SEDAR

Southeast Data, Assessment, and Review

SEDAR 49

Gulf of Mexico Data-limited Species:

Red Drum, Lane Snapper, Wenchman, Yellowmouth Grouper, Speckled Hind, Snowy Grouper, Almaco Jack, Lesser Amberjack

SECTION V: Review Workshop Report

November 2016

SEDAR
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405

Table of Contents

Table of Contents 2
1. INTRODUCTION 2
1.1 WORKSHOP TIME AND PLACE 2
1.2 TERMS OF REFERENCE 2
1.3 LIST OF PARTICIPANTS 3
1.4 LIST OF REVIEW WORKSHOP WORKING PAPERS AND DOCUMENTS 4
2. REVIEW PANEL REPORT 5

1. INTRODUCTION

1.1 WORKSHOP TIME AND PLACE

The SEDAR 49 Review Workshop was held November 1-3, 2016 in Miami, Florida.

1.2 TERMS OF REFERENCE

1. Review any changes in data following the Data/Assessment workshop and any analyses suggested by the workshop. Summarize data as used in each assessment model. Provide justification for any deviations from Data/Assessment Workshop recommendations.
2. Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions, and consider the following:
   a) Are data decisions made by the DW and AW sound and robust?
   b) Are data uncertainties acknowledged, reported, and within normal or expected levels?
   c) Are data applied properly within the assessment model?
   d) Are input data series reliable and sufficient to support the assessment approach and findings?
3. Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data, and considering the following:
   a) Are the data-limited methods scientifically sound and robust?
   b) Are the methods appropriate given the available data?
   c) Are the data-limited models configured properly and used in a manner consistent with standard practices?
   d) Are the quantitative estimates produced reliable? Does the method produce management metrics (e.g. OFL, ABC) or other indicators (e.g. trends in F or Z,

probability of overfishing) that may be used to inform managers about stock trends and conditions?

4. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
   - Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
   - Ensure that the implications of uncertainty in technical conclusions are clearly stated.
5. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.
   - Clearly denote research and monitoring that could improve the reliability of future assessments.
   - Provide recommendations on possible ways to improve the SEDAR process.
6. Consider whether the stock assessment constitutes the best scientific information available using the following criteria as appropriate: relevance, inclusiveness, objectivity, transparency, timeliness, verification, validation, and peer review of fishery management information.
7. Provide guidance on key improvements in data or modeling approaches that should be considered when scheduling the next assessment.
8. Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference.

1.3 LIST OF PARTICIPANTS

Workshop Panel

Luiz Barbieri, Chair ..... Gulf SSC
Panayiota Apostolaki ..... CIE Reviewer
Yong Chen ..... CIE Reviewer
Jamie Gibson ..... CIE Reviewer
Kai Lorenzen ..... Gulf SSC
Joe Powers ..... Gulf SSC

Analytic Representation

Skyler Sagarese ..... SEFSC, Miami
Jeff Isely ..... SEFSC, Miami
Shannon Cass-Calay ..... SEFSC, Miami

Appointed Observers

Ben Blount ..... Gulf SSC
Claudia Friess ..... Gulf Appointee

Attendees

Shanae Allen .....FWRI
Jay Grove .....FWC
Bill Harford.....Univ. of Miami
Matthew Johnson .....SEFSC
Mike Larkin ..... SERO
Michelle Masi .....FWRI
Kevin McCarthy.....SEFSC
Michael Schirripa.....SEFSC
Matthew Smith.....SEFSC
Beth Wrege .....SEFSC

Staff

Julie Neer ..... SEDAR
Ryan Rindone.....GMFMC Staff
Charlotte Schiaffo .....HMS

1.4 LIST OF REVIEW WORKSHOP WORKING PAPERS AND DOCUMENTS

Documents Prepared for the Review Workshop			
SEDAR49-RW-01	Revised Results for the Generic Implementation of Itarget0 and Ltarget0 for Lane Snapper, Wenchman, Lesser Amberjack, and Almaco Jack	Skyler R. Sagarese, J. Jeffery Isely, and Matthew W. Smith	21 October 2016

2. REVIEW PANEL REPORT

Executive Summary

The Review Workshop Panel was presented outputs and results of the SEDAR 49 stock assessment of Gulf of Mexico data-limited species: Red Drum, Lane Snapper, Wenchman, Yellowmouth Grouper, Speckled Hind, Snowy Grouper, Almaco Jack, and Lesser Amberjack. Multiple analytical models were used to conduct this assessment. The Data-Limited Methods Toolkit (DLMtool), a software program that allows evaluation of the performance of multiple data-limited assessment models in a simulation environment using management strategy evaluation (MSE), was the primary modeling platform used in this assessment to estimate reference or target catch levels. In addition to the DLMtool, a mean length estimator approach assuming non-equilibrium conditions was used to estimate total mortality from length-frequency data. Lastly, a catch curve analysis was employed where possible to estimate the total mortality rate (Red Drum only). Data used in the assessment include stock identification and life history information, fisheries catch and effort data, abundance indices, as well as assumptions about stock depletion and for some methods, choice of a reference period for indices or mean length information. In general, the assessment input data series are reliable and were applied properly given the data-limited assessment approach used. The data and information requirements for use of the DLMtool appear substantial relative to the information that is available for these specific stocks. Uncertainties in most data inputs were acknowledged and reported and most are within expected levels. Possible exceptions are uncertainties in inputs regarded as 'assumptions' rather than data, such as the depletion level or choice of index reference period. Although the Review Panel concluded that the SEDAR 49 assessments represent the best scientific information available it also recognized that the methods used only provide general guidance towards catch advice. Therefore, the outcomes of this analysis do not correspond to the traditional management estimates produced in data rich assessments (e.g., MSY or its proxy). Further, the DLMtool approach is still under development and adjustments to better fit Gulf of Mexico stocks are still required.

