growth curve formulations. Schueller et al. (2014) offers a potential for bias correction to consider. Alternately, Lester et al. (2004) offers a growth model that specifically models the pre-maturation phase of growth separately from the mature phase that could produce a better fit to the data. The RP believes the two approaches should be considered to better model size at age. However, the RP notes several aspects of the size at age data that could result in biased growth parameters, regardless of the model chosen.

a. Variability in size at age declines with age, an unlikely relationship that may be a result of gear bias; in most fishes, variability in size at age is constant or increases with age
b. Drum in the 70-90 cm size range are not well sampled, likely a result of gear bias; and
c. The RP also believes future explorations of size at age for red drum stocks should evaluate existing growth increment data from the tagging studies, to further elucidate growth patterns.

Thus, the data available to analyze growth for red drum are likely problematic and need further consideration and analysis. Traditionally, the expectation is that as age increases, the variability in size is likely to remain constant or increase. The lack of such a trend in the data suggests there is a bias in data collection and the full variability of size at age is not being sampled. One potential bias could be a gear bias where a certain survey or fishery gear doesn't sample specific sizes well. Another example could be a bias due to spatial dynamics of the population and no sampling occurring within a given area or time frame. This potential bias in sampling leads to a potential bias in the estimation of the growth curve parameters. When estimating growth, one assumption is the data at age are representative of the range of sizes at that age.

The Assessment Committee made the assumption that the data reflect the true size distribution at age and corrected the growth curve estimation by allowing for an age-varying K parameter for the von Bertalanffy growth curve. An alternative explanation is the data are not representative of the full distribution of sizes at a given age. If this is the case, the estimation of the von Bertalanffy growth curve should be bias corrected such that all of the parameters would be estimated in an unbiased manner. A tested method to bias correct growth curve estimations can be found in Schueller et al. (2014). The RP recommends bias correcting the growth data given the lack of samples in the 70-90 cm range, which indicates the full size range at age is unlikely to be sampled across all age classes. In addition to the bias correction, the data should be explored over time to assess the possibility of time-varying growth. However, the RP recognizes that considering and correcting for bias in the growth data could be beyond the scope of the simulation assessment. The RP suggests further exploration of how bias in growth parameters could influence the simulation assessment model results.

The RP also notes that potential growth information from tagging data has not been investigated in past stock assessments due to availability of traditional age-length data. Various tagging programs for red drum have been conducted in multiple states. There is a substantial amount of tagging data available, including information on large and old individuals. The RP recommends analyzing the size increment data from tagging programs. For example, analysts can fit the growth increment form of the von Bertalanffy function (Fabens 1965) to the size increment data. The estimated von Bertalanffy parameters (K and Linf) can then be compared with those obtained from the age-length data. The comparison may shed light on the representativeness of the age-length data. Furthermore, it may be worthwhile to fit the von Bertalanffy growth curve using both size increment and age-length data (Kirkwood 1983). Again, this is a recommendation for future assessments.

The survey index data for the northern model were appropriate and were limited to one index for recruitment, one index for sub-adults, and one index for mature adult abundance. The approach used in the northern region uses the available data to the extent scientifically possible.

The survey index data for the southern model were more plentiful and complex. The base configuration of the southern model included eight index data sets. The model included three indices of recruitment, two indices for sub-adults, and three indices representing mature abundance. When multiple indices are included that represent the same segment of the population, the estimation model will find similar trends, but will also have a difficult time fitting the data if the same underlying trends are not informing the data. Moving forward with the estimation model, analysts should consider providing the best information available on trends in abundance over time for the given size and age ranges. With multiple possible data sources, analysts should consider prioritizing the data and using the longest time series and largest, most representative spatial scales. If that is not an option and all data are equally valuable, analysts could consider combining indices using a variety of different options such as the Conn method (Conn 2010), VAST (Thorson 2019), hierarchical modeling, or dynamic factor analysis. In addition, exploring the relationship of the indices to each other through correlation analyses, with appropriate lags to account for size or age class differences, is critical to determining if the estimation model inputs provide a cohesive picture of the stock dynamics.

Natural mortality is one of the most critical parameters influencing the identification of sustainable harvest levels. The RP feels the simulation assessment handled natural mortality appropriately using surrogate measures for M and size dependency in M. Overall, the natural mortality approach used in the models was appropriate.

