Panel Review Report

October 8, 2018

Review of

A Stock Assessment of Eastern Oyster, *Crassotrea virginica*, in Maryland Waters of Chesapeake Bay

August 27-29, 2018

Annapolis, MD

Panel Members

Paul Rago[1] (Chair)

Daniel Hennen[2]

Daphne Munroe[3]

[1]MidAtlantic Fishery Management Council Science and Statistical Committee, retired, Northeast Fisheries Science Center, National Marine Fisheries Service, Woods Hole, MA

[2]Northeast Fisheries Science Center, National Marine Fisheries Service, Woods Hole, MA

[3]Haskin Shellfish Research Laboratory, Rutgers University, Port Norris, NJ

REPORT OUTLINE

Section 1.0 Executive Summary

The decline of oyster abundance in Chesapeake Bay is well known and significant public and private resources have been invested to support its revival (Kennedy et al., 2011). With considerable foresight, Maryland began a program to monitor abundance in 1939. Ironically there have been no stock assessments of the overall resource. This report represents the results of an external peer review by a panel of independent experts of an assessment of oysters in Maryland's portion of Chesapeake Bay. The assessment and review were mandated by the Sustainable Oyster Population and Fishery Act of 2016 (Maryland General Assembly). The assessment was conducted by the Maryland DNR in consultation with the University of Maryland Center for Environmental Science (UMCES) and was completed in early August, 2018. The review was conducted in late August, 2018 following Terms of Reference reviewed by the Maryland Oyster Advisory Commission.

Primary Conclusions

Overall, the Review Panel concluded that all Terms of Reference had been met. The Review Panel supported the conclusions of the Assessment Team and agreed that they had fully utilized the available data at an appropriate temporal and spatial resolution. The modeling approach is innovative and the results can serve as an adequate basis for management decisions. All stock assessments, however, represent a compromise between the ideal and the realized. Changes in data quality over time, lack of sufficient spatial resolution in the characterization of removals, significant but variable impacts of disease, observation error in monitoring programs, habitat loss, and trends in ecosystem conditions all influence the oyster assessment. It is the opinion of the Review Panel that this assessment deals with these compromises in a rigorous and scientifically credible way.

Review Process

The review process comprised several conference calls prior to the onsite review, a three-day meeting attended by the Assessment Team which included staff from Maryland DNR and from UMCES, and a follow up writing period by the Review Panel. A complete list of the participants at the meeting and their roles may be found in Appendix 3. The Review Panel's report was submitted to Maryland DNR for review but only factual errors were revised. No changes were made to the opinions or conclusions of the Review Panel. Any factual errors remaining in the report are the responsibility of the Review Panel.

Data Considered

The Assessment Team conducted a thorough review of all primary sources of data for oysters in Maryland. These included both fishery-dependent and fishery-independent data as far back as 1889. Not surprisingly, the utility of these time series for stock assessment and modeling purposes varied over time. After extensive analyses, the stock assessments were based on 36 spatially discrete units based on removals recorded at the level of NOAA Codes. Oysters are harvested from well-known beds, defined over a century ago by Yates. One or more oyster beds occur within the NOAA Code subareas but official landings could not be resolved to a finer scale. Conversely, fishery independent time series of relative abundance which might have been disaggregated to a finer spatial scale, could be combined in a scientifically credible way for consistent measures of trend. Based on these considerations, the assessment period is restricted to 1999 onward.

Modeling Approaches

In contrast to most stock assessments, the natural mortality rate of oysters is both variable and high relative to fishing mortality. Diseases (MSX and Dermo) vary in intensity over time and along salinity gradients within the Bay (Bushek et al., 2012). Consistent long-term monitoring of oyster boxes (i.e., dead oysters whose shells remain hinged) allowed the Assessment Team to independently estimate annual natural mortality rates apart from the stage-based model. Three separate methods were used, allowing for valuable insights into model performance.

Oysters growth varies seasonally and annually, making age determination difficult (Kraeuter et al., 2007). The assessment relies on a novel stage-based population model that also includes the dynamics of habitat. Inclusion of habitat allows for prediction of increases due to shell supplementation programs, but otherwise habitat is assumed to decline based on contemporary analyses of Bay-wide habitat degradation. An equally novel model for determination of fishing mortality reference points is developed. It also models habitat changes but in a conceptually different manner (i.e., density dependent dynamics). In contrast to the population model, the habitat state variable in the reference point model can increase in response to an intrinsic rate of growth as well as habitat supplementation. The Panel expressed some reservations about these differing conceptual bases and the use of stage-based model results as input to the biological reference point model. At the current level of oyster abundance these concerns are considered minor but future assessments should attempt to reconcile these differences. Moreover, reliance on external parameters derived from the literature and use of strong penalty functions in the estimation methods should be reviewed in future assessments.