2.1 Statements Addressing Each ToR

1. *Review any changes in data following the Data/Assessment workshop and any analyses suggested by the workshop. Summarize data as used in each assessment model. Provide justification for any deviations from Data/Assessment Workshop recommendations.*

The data and analyses reviewed and initiated during the Data/Assessment workshops were examined during this Review Workshop. There were a number of additional tests that were requested to help guide the Review Panel in deliberations on the efficacy of the analyses. All of these may be categorized as either: (1) additional diagnostics for evaluating the methods used; or (2) additional sensitivity analyses to better understand the uncertainty of the methods as they

were applied to the species/stocks of concern. All sensitivity analyses were conducted for all species except for alternative values of M for red drum.

The additional diagnostics that were requested were to include the interquartile range in the simulation trajectories instead of 5th and 95th percentiles; and to examine the trajectories of selected individual simulation runs. Both of these cases were desired since the trajectories expressed as simply a median and 5th and 95th percentiles for each year tends to mask the dynamic behavior that might occur in the simulation trajectories. Interquartile ranges narrow the interval and are probably more akin to risks addressed in most decision-making frames. Adding the interquartile range to the plots with the 5th and 95th percentiles would better display the distribution of the simulation results.

Examination of individual runs are useful for evaluating the variability exhibited by a simulated population from year to year and provides a qualitative method for evaluating the plausibility of the simulated population trajectories.

Additionally, simulation sensitivities were included. The first was to impose annual variability in growth parameters, specifically L_∞ of the von Bertalanffy growth equation. This was done because growth is known to vary through time for some species. Also, the panel wanted to look at potential differences in outcomes if there was plasticity in growth that responded to the environment. The variability imposed was approximately 15-20%. Results show an increased uncertainty in performance measures and the probability of achieving them.

The panel also wanted to better understand the performance of the methods when the index of abundance was of poorer quality. The original tests were conducted with a CV of 24%. Therefore, additional runs were done in which the CV of the observation error on the index was increased from the original tests to either 50% or 100%. Performance measure uncertainty increased, but not greatly so.

Ideally, one would do a stock assessment on a regular basis and then change catch recommendations based on the stock assessment. With data poor species this is not possible, so the methods being utilized are designed to provide a catch recommendation based on limited index information from a fixed decision interval. The initial tests used ten years. At the request of the Review Panel a three-year decision interval was conducted rather than the original 10-year interval. Results from these sensitivity analyses were not consistent for all species. Increasing the frequency of the assessment did increase the frequency at which catch recommendations could change, which for some species did change the longer term yield (e.g. Red Drum).

The methods utilizing an index are predicated on the index being proportional to abundance. But what happens if it is not? The beta parameter is a simple way to impose a nonlinear relationship between abundance and the index. This is done by making the index proportional to the abundance exponentiated by beta. Betas less than 1 imply hyper-stability in which large changes in biomass are not reflected by large changes in the index (the index is more stable than the biomass). In base model runs, a uniform distribution for beta was assumed with bounds of 0.33

and 3.0. At the request of the Review Panel betas values of 1.0 were examined. In general, the probability of not overfishing, the probability that the biomass was greater than 50% of Bmsy and yields all increased very slightly when beta was fixed at 1.0, although differences were negligible. .

The Panel wanted to see a run in which the steepness parameter is fixed at the lower bound of the range that was originally tested. As expected, in most, but not all cases (e.g. Lane Snapper, LstepCC0 model), the probability of not being overfished decreased relative to the base case. For some species, the number of operating models meeting the selection criteria changed when steepness was fixed at a lower value..

The base case red drum utilized a natural mortality rate, M, of 0.06. The panel felt that given that most of the fishery is inshore where natural mortality rates would be expected to be higher. Therefore, a sensitivity run in which M ranged from 0.16 to 0.184 was conducted.

In general, these analyses showed some sensitivity in the results to input variation. Nevertheless, key performance measures such as the probability of not being overfished were usually in an acceptable range (~80%), although in some cases the options for operating models did change. However, this is predicated on the specification of initial depletion related to the classification of species as specified by the Gulf Council's ABC Control Rule tiers 3-A or 3-B.

2. *Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions, and consider the following:*
   a) *Are data decisions made by the DW and AW sound and robust?*

Data used in the assessment include stock identification and life history information, fisheries catch and effort data, abundance indices, as well as assumptions about stock depletion and for some methods, choice of a reference period.

Stock Identification

For all stocks, a single, separate, Gulf stock was assumed. For Red Drum, there is evidence of genetic divergence in the northern GOM, but specific populations have not been delineated. Some uncertainty was noted with respect to stock identification of Lane Snapper (some genetic evidence for separate Western and Eastern Gulf stocks, as well as for hybridization with Yellowtail Snapper). No genetic or other data suitable for stock identification were available for the remaining species considered in this assessment (Wenchman, Yellowmouth Grouper, Snowy Grouper, Speckled Hind, Lesser Amberjack, and Almaco Jack). Stock structure information for more extensively studied, related species was used to support the single stock assumption for the latter species and or Lane Snapper (SEDAR 2016). The stock identification decisions appear practical in the light of very limited data.

Life History Information

<u>Growth</u>: Body growth data were available for all stocks except Lesser Amberjack (growth information from the South Atlantic used) and Almaco Jack (no growth information at all,

growth inferred from information about Greater Amberjack). Growth was described using a constant, standard von Bertalanffy growth function. This is the only growth modeling option offered by the DLMtool and is in line with common practice even in data-rich assessments in the Southeastern U.S. Growth in Red Drum has previously been shown to be better described by a bi-phasic model (Porch et al. 2002). Simulation testing during the Review Workshop of implications of temporal variation in growth for the performance of management procedures has shown that growth variation reduces the performance of procedures involving the use of mean length indices. Empirical analysis of temporal growth variation at the stock level could help identify stocks in which management procedures involving size indices may perform well (i.e. stocks with limited temporal growth variation).