During discussions, the RP learned that much more tagging data exists that could provide better informed estimates of fishing mortality, particularly in North Carolina and South Carolina. There are evidently data that correct for non-reporting of tags and thus could be very useful. The RP encourages new analyses of the tagging data to obtain estimates of harvest rate information (F) that could improve future assessments.

Finally, the discard mortality rate was a key uncertainty in this assessment, as well as the number and size composition of released fish that ultimately would be exposed to discard mortality (currently set at 0.08). There is a key need to better quantify the number and sizes of discarded catch, particularly given the apparent recent increase in anglers targeting large, spawning fish offshore. The RP recommends better data collection of discard numbers and sizes as a high priority for future assessments, including the use of angler phone apps and other tools to measure the size and number of discarded fish. Further, the assessment could benefit from more sensitivity analyses to evaluate how the size and number of discarded fish could influence the assessment trends and reference points. Finally, the RP believes the discard mortality rate of 0.08 could be a bit high, and should consider the effects of lower values (e.g., 0.04). That said, the number and size of discarded fish is a major uncertainty that if quantified, would improve future assessments.

2. Evaluate the thoroughness and appropriateness of information used to parameterize simulation models.

The RP feels the SAS did a very thorough job of parameterizing the models, including critical parameters of natural mortality and recruitment compensation. There is some uncertainty in how selectivity from the different regions is influencing model outputs, as regulations changed through time and were different across the states. This creates uncertainty in the models because the north and south stocks have different selectivities, likely operating within different states for each region (north and south). Selectivity is particularly concerning for the southern stock where size and bag limits varied through time and across the states of South Carolina, Georgia, and Florida. An amalgamation of selectivities could contribute to uncertainty and possibly bias in the southern stock SS model. The RP recommends further sensitivity analyses to explore how changes in the selectivity curves influence model predictions when given perfect data.

3. Evaluate the appropriateness of models for simulating red drum populations and generating data sets sampled from the simulated populations.

The Stock Synthesis simulation package (SSsim) is used to simulate red drum populations and create data sets from the operating models. The RP agrees this is an appropropriate model or method for simulating red drum populations and generating data sets for use in the estimation models. Overall, the uncertainty in the operating model represents the observed uncertainty in the data.

4. Evaluate the incorporation and treatment of uncertainty in simulated populations.

The RP feels that uncertainty was handled appropriately overall. The SAS includes uncertainty through variable population dynamics scenarios in the operating model (OM). These include a scenario in which fishing pressure is increased in the projection period, an increase in the selectivity at age of older fish through a catch and release trophy fishery, a scenario in which natural mortality is lower than expected, and a time varying realized recruitment scenario. The incorporation of uncertainty into the simulated populations in the operating model is well described and appropriate for red drum.

5. Evaluate candidate assessment methods and application of assessment methods to data sets sampled from simulated populations.

The Stock Assessment Subcommittee (SAS) explored a few assessment methods within each of the estimation models (EM). Exploration of assessment methods is constrained by the limitations of each EM framework and by the requirement that any model configuration has to be flexible enough to fit the data provided by each of the scenarios developed in the OM.

In general, the assessment methods available for exploration in the SCA are limited compared to those available in the SS EM. For example, the SAS explored estimating time varying equilibrium recruitment (R0) in SS as an attempt to fit the data produced by the OM, which has a temporally varying stock recruitment relationship. The SCA EM does not estimate stock recruitment parameters and so no such exploration is possible.

The assessment methods available in SS are many and varied. The SAS chose to limit the tuning of SS models to configurations that would fit all of the runs from each of the OM scenarios. The approach means that some individual runs and scenarios could be fit better, and results for the SS models are possibly less precise than they could be. However, the RP recognizes it would be unreasonable to attempt to tailor each fit to the hundreds of OM runs. SS employs parameter penalties to help with estimation. The penalties can be (mis)used to direct the EM to a particular solution on the likelihood surface, inflating the perceived stability of the model. The SAS does not misuse this feature of SS. They employ appropriate penalties on parameters that are weak enough to allow a broad array of solutions and provide enough guidance to help with model convergence. Other choices made in configuring SS for each OM scenario are reasonable and would likely have been employed by other competent stock assessment scientists given similar datasets.

The RP finds the application of assessment methods to be appropriate and representative of the choices made by professional stock assessment scientists. However, a few additional items could be considered.

First, further examination of the SS estimated stock-recruit relationship, including the steepness parameter, is recommended. The estimated steepness values are unexpectedly low for both north and south stocks, causing an estimated SSB ratio with considerably high bias (e.g., scenario ***Depr R***). The RP feels the assessments do not appear to have data to inform steepness, and thus recommend fixing steepness at 0.99. However, the RP recognizes that such model configuration, in conjunction with other fixed life history parameters (e.g., natural mortality), could constrain the calculation of potential reference points. Fixing several parameters limits the flexibility for reference points to be informed by the data.

Second, the RP recommends exploration of the start year of the model. Given the time series available, the model could be started earlier than 1989 or later than 1989. The model could be started earlier, for example 1950, in order to capture the decline in the population with increased catches by both the commercial and recreational fisheries, and to leverage all of the available data. In addition, a later start year of 1991 could be considered if tagging data were to be used. Parameter estimates from the tagging data during the earliest years were quite uncertain. Censoring those earlier years may help with parameter estimation and model performance. Additional sensitivity runs should be used to diagnose the robustness of the model outcomes to the decision of the starting year of the model. In some cases, the choice of start year can lead to difficulties initializing the model at the appropriate scale of abundance given the data available and the level of depletion.

6. **Evaluate the choice of reference points for characterizing stock status of simulated populations. Recommend alternatives if necessary.**

In general, the RP feels the reference points selected by the SAS are appropriate. The RP agrees that an escapement reference point is vital to assessing a stock primarily driven by recruitment. The RP recommends monitoring both an annual and 3-year moving average measure of SPR status. The three year moving average introduces some inertia into the management process and reduces the probability that management actions are based on noise rather than signal. The annual measure can be important to balance that inertia with the ability to detect rapid changes in SPR status that might require immediate attention.

SSB status could turn into a trend-based reference point, but the SAS would need to select a reference time period. A general result of the simulation exercise is that trend was more stable than scale in SS models for both regions. If this result holds, once a final version of the SS model for the south is configured, there is a possibility of using an SSB reference point based on trend for management. Trend-based SSB reference points require a reference period for internal comparison. Identifying an appropriate reference period would require further study by experts in the fishery and is outside the purview of the RP.

The SS model for the south appears biased in scale, but demonstrates a stable trend. This result indicates trend based reference points could be useful for management. The RP thinks trend based reference points are a potentially useful tool to mitigate a model that shows scale instability. However, the Panel recognizes there may also be trends in bias. Once the SAS has demonstrated that the EM for the south can reproduce the dynamics of the OM when given data without observation error, it will be possible to determine if there are trends in bias and by extension whether or not trend-based reference points are appropriate.

7. Evaluate the choice of metrics used to evaluate performance of each candidate assessment method for estimating the population dynamics and stock status of simulated populations. Recommend alternatives if necessary.

The selected performance metrics are appropriate and represent standard reference points for diagnosing overfishing in stock assessments. The escapement goals for red drum are sound performance metrics for a stock with dome-shaped selectivity that focuses harvest on juvenile fish.

The SAS conducted 100 iterations for each scenario and computed relative error and error rates (Type I and Type II) as metrics for each EM and scenario. Given that process error (recruitment deviations) is also introduced to the simulation, the RP feels that 100 iterations is a low number and simulation results might reflect a substantial amount of randomness. The RP notes the actual number of iterations is lower because non-converged runs were excluded. The RP recommends the following two exercises to explore the impacts of number of iterations: (1) for a given scenario, increase the number of iterations to 200 and compare the results with 100-iteration results; (2) for a given scenario, perform several runs of 100 iterations and check variability in produced relative errors and error rates among runs. For the purpose of model comparison, however, the RP thinks 100-iteration results will likely indicate the difference in performance among EMs. For a given scenario the SAS fit all of the EMs to the same <100 datasets. The EMs used the same datasets and comparisons were based upon medians.

8. Evaluate the choice of the preferred assessment method(s) for characterizing stock status. Recommend alternatives if necessary.

The RP evaluated all three assessment methods presented by the SAS: TLA, SCA, and SS. Overall, the RP does not recommend further exploration of the SCA model. The RP recommends the use of the SS model for future analyses and assessments, and recommends use of the TLA as an accessory model.