Natural mortality: Natural mortality (*M*) was described by a constant rate for the exploited size/age groups, the only option available in DLMtool. Conversely, many data-rich assessments in the Southeastern region account for size/age-dependence in mortality rates. This is unlikely to be a major concern for the data-poor assessments, however. The only case where use of an age-dependent *M* could affect results is in the case of Red Drum since harvesting is largely restricted to juveniles which may have different natural mortality rates than the larger/older individuals in the spawning stock. No direct estimates of *M* were available for any stock except for Red Drum where, given closure of the fishery in federal waters, total mortality *Z* in the spawning stock may approximate *M*. Only one empirical estimator was used to generate *M* estimates: the revised Hoenig estimator of Then *et al*. 2014. Uncertainty in *M* was characterized as the range of point estimates obtained from the revised Hoenig estimator for plausible values of maximum age. The uncertainty generated in this way is likely to underestimate true uncertainty in *M* because the *M* estimator itself is associated with prediction uncertainty not reflected in the range of point estimates. Moreover, the use of only one empirical *M* predictor, as opposed to multiple predictors based on a suite of different life history characteristics, may underestimate uncertainty and represents a departure from previous practice. There is ongoing research about the most appropriate approach to estimation of *M* and best practices are expected to continue to evolve..

Maturity: Maturity (length at 50% maturity) information was available from biological sampling for all species except Wenchman, Lesser Amberjack and Almaco Jack. Information from related or similar species was used where information from direct sampling was not available. An alternative approach would have been to use life history invariants.

Steepness: No direct estimates for steepness (*h*) were available for the stocks considered. Plausible ranges for *h* were determined from reviews conducted as part of previous SEDAR assessments and from comparative information on related species.

Removals (Landings and Dead Discards)

Total removals (in weight) were calculated as the sum of commercial landings + commercial dead discards + recreational landings + recreational dead discards. Uncertainty in total removals was estimated by propagating uncertainty estimated for individual components.

Commercial landings were constructed using data housed in NOAA's Southeast Fisheries Science Center's Accumulated Landings System (ALS). The ALS includes landings data beginning in 1962, the terminal year for SEDAR 49 was 2014. Uncertainty estimates were provided for the landings of each species and accounted for species misidentification, landings reported by species group, and differences among states in the implementation of trip ticket programs. The workgroup used expert opinion to estimate landings uncertainty for each species. For most species, the commercial landings data were considered adequate for assessment analyses.

Recreational landings were obtained from multiple sampling programs including the Marine Recreational Fisheries Statistics Survey (MRFSS) and the Marine Recreational Information Program (MRIP), the Southeast Region Headboat Survey (SRHS), the Texas Parks and Wildlife Department (TPWD) and the Louisiana Creel Survey. The MRFSS/MRIP provided a long time series of estimated catch-per-unit effort, total effort, landings, and discards for six two-month periods (waves) each year, starting in 1981.

Annual removals associated with dead discards were obtained by multiplying annual numbers of discarded live fish with recommended discard mortality rates and average weights of discarded fish. Discard mortality rates were determined by consensus agreement among data workshop attendees. The recommended values were based on direct fisher input and review of relevant studies. For most of the species, field estimates of discard mortality rates were unavailable and mortality rates associated with similar species were discussed as proxies.

Fishing Effort

The fleet that accounted for the largest proportion of the total removals was selected as the representative fleet for each species. Fishing effort was summed by year for each of the representative fleets. The recreational fishery was recommended by the DW to be the most representative for Red Drum, Lane Snapper, Almaco Jack, and Yellowmouth Grouper. Commercial fisheries were recommended as the most representative for Speckled Hind (bottom longline), Snowy Grouper (bottom longline), Lesser Amberjack (vertical line), and Wenchman (finfish trawl). The effort time series was selected based on concurrent landings information from both the commercial and recreational fisheries. Effort data decisions are sound and well documented.

Abundance Indices

Abundance indices were potentially available from a variety of fisheries-independent and fisheries-dependent surveys. Fisheries-independent surveys considered included:

<u>SEAMAP Summer Groundfish Survey</u>: A collaborative effort between federal, state and university programs, designed to collect, manage and distribute fishery-independent data throughout the region. This semi-annual groundfish trawl survey provides a valuable source of fisheries-independent information on many commercially and recreationally important species throughout the northern Gulf of Mexico (GOM).

MSLABS Small Pelagics Survey: The Southeast Fisheries Science Center (SEFSC) Mississippi Laboratories (MSLABS) Small Pelagics Survey was initiated in October of 2002 as an outer shelf and upper slope survey (i.e., between 110 and 500 m station depth). The MSLABS Small Pelagics Survey was selected to provide an abundance index for Wenchman Snapper.

SEAMAP Reef Fish Video Survey: The SEAMAP reef fish video survey provides an index of the relative abundances of fish species associated with topographic features (e.g., reefs, banks, and ledges) located on the continental shelf of the Gulf of Mexico (GOM). SEAMAP Reef Fish Video Survey indices were produced for all SEDAR 49 species with the exception of Red Drum. The SEAMAP Reef Fish Video Survey was selected to provide abundance indices for Yellowmouth Grouper, Snowy Grouper, Lesser Amberjack and Almaco Jack.

NMFS Panama City Laboratory Trap and Camera Survey: Fishery-independent trap survey of natural reefs on the inner shelf of the eastern Gulf of Mexico off Panama City, FL. This survey provides and age-based annual index of abundance for pre-recruit (age 0-3) reef fish. No abundance indices based on this survey were recommended for the data-poor stocks in SEDAR 49.

DISL Bottom Longline Survey: Bottom longline survey operating monthly in the coastal waters of Alabama and Mississippi as well as federal offshore waters from May 2006 through the present by the Dauphin Island Sea Lab (DISL). This survey provides nominal catch per unit effort (CPUE) for Red Drum. The DISL survey was selected to provide an abundance index for Red Drum.

Fishery-dependent surveys included:

Headboat Survey: The Headboat Survey covers the Gulf of Mexico headboats starting in 1986. Total catch per trip is reported in logbooks provided to all headboats. The Headboat survey was selected to provide an abundance index for Lane Snapper.

Marine Recreational Fisheries Statistics Survey (MRFSS)/ Marine Recreational Information Program (MRIP): The MRFSS began in 1981 and provides information on participation, effort, and species-specific catch. No abundance indices based on this survey were recommended for the data-poor stocks in SEDAR 49.

Commercial Logbook: The NMFS Gulf of Mexico Reef Fish Logbook Program collects catch and effort data by trip for permitted vessels that participate in fisheries managed by the Gulf of Mexico and South Atlantic Fishery Management Councils. No abundance indices based on this survey were recommended for the data-poor stocks in SEDAR 49.