The RP recommends that the SCA model should not be further explored for red drum stocks because the SCA seems to be intrinsically biased even when using perfect data from the operating model. The RP notes the initialization of the SCA and the bias associated with it could be remedied with alternative approaches to initialization (Figure 1). Additionally, the RP notes the SS model is essentially an SCA approach with more flexibility. While the RP agrees that, with more work, the SCA model is likely to be able to produce robust, unbiased estimates, the time and resource commitment is not worthwhile. Ultimately, the RP recommends not pursuing the SCA model further for the red drum stocks.

The RP expects the SS model to produce unbiased and robust estimates of the red drum stocks given that the Operating Model producing data was SSsim. The SS model for the northern region appeared to be unbiased when using perfect data from the operating model (Figure 2). The SS model for the southern region needs additional work to determine if the model can produce unbiased estimates while using perfect data from the operating model. The conclusions from the report for the SS south model are potentially uninformative because of the lack of a working model using the perfect data from the OM. The expectation is that the SS model will be able to reproduce the OM with further work. At that time, the sensitivity runs may need to be redone to reassess conclusions. Options to explore as the SAS determines what is leading to the inability to reproduce OM results include: 1) more years of model sensitivity runs, 2) consider impacts of growth curve biases on the results, and 3) explore the effects of different selectivity curves through time used for South Carolina, Georgia, and Florida. It would be worth exploring how the selectivity parameters influence model results, particularly given the changes in selectivity through time and across states in the southern region. In the absence of other ideas for improving the fit to data without error, it would be worth fixing all but one of the scaling parameters at their true values to make sure there are no gross specification errors present in the EM model configuration. If the one estimated scaling parameter (for example R0) is accurately reproduced in the EM, the remaining parameters could be iteratively opened to estimation in order to track down which ones are introducing bias into the model. Additional penalties (parameter priors) on troublesome parameters may be warranted.

The RP is particularly concerned with the unexpected outcomes in the "sensitivity runs" that remain unexplained for both the north and the south SS models. For the northern model, incorporation of the B2 (recreational live discards) composition data improves characterization of discards but results in more biased results, rather than less biased results. For the southern model, the use of the true growth information or model from the operating model does not improve the robustness of estimates.

Finally, the TLA may be a useful accessory tool because it shows no bias and provides recruitment information. TLA could be used as an annual, interim tool between assessments, as recommended by the SAS. TLA provides information on Recruitment Condition and SSB status and could be used as a tool to indicate the need for an assessment during periods of poor recruitment. The RP expresses concern over the methods for determining the reference points used in the evaluation of TLA performance. The grid search method uses information from the entire time series of the simulation, including the projection years. Therefore the TLA leverages information not available to the other models and would not be available to a TLA based on 'in situ' data. It would be informative to repeat the grid search using only the 'burn in' and pre-2023 periods to see if the reference points identified were similar to the ones identified in the presented assessment. The reduced time series grid search would be more directly comparable to the other assessment models and would be representative of options available in an 'in situ' application of the TLA.

During the Review Workshop, the RP made analytical requests to the SAS that were informative for determining the status of each of the models for use in red drum assessment and management. During Day 1, the RP requested running each of the estimation models with perfect or near perfect data from the operating model. This would allow the RP to assess how well the estimation model performs given a perfect dataset. The RP requested running data from the operating model with no error in the SS and SCA estimation models using only one iteration each for the north and south. The request included using all data from all years but with no observational error. The SAS provided the results, leading to the conclusions above regarding use of the SCA and SS models.

On Day 2 of the Review Workshop, the RP made additional analytical requests. The first was to continue to run the perfect data from the OM in the estimation model configurations for the southern region. In addition, the RP made requests intended to sleuth out why the southern region was not performing as expected or why the SCA model was not matching the operating model data well. First, the RP requested fixing the initial numbers at age at the true values for the northern SCA model in order to help with model initialization (Figure 1). Second, the RP requested fixing M at the true value whereby the value for the Age 7+ group was averaged across all of the available ages. The preferred average was the numbers-weighted M for the Age 7+ group for the south and north using the base model.

On Day 2, the RP also requested additional figures for consideration. First, the RP wanted to see the annual SPR values instead of the three year average SPR values. Second, the RP wanted to double check what SS was doing with the SPR calculations and requested the values be computed using a manual SPR calculation in a spreadsheet. Finally, the RP requested that growth and B2 be calculated annually. These requests were made in order to guide future work on the models in preparation for future red drum stock assessments.

Finally, the RP recognizes the spatial structure of the models needs further exploration and future assessments may or may not have the same structure explored here. Given the analyses explored for the simulation assessment, it was difficult to properly evaluate the most robust choices for spatial delineation and spatial assumptions within the modeling framework. Future exploration of the decisions regarding spatial assumptions should include analyses of the tagging data and the consideration of one model versus separate northern and southern models. Several capabilities within Stock Synthesis could be explored. One example could be one model with limited movement, but two separate areas for estimation of life history parameters and fishing mortality rates, plus the incorporation of tagging data. The single model could be set up to leverage all of the data available for the species while still allowing for differential management and population dynamics of red drum in the north versus the south. Another example could be two separate models, as presented here, one each for the north and the south, with tagging data incorporated.

9. **Review recommendations on future monitoring provided by the Technical Committee and comment on the appropriateness and prioritization of each recommendation. Provide any additional recommendations warranted.**

This TOR is partially addressed. The RP could not fully evaluate the simulation results for the southern area due to lack of a converged model that could accurately reproduce the OM when given data without observation error. Results from the future monitoring prioritization study are counter-intuitive and therefore could not be fairly interpreted. The RP feels the longline survey is very likely to be important to the assessment because it is the only source of information for adult fish. However, the simulation study indicates the long line data are not helpful to the assessment. Removing long line data made little or no difference to the results. Also, the RP feels additional length composition data from recreational discards should help the model inform recreational discard selectivity, and improve model performance. Counter to expectations, simulation results show increased bias relative to the OM when recreational discard composition data are added to the northern model. The RP feels it is important to understand why these results occurred before recommending a prioritization of future monitoring efforts.

One additional option to explore is the creation or collection of data to inform trends and selectivity of fish in the 70-90 cm range. The sampling gears and methods used to collect data for red drum generally do not catch large numbers of fish in the 70-90 cm range. The RP is concerned the range of ages in that size class is not well characterized. Collection of data from the 70-90 cm size range (28-35 inches) will likely provide information on age, trends in abundance, and selectivity across gears. This information will in turn lead to better, more robust analyses of growth.

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

This peer review panel report fulfills the requirements under this term of reference. The RP has provided detailed information for each review panel term of reference. The report was completed in the allocated time frame.

Following the Review Workshop, the Assessment Committee needs to work on fitting the SS southern model to the "perfect" data from the operating model, in order to show the estimation model can reproduce the truth from the operating model. Once that work is done, the Committee can move forward in considering our recommendations for the assessment of red drum in the northern and southern regions.

REFERENCES

Conn, P. B. 2010. Hierarchical analysis of multiple noisy abundance indices. Canadian Journal of Fisheries and Aquatic Sciences 67: 108-120.

Fabens, A. J. (1965). Properties and fitting of the von Bertalanffy growth curve. *Growth*, *29*, 265-289.

Kirkwood, G. P. (1983). Estimation of von Bertalanffy growth curve parameters using both length increment and age–length data. *Canadian Journal of Fisheries and Aquatic Sciences*, *40*(9), 1405-1411.

Lester, N. P., B. J. Shuter, and P. A. Abrams. 2004. Interpreting the von Bertalanffy model of somatic growth in fishes: the cost of reproduction. Proceedings of the Royal Society https://doi.org/10.1098/rspb.2004.2778.

Schueller, A.M., E.H. Williams, and R.T. Cheshire. 2014. A proposed, tested, and applied adjustment to account for bias in growth parameter estimates due to selectivity. Fisheries Research 158: 26-39.

Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Research*, *210*, 143-161.