Despite of consideration of the above, wide range of surveys, no abundance indices were recommended for Snowy Grouper and Speckled Hind. Potential abundance indices were constructed evaluated carefully and the decision process is well documented.

Size-Structure Indices

Some data-limited approaches in the DLMtool use length composition in conjunction with the mean length estimator to calculate current stock abundance or current stock depletion. Length samples were obtained from a variety of fishery-independent and fishery-dependent data sources for all eight species under assessment.

Size structure information was obtained from the NOAA Fisheries Trip Interview Program (TIP) for commercial landings. Length samples for recreational fisheries were obtained from the MRFSS/MRIP surveys, the Southeast Headboat Survey, the TPWD, the Florida Fish and Wildlife Conservation Commission (FWC), the Gulf States Marine Fisheries Commission FIN database (GFIN), and the TIP database.

Where available, length samples were also obtained from fishery-independent surveys including the NMFS small pelagics survey, SEAMAP groundfish survey, SEAMAP reef video survey, Panama City video survey, and Panama City trap survey. For all species except Red Drum and Wenchman, annual sample sizes were too small for analysis.

In addition to length composition data, the DLMtool and mean length estimator approach require information on the selectivity at length including the size at first capture (or size at first recruitment to the gear) and the size at full recruitment to the gear. Length frequency plots for each fleet and gear were used to inform decisions about the size at full recruitment for each species and gear since the assessment approach requires the characterization of a fleet considered most representative in terms of selectivity and exploitation pattern for the simulation.

Stock Depletion

The evaluation of management procedure in the DLMtool requires an estimate of current depletion of the stock. Estimates of current depletion were not available for the majority of the species under assessment during SEDAR 49. An estimate for Red Drum was available from the 2015 FWC assessment which assessed the stock status in Florida waters. For the remaining species under consideration for SEDAR 49, depletion estimates were derived by using ‘similar species/stocks’ that have been assessed using Stock Synthesis as proxies.

The rationale underlying the choice of proxy stocks for depletion estimates is stated only in general terms, selection criteria for the decision process are not well documented. Since the identification of proxies involves consideration of the fishery as well as biological characteristics, such decisions are potentially complex and should be guided by well-defined criteria. Moreover, since stocks assessed by Stock Synthesis tend to be strongly targeted and carefully managed, it is unclear how representative the depletion levels of such stocks are for the data limited, often non-targeted and barely managed stocks.

Reference Period

Some management procedures rely on comparison of abundance or mean length indices to index values derived for a reference period that essentially provides a baseline status associated with the index values during that time period. The choice of a reference period then becomes an

important input. Reference periods specified in GMFMC (2011) for seven of the eight species were used, an approach that facilitated comparison of method performance between feasible methods considered during SEDAR 49 and the method currently being used. A reference period for the eighth species, Red Drum, was chosen at the Assessment Workshop. The analytical team explained that the periods were chosen to represent periods of approximately constant catch. The rationale for this criterion is not entirely clear. Indeed, different arguments have been advanced for this criterion, including that stable catches represent conditions associated with MSY, a sustainable catch level, or at least stability in exploitation. None of these arguments are necessarily true: stable yield and biomass can be achieved at any sustainable level of exploitations (not just MSY), including in a state of severe yet sustainable overfishing. Stable catches can also be associated with increasing exploitation levels in a declining stock, i.e. they are not even necessarily associated with stability in the fishery. It is important that the rationale for the setting of reference periods and the specific criteria and decision processes are more explicitly motivated and reported.

b) Are data uncertainties acknowledged, reported, and within normal or expected levels?

Uncertainties in most data inputs have been acknowledged and reported and most are within expected levels. Possible exceptions are uncertainties in inputs regarded as ‘assumptions’ rather than data, such as the depletion level or choice of index reference period. These issues have been discussed in more detail above. Since such assumptions are important inputs to data-limited assessment, greater efforts should be made in future assessments to deal with uncertainty in ‘assumptions’ in the same rigorous and structured manner as is common practice with inputs regarded ‘data’.

The Data Workshop provided a particularly rigorous evaluation of the potential sources of life history. Sources of information were identified via a literature and a reliability rubric, based on sampling considerations, the quality of the data collection and analysis, and the overall reliability of the work, was used to score the work for providing life history parameters for use in the DLM tool. In this way, uncertainties in the life history information was fully acknowledged and clearly reported.

Similarly, the Workshops provided clear explanations of decisions about which surveys were most appropriate for each stock.

Uncertainties are broadly within expected levels. It should be noted that, due to the non-target nature of many of the fisheries and relatively low rates of encounter for many species, uncertainties are expected and found to be fairly large. The possibility of unquantified biases, e.g. due to misidentification of rare species, has been noted.

c) Are data applied properly within the assessment model?

The data are properly applied within the DLMtool, following guidance developed by the tool’s developers and other experts.

d) Are input data series reliable and sufficient to support the assessment approach and findings?

The input data series are generally reliable and sufficient to support the assessment approach and findings. However, the data and information requirements for use of the DLMtool are in fact quite variable and can be substantial for some of the more rigorous applications. The types of information available for the SEDAR 49 stocks shows that some of the Gulf stocks are too data poor for the more rigorous applications of data-poor methods. For stocks where data are sufficient to support use of the DLMtool, substantial efforts are required to prepare data inputs and arrive at well-founded assumptions.

3. Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data, and considering the following:

Overall, the methods used represent reasonable choices given paucity of data that limits the spectrum of quantitative approaches that could be used. The adoption of a simulation evaluation approach in the context of the DLMtool allows for a relatively abstract and high level consideration of management procedures which reflects the knowledge gaps and adds value to the assessment since it gives an indication of the procedures that might be fit for the nature of species and fisheries studied. However, the adopted approach does pose certain challenges both in terms of the way it is implemented/designed as well as its capabilities for simplifying assessment and development of catch recommendations for management. Furthermore, the Review Panel notes that this method is not meant to replace standard stock assessments and its use does not mean that data collection and knowledge should not be improved. More detailed analyses of this ToR is provided under each of the four questions below.

a) Are the data-limited methods scientifically sound and robust?

The DLMtool is the main package used in this analysis. The main equations in the population dynamics model are standard formulae that have been used extensively in the past and are scientifically sound. The management procedures (MSEs) considered have also been used in other studies and peer-reviewed as part of previous work (Geromont and Butterworth, 2014) so, the general concept is sound. However, those are empirical MSEs and their parameterization, as used in the DLMtool, has been adopted to support management of severely depleted stocks of medium productivity.

In addition to the DLMtool approaches, catch curve analysis was also employed to calculate total mortality. In principle, there is good understanding of the scientific basis and use of the operating model and MSEs in the DLMtool but there are still concerns about the implementation of the approach. Specifically:

- The translation of all the mathematical formulae into R code has not been checked by this Panel. Some parts of the code were discussed especially those involving internal boundaries

and checks that are hardwired into the code and lead to diversions from the main formulae influencing the results. An example is the adjustment of the fishing mortality to avoid extinction that the model does automatically. Those need to be checked and documented in detail to ensure that the performance of tested MSEs is not artificially enhanced. In particular, the equations in Section 3.1 of the SEDAR 49 Addendum that include the catch recommendation smoothing parameter, *w*, should be checked to see if they are appropriate for values of *w* not equal to 0.5 (not that this would not affect results in SEDAR 49 where 0.5 was the only value used).

- The approach is still under development and requires a very good understanding of the underlying concepts and their translation into source code, and that takes a lot of time so, it is not a quick shortcut to assessing data-poor species.
- The scalars of the formulae used to describe the MSEs tested in this assessment were generally default values intended to provide a generic approach to help overexploited species to recover. Therefore, there is no evidence that the same values for the scalars represent the best option for the type of species assessed in this exercise (*see also comments below*).
- See comments under ToR 7 below for a discussion of the parametrization of the stock-recruitment function. Although the use of the steepness parameter (*h*) was considered acceptable here the Review Panel recommends the use of a different approach for the next application of the DLMtool approach.

b) Are the methods appropriate given the available data?

The methods proposed aim to address paucity of biological and other information in data limited species. All the species considered here, with the exception maybe of one, could be assigned to that category so, the use of DLMtool is generally appropriate. The volume and extend of data for red drum was relatively greater than for the other species and that warrants further consideration to decide whether this stock can be treated as data-moderate. This does not render the DLMtool inappropriate but suggests that consideration of additional quantitative approaches could be of value to identify those that are more suitable.

The Review Panel did consider the role of the species in the context of the fisheries that affect them (e.g., by-catch species) and their dynamics and it is not clear whether the current methods provide a flexible enough mechanism for capturing those characteristics.

For some species, several management approaches met the performance criteria, however, the Review Panel believed that the management approaches that made use of relative abundance information in the form of a CPUE index or mean length information to provide a signal about the population response to future exploitation would better safeguard populations than those that do not use relative abundance information.

In particular, the following points are made:

- The methods and parameterization of models has been designed for target species and at least half of the species considered in this assessment were not targeted species. So, the type of information available or of use for this assessment differs from that for targeted species. For example, effort patterns characterizing the target species in the relevant fishery in which the study species are caught is an alternative source of information in addition to indices for the study species. The former could be included into the model but the current configuration accepts only one data series for fisheries so, that is not possible.
- The MSEs are configured with overexploited species in mind and mainly to describe long-lived species. This does not fit well the dynamics of some of the species considered here so, further work is needed to identify smoothing parameters and scalars that are more appropriate for short-lived species or species that are not heavily fished or targeted.
- For the Red Drum, the von Bertalanffy growth model is not the most appropriate as it does not describe the gender-changing characteristic of that species and the effect it might have on growth so, that is a limitation.

c) Are the data-limited models configured properly and used in a manner consistent with standard practices?

The values of the model parameters reflects the recommendations of the data workshop and in that sense, it is properly configured but given concerns about parameter values selected by the data workshop there are recommendations for further work to address them. Those include:

- The choice of L_∞ is not supported by catch at age data that for most of the species considered appear to include considerably higher values for fish length.
- Similarly, the CV for the growth parameters are unrealistically small so, this part of the model configuration needs to be revisited.
- With the exception of Red Drum, all simulations used the reference periods adopted by the Gulf Council and used them to determine changes in future catches. However, there is very little information about the state of nature that reference period represents and no clear justification for the choice of that reference period. In conventional assessments the reference period is set at a much earlier time period and is assumed to either reflect the state of the population that led to optimum production or, in some cases, the state of the population at almost unexploited conditions. The interpretation of model predictions will be affected by those assumptions and, therefore, the choice of the reference period need to be substantiated and an explanation provided for what state of nature it is supposed to represent.
- Temperate species were included in the meta-analysis used to find plausible values for the length at age equation and that might have introduced bias in the range of plausible values. However, given that understanding of the dynamics of the 8 stocks is limited the approach applied is still considered reasonable.

- As pointed out in ToR 2 above, the Hoenig estimator was chosen to calculate *M* values and that does not reflect common practices that consider more than one methodology to find estimates of *M*. The latter provides a more thorough view of the plausible range of values for *M* and it is recommended.
- From the relevant documents and discussions during the review workshop, it transpired that more data than those used for red drum existed. This suggests that the model for red drum does not reflect best available knowledge. It is understandable that given the large number of species being assessed here compromises in the data compilation and hence model configuration were inevitable. However, that weakened the value of the analysis. It is recommended that future assessments allow enough time to identify and compile all available data to strengthen model configuration.

The timeframes for the MSE simulations do not reflect the dynamics of some of the stocks (e.g. 40 years for a species that live for 5 years). It is recommended that simulation time be calculated as a function of generation time or a similar constant to better reflect the biology of assessed stocks.

d) Are the quantitative estimates produced reliable? Does the method produce management metrics (e.g. OFL, ABC) or other indicators (e.g. trends in F or Z, probability of overfishing) that may be used to inform managers about stock trends and conditions?