FIGURES

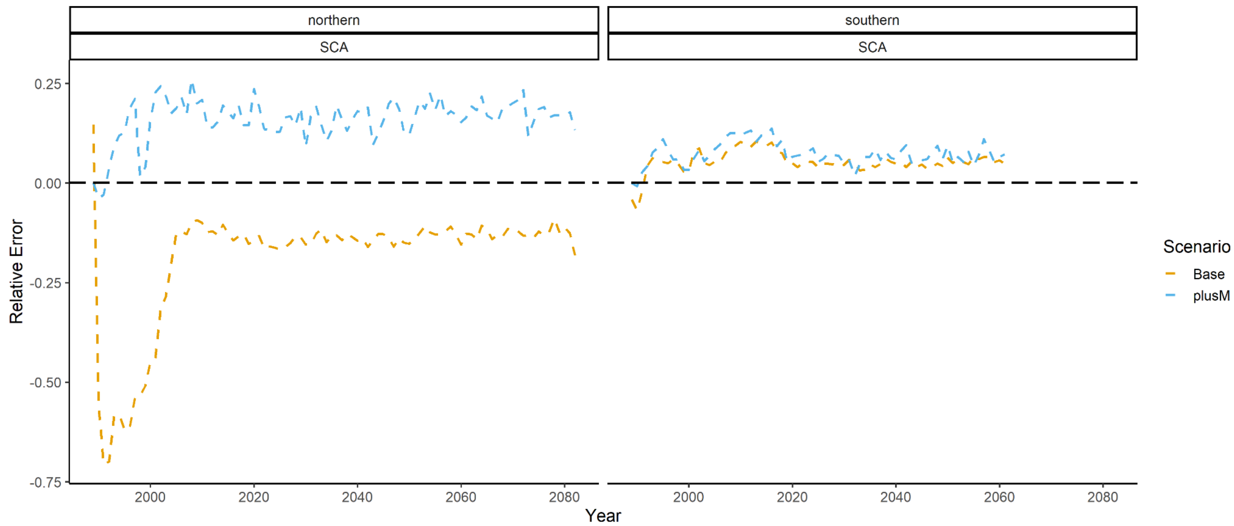


Figure 1. A plot of the relative error in sub-adult abundance for the northern and southern SCA models demonstrating that fixing parameters can lead to reduced bias in the early part of the time period for the north. This likely indicates something amiss with the initialization.

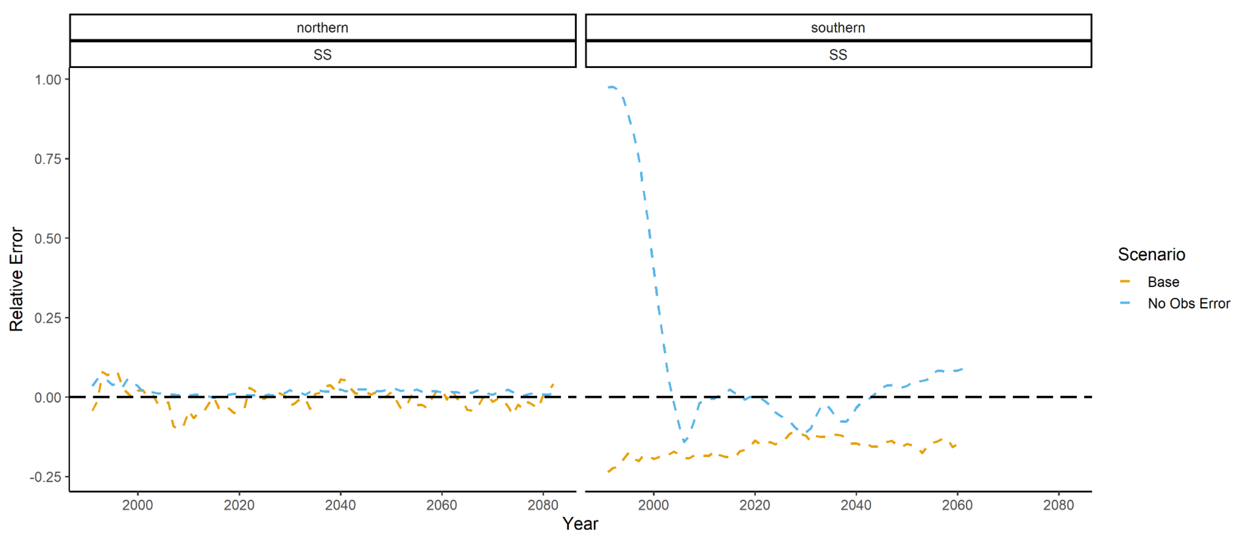


Figure 2. A plot of the relative error in the three year F ratios for the northern and southern SS models demonstrating that the northern model was able to produce unbiased results when using the perfect data from the operating model.