Yes, within the context of data limited approaches it provides guidance on management approaches that can be effective and if those are adopted they can be used to guide the decision for ABC. However, the outcome of these analyses do not correspond to the traditional matrices produced in data rich assessments (e.g., MSY, OFL, or ABC). The estimates produced in this assessment mainly concerned metrics that described the performance of alternative management procedures. That included probability of the population and yield to be above a pre-specified reference point (MSY), probability of not overfishing, and probability that the biomass will fall below a pre-specified limit for each of the MPs considered. Although the assessment does provide catch estimates, it is not clear how they relate to management quantities (OFL, ABC, etc.), so, the metrics are useful to inform managers but the way in which they will be incorporated into the decision-making process has not been clarified yet and it is expected to require an adjustment in the current procedure for setting catch quotas. In that context the Panel has made the following points:

- This is a methodology to guide decisions and help avoid overexploitation while the knowledge is built to develop a robust assessment.
- The interpretation and use of these results requires a different management paradigm as the tested methodology does not produce the metrics that calculated in a conventional stock assessment (e.g., B_{MSY}).
- However, it is of value since it provides signals about stock status and exploitation levels in the absence of absolute estimates about stock size and exploitation.

Further, in terms of informing management decisions, the choice of the metrics does not reflect well the fact that most of the species are by-catch/no-target species. For example, it is questionable whether achieving MSY is a realistic or relevant objective in these fisheries, although knowledge of MSY would help ensure stocks are not overexploited. Hence, there is a need to define how the relevant metrics are expected to inform management decisions and whether all metrics that have been calculated should be given the same weight when one decides on the best management procedure to use. Therefore, the Panel recommends that performance metrics and additional criteria are revisited and possibly adjusted to reflect the fact that these stocks are bycatch species and because of that, certain objectives such as avoiding overexploitation could be more important or relevant than achieving MSY.

The evaluation outcomes were tested under a range of scenarios and uncertainty levels and the main conclusions were not affected. That provides some assurance about the robustness of the estimates and the reliability of the outcomes of the MSEs in terms of the management procedures that are more appropriate for the assessed stocks. However:

- The influence of the constraints of the model (see previous section about hardwired checks in the source code) on probability density functions reduces the reliability of the results.
- Combining probability density distributions for catches that come from different MSEs is an arbitrary choice that does not have a clear justification and leads to recommendations for catches that have not been tested in the simulation evaluation exercise. Therefore, the Panel did not agree with the proposed approach that combined catch results from two or more management procedures. Additionally, if the catch recommendation is a single value associated with a single procedure, testing of that value in a simulation evaluation would also be warranted, particularly if the probability density distribution is wide.
- Also, see comments in ToRs 4 and 7 below regarding the potential impact of covariance in life history parameters on the outcome of model results.

4. *Consider how uncertainties in the assessment, and their potential consequences, are addressed.*

Uncertainty associated with the population, data, and assessment models was addressed via Monte Carlo simulation and sensitivity analysis in the SEDAR 49 stock assessment.

a) *Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.*

The SEDAR 49 assessment developed and employed a structured approach to systematically evaluate possible impacts of uncertainties associated with the parameters in the operating models and variables/data used in developing catch advice. This approach includes:

- Monte Carlo simulations were conducted to capture the uncertainties associated with the parameters used in the operating model to simulate the fisheries for evaluating the performance of 11 methods considered for developing catch advice. Uncertainties associated with some key life history parameters (e.g., von Bertalanffy growth parameters L_∞, K and t_0, with correlations of these three parameters being considered in random sampling; natural mortality rate (*M*); steepness (*h*); the beta parameter defining hyper-depletion/hyper-stability) and fishery parameters (i.e., total removals, length at first capture, and length at full capture) were quantified with lower and upper boundaries (or CVs) largely defined based on meta-analyses of existing data, previous studies and expert opinions. One thousand simulation runs were conducted with these model parameters being randomly drawn from the uniform distributions defined by these lower and upper boundaries.
- Uncertainty associated with the total removals for all the eight species was quantified with CVs defined in the Data Workshop based on the values defined for total commercial and recreational catches, and discard mortalities. The abundance indices from fishery-independent and/or fishery-dependent programs were also quantified for all the eight species based on the best information available at the Data Workshop.
- The quality of different data was quantified with reliability scoring systems at the Data Workshop based on source of the data, spatio-temporal coverages of sampling programs, sample sizes, likelihood of species misidentification, and other factors (e.g., changes in fishermen's fishing behaviors as a result of changes in management regulations). The semi-quantitative scores of data quality were used in the selection of feasible methods for catch advices.
- Sensitivity analyses were also conducted to evaluate the robustness of performance of feasible catch advice methods regarding uncertainties associated with scalars built in various methods. However, sensitivity analyses for the catch recommendation smoothing parameter *w*, which determines how the catch advice changes relative to the value of the abundance index, were not carried out. The AW did recommend additional tuning to meet specific performance criteria in future evaluations and this sensitivity could be carried out as part of this tuning. This parameter determines the catch rule.
- The simulation of fishery by the operating model is conditional on the assumed depletion level which is usually unknown. Possible impacts of violating the assumed depletion level were evaluated by running all three possible depletion scenarios (i.e., lightly, moderately, and heavily depleted) for each method identified as feasible for each species.
- All the methods for catch advice that were deemed feasible based on the data availability and quality were considered and evaluated for all the eight species in this study, indicating that variability associated with choices of catch advice methods were considered.

Although the coverage of uncertainty sources is very comprehensive for all the eight species in SEDAR 49, some extra analyses can be done to further improve our understanding of the impacts of uncertainties on the development of catch advice using the DLMtool:

- Evaluate all the default values and built-in constraints used for the methods included in the DLMtool software because these methods were developed for fisheries outside the Gulf of Mexico and their associated parameters are likely inappropriate. There is a need to carefully evaluate their suitability for Gulf of Mexico fisheries.
- Current simulations were run with uncertainty of all the sources being incorporated, which may make the identification of impacts of a single uncertainty source difficult, and a structured simulation design may be needed to isolate and identify impacts of an individual uncertainty source.
- Different levels/forms of uncertainty for some key parameters/data (e.g., annual variability in growth parameters, annual variability in total removals, different levels of variability in the index of abundance etc.) need to be considered to have a better understanding of impacts of these uncertainties.
- Possible correlations between the S-R parameter *h* versus *M*, and parameters quantifying reproductive potential may need to be considered in the simulations (see more detail on this topic under ToR 7 below).

b) Ensure that the implications of uncertainty in technical conclusions are clearly stated.

The possible implications of uncertainty of various sources in technical conclusions are clearly stated in the selection of methods for developing catch advice, and the relevant mechanisms were discussed in the Review Report and at the Review Workshop.

5. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.

a) Clearly denote research and monitoring that could improve the reliability of future assessments.

b) Provide recommendations on possible ways to improve the SEDAR process.

Sea sampling programs to better quantify discards and discard mortality for all the eight species.

The choice of reference time period for Tier 3A and Tier 3B stocks needs to be re-visited given the new information available and possible changes in the ecosystems.

The operating model simulates the population dynamics of a given species conditional on the assumed depletion level which is usually unknown. Although the base case scenario for depletion level was developed for each species based on the best available information and a sensitivity analysis was conducted for alternative depletion levels, a reality check may be

necessary to help simulate a fishery that realistically reflects the dynamics of fishery of interests. Reliable information on the fishery and population (e.g., temporal trend of fishing efforts, fishery-dependent and fishery-independent abundance indices and biological information such as age- and length compositions) needs to be collected to help define possible depletion level. These data can be used to tune the operating model parameterization to improve the fishery simulation realism by the operating models. Further, a number of surveys were considered at the DW but not all of them were deemed appropriate to inform a stock assessment. It is important to revisit the design of the surveys to ascertain whether changes could be made to get more value out of those surveys. The Review Panel also recommends that more time is spent to identify the methodology and indicators that are best for the type of exploitation and species we have. Trying to calculate MSY and other conventional metrics might not be the most appropriate approach especially for species that are caught as bycatch. Similarly, collecting all the data that are needed to do a proper stock assessment is a very big task and it is important to identify some interim approaches such as using indicator species (to represent a complex of species) or maybe use the status of the targeted stock as a proxy for the status of the by-catch species.

6. *Consider whether the stock assessment constitutes the best scientific information available using the following criteria as appropriate: relevance, inclusiveness, objectivity, transparency, timeliness, verification, validation, and peer review of fishery management information.*

The Review Panel considers that the SEDAR 49 assessment constitutes the best scientific information available, and fulfils the following criteria:

Relevance: application of the DLMtool to provide quantitatively-based catch advice (albeit data-limited in nature) to Gulf of Mexico stocks is a highly relevant step in the evolution of stock assessments in the region.

Inclusiveness: in general, analyses conducted during SEDAR 49 include all data that have been quality assured and proved adequate for use in the assessment. This includes data from State as well as Federal sampling schemes, where needed. Additionally, there are opportunities for stakeholders or the public to provide input into the process.

Objectivity: the DLMtool is a highly objective procedure based on well tested statistical modeling principles, and using data sets and principles that have been well documented and reviewed through the SEDAR data and assessment process. Possible exceptions are uncertainties in inputs regarded as ‘assumptions’ rather than data, such as the depletion level or choice of index reference period. .

Transparency: all outputs of the data, assessment and review workshops in SEDAR 49 are fully documented and publicly available. The discussions at the review workshop are also recorded for the administrative record. All data sets are thoroughly explored and the quality of data on which the assessment is based is documented and transparent, as are all decisions

related to the choice of assessment model, how it is implemented, and the results of the different runs and sensitivity and uncertainty analyses.

Timeliness: The SEDAR process in general is arranged to provide timely fishery management advice where it is needed, and to ensure that assessments are benchmarked and reviewed at appropriate intervals.

Verification: The SEDAR 49 assessment was structured and conducted as to provide deliverables that comply with legal requirements under the Magnuson Stevens Act (2007) for developing and monitoring of fishery management plans and providing information on stock status. However, given the data-limited nature of the methodologies applied estimation of standard reference points for catch advice was not achieved.

Validation: The SEDAR 49 assessment process was implemented to meet the needs of fishery managers for peer-reviewed stock assessments and associated catch advice. The process is open and fully transparent to the fishery managers and to stakeholders from commercial and recreational fisheries, conservation groups or others with a stake in the outcomes and who have opportunity to give their views on record.

Peer review: The SEDAR 49 assessment process includes full peer-review by experts appointed from the Center for Independent Experts (CIE, University of Miami) and the GMFMC SSC. The review panel report and the independent CIE reviews are publicly available.

7. *Provide guidance on key improvements in data or modeling approaches that should be considered when scheduling the next assessment.*

The eight species assessed during SEDAR 49 are all data poor and therefore improvements to the data for these species would be expected to improve their respective assessments. Opportunities for collecting samples to improve life history parameter estimates are outlined under ToR 5 above (Research Recommendations). In summary, these include increased dockside and/or at-sea sampling for most of the species, and development of sampling protocols for species encountered in existing surveys such as the NMFS Pascagoula Groundfish and Small Pelagic surveys.

With respect to removals by commercial, recreational and other fisheries, discard mortality and quantification of uncertainty in the discard estimates are two sources of uncertainty in the assessments. Because fishery removals play a key role in determining current abundance levels, improvements to the removal estimates would be expected to improve the catch recommendations. Additionally, improvements in the information about the size and age of fish removed by the fisheries should lead to better estimates of fisheries selectivity, thereby reducing uncertainty in the population-level effects of different catch levels.

Although the DLMtool does provide a mechanism for evaluating management procedures and operating models with very limited amount of data, it does not provide a real-world evaluation

of whether the procedures and models are achieving the management objectives. Although metrics such as mean length can potentially be used as an abundance proxy, it is not clear, at least for some species, whether a change in mean length might be indicative of increased survival, a change in recruitment, or pulsed recruitment. Indices of relative abundance (particularly in combination with age or length data) would be expected to be most indicative about changes in abundance and whether management goals are being achieved. For species for which abundance indices are available, ensuring that relative abundance indices are indicative of abundance (by ensuring all habitat types are appropriately sampled, for example) would strengthen the assessments. Additionally, given both their within-year and among-year variability, evaluation of the precision of indices with respect to their utility for detecting changes in abundance would also be expected to improve the assessments.

As used for this assessment, the DLMtool was configured to primarily evaluate management procedures rather than to specifically provide catch advice. For this reason, the DLM tool produces very different outputs than traditional assessment models or the approaches currently used to provide catch recommendations for data-limited stocks. The long-term simulations are, in many ways, more like population viability analyses used in conservation biology than traditional fishery stock assessment models. The Review Panel suggested that some modifications and additions to the approach would be expected to significantly improve catch recommendations using the method.

The DLMtool does provide an evaluation of potential operating models based on a set of performance metrics. As applied in SEDAR 49, there was a constraint on the amount the catch could change, which limited the set of operating models deemed acceptable. For example, an extremely low constant catch would meet the three performance metrics used to choose potential operating models, but the constant catch scenario was not always considered appropriate likely due to this constraint. A broader range of catch recommendations for each operating model might increase the number of options available for operating models. Additionally, the effect of constraints on the catch recommendations from a single operating model was not fully explored during SEDAR 49, but is necessary to be able to interpret the probability distributions for the catch recommendations.

The model output includes a probability distribution for the catch recommendation associated with each potential operating model. However, because the tool is evaluating potential operating models, it is not evaluating whether a specific catch recommendation would meet the performance metrics (there is uncertainty associated with the catch recommendation). An additional step, involving feeding the specific catch recommendation back into the operating model would help ensure that performance metrics are met given the uncertainty in the operating model input parameters (use different random values). Sensitivity analyses to assumptions about depletion levels and other assumptions could also be carried out at this step.

As implemented in SEDAR 49, the probability of meeting the performance metrics was calculated across all years and simulations independently. However, each simulation is a potential realization of future conditions that either meets management objectives or does not. Performance metrics and standards, based on management goals, should be two tiered, including criteria that are applied to each individual simulation to determine whether it meets the metric or standard, as well as risk acceptance criteria applied across simulations based on the probability that the standard is met. For example, depending on management objectives, criteria applied within a single simulated trajectory could include: the proportion of the years during which the population is over-fished; the proportion of the years the population is in an over-fished state; the proportion of the years that the population is above or below some abundance threshold; or, in the case of rebuilding, whether a simulated population rebuilds within a specified timeframe. Each simulated population trajectory either meets the objective, or does not. The probability of meeting the objective can then be calculated as the proportion of simulated populations that meet the objective. This probability can then be compared with a risk tolerance criterion for that performance metric.

This application of the DLMtool provided much more information than was previously available for these species. For this reason, performance metrics could be developed that are situation specific (e.g. dependent on the depletion level, life history, whether fisheries are targeted or bycatch, distribution of fishing effort, trends in indices, etc.). Additionally, throughout the SEDAR 49 review meeting it was not clear how the catch recommendations should be used. The development of guidance on the interpretation of the catch recommendations from the tool as an OFL, ABC, ACL, ACT or some other value would aid in the utility of assessment results from the tool. Particularly given that a probability distribution is produced for the recommendation, the potential to use different percentiles from the distribution for different metrics could be explored. The interpretation of the output might also be situation-specific, and might differ among populations.

Life history parameter covariance is difficult to incorporate into the simulations and, if not fully specified, could result in parameters combinations that are biologically unrealistic. For example, the steepness parameter in the stock-recruitment (S-R) relationship depends on the slope at the origin of the S-R relationship, but also on the natural mortality rates, growth parameters, maturity parameters and length-weight conversion parameters, most of which were assumed uncorrelated in SEDAR 49. This individual issue would be addressed if the S-R relationship was parameterized in terms of the slope at the origin. More broadly, calculation of $SPR_{F=0}$, lifetime maximum reproductive rates or other similar metrics could provide a mechanism for filtering out combinations of parameter values that are biologically unrealistic, if limit values were included in the model.

For many of the operating models there are many control values that can be set which influence how the catch recommendation changes though time in each simulation. For example, in the case where abundance metrics or proxies (e.g. mean length), are available,

there are options for choosing the index limits and smoothing parameter values that determine the harvest control rules. In SEDAR 49 sensitivity analyses were carried for many of these options, although the model performance metrics were applied using default settings. Optimization methods could be developed that would choose the most appropriate values (given management goals and performance metrics), including the catch recommendation and assessment frequency, reducing the need to run a potentially large number of sensitivity analyses separately to find the best values.

Overall, the Review Panel believes that the data synthesis that occurred as part of SEDAR 49, and the application of the methods in the DLMtool has provided a lot of information that was previously unavailable for these species. However, tailoring the approach used in SEDAR 49 specifically for the provision of catch advice, along the lines suggested above, would be a significant step prior to the next assessment for these species.

8. *Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference.*

This report constitutes the Review Panel's summary evaluation of the stock assessment and discussion of the Terms of Reference. The Review Panel will complete edits to its report and submit a final document to the SEDAR program for inclusion in the full set of documents associated with SEDAR 49.

References

Geromont, H., and D. Butterworth. 2014. Generic management procedures for data-poor fisheries: forecasting with few data. ICES Journal of Marine Science 72:251-261.

Gulf of Mexico Fishery Management Council (GMFMC). 2011. Final Generic Annual Catch Limits/Accountability Measures Amendment for the Gulf of Mexico Fishery Management Council's Red Drum, Reef Fish, Shrimp, Coral and Coral Reefs, Fishery Management Plans (Including Environmental Impact Statement, Regulatory Impact Review, Regulatory Flexibility Analysis, Fishery Impact Statement). Gulf of Mexico Fishery Management Council, Tampa, Florida. 378 pp.

Then, A. Y., J. M. Hoenig, N. G. Hall, and D. A. Hewitt. 2014. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science.

SEDAR

Southeast Data, Assessment, and Review

SEDAR 49

Gulf of Mexico Data-limited Species:

Red Drum, Lane Snapper, Wenchman, Yellowmouth Grouper, Speckled Hind, Snowy Grouper, Almaco Jack, Lesser Amberjack

SECTION VI: Post-Review Workshop Addendum Report

November 2016

SEDAR
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405