Restriction of the assessment period to 1999 onward precludes the ability to estimate historical abundance levels, say in the late 1890's. Any such exercise is unlikely to yield precise estimates. Moreover, it can be argued that the environmental and ecological conditions that obtained nearly 150 years ago are unlikely in 2018 onward and are therefore not useful as biomass targets. Despite these limitations and differences in size limits over time, it is relevant to note that the estimates of market oyster abundance of about 300 million market oysters in 2018 is less than 10% of the quantity ***harvested*** annually before 1900.

The Assessment Team used the minimum abundance estimated between 1999 and 2017 as the abundance threshold for each NOAA code. This was based on the assumption that if abundances as low as those observed previously have not so far caused a population crash, they should be sufficient to prevent a crash in the future. This approach is often used in European assessments where the lowest observed abundance provides an estimate of the threshold for recruitment failure. Recruitment failure per se is unlikely in oysters but the

Review Panel agreed that this threshold criterion appropriately balanced the information content of the assessment with a longer term perspective on abundance. The Review Panel concluded that a determination of the carrying capacity of Chesapeake Bay under prevailing environmental conditions (particularly disease prevalence) is beyond the scope of existing data sources and scientific understanding.

Terms of Reference 4 and 5 (see Appendix 2) were particularly challenging and the Assessment Team did an exceptional job of assessing the efficacy of various management policies implemented by the State. Where data allow, the quantitative impacts of these measures are explicitly incorporated into the model's interpretation of habitat changes, exploitation estimation, and reference point determination. The MD DNR has in place a number of long term studies that may ultimately allow for quantification of the utility of these measures and improvements in approaches. Rigorous monitoring will be essential. Well designed and monitored management experiments within NOAA code areas may prove useful for improving management interventions. The Review Panel recommended that sanctuary and habitat plantings, and aquaculture operations should not be considered a part of the standing stock of the fishery, nor part of the reproductive capacity of the fishery. Doing so will overestimate the spawning potential, and the contributions of sanctuaries, habitat plantings and aquaculture are as yet unclear and likely vary greatly by source.

Research Recommendations

The Review Panel endorsed the recommendations of the Assessment Team. In addition, the Panel's recommendations include:

1. Implement an annual dockside monitoring program to establish the number of small oysters being caught per bushel, and to estimate the size and number of oysters per bushel for each NOAA code over the course of a fishing season.
2. Conduct experiments to estimate of dredge efficiency for the survey.
3. Compare existing independent experimental estimates by gear type of abundance within some NOAA codes
4. Consider re-running the Bayesian model using data from the patent tong survey in key areas if and when sufficient data are available. Results could be compared to the existing estimates of mortality from the Bayesian model that uses the fall dredge survey.
5. Conduct a detailed examination of trends from survey based disease incidence and rates of natural mortality. Any evidence of disease resistance or changes in virulence should be thoroughly examined.
6. Examine potential retrospective patterns in terminal year estimates of biomass and fishing mortality to address uncertainty concerns for management.
7. Review the performance of the assessment and reference point models by examining likelihood profiles for key parameters and the influence of penalty functions on parameter estimates. Further simulation testing would be valuable.

8. Improve the habitat dynamics model, possibly allowing for regeneration of habitat through population growth and replenishment of shells through natural mortality of live oysters.
9. Develop an assessment model with the capability of estimating the reference point parameters internally.

In summary, the Review Panel commends the Assessment Team for a job well done. The assessment is an important step forward for improving the management of oysters in Maryland. Further work on improving data, enhancing monitoring, conducting experiments and model structure will improve our understanding of oyster dynamics.

Section 2.0 Introduction

Call for Stock Assessment

Nearly 150 years ago the Chesapeake Bay was the world's leading producer of oysters. Overexploitation, changes in water quality and the introduction of diseases led to a century long decline with recent landings at only about 3% of historic peaks (Rothschild et al., 1994; Kennedy et al., 2011). The decline in the resource has led to initiation of several long-term monitoring studies, a variety of intensive efforts to rehabilitate the resource (Kennedy et al. 2011), and many concerns about the health of the Chesapeake Bay ecosystem. In 2016 the Maryland General Assembly passed the Sustainable Oyster Population and Fishery Act which in turn led to a request for a formal assessment of the Maryland oyster stock. The request included explicit terms of reference (Appendix 2) and a review by an external independent panel of experts (this report).

The stock assessment was completed and a report was prepared by a stock assessment team consisting of shellfish scientists, statisticians and stock assessment experts from Maryland DNR and the University of Maryland Center for Environmental Science (UMCES). A complete list of participants in the stock assessment project is provided in Appendix 3. The Assessment Team's report MD Department of Natural Resources, 2018represents the first formal assessment of the Maryland oyster population. As such it provides the quantification of current stock size, rates of exploitation and target biological reference points. Current **stock sizes** are estimated to be less than 10% of the peak historic landings, suggesting that the decline in landings mirrors a comparably precipitous decline in overall abundance.

This report provides a record of the review conducted August 27 to 29, 2018 in Annapolis, MD. The Review Panel considered a wealth of written material and presentations made by the Assessment Team. The Assessment Team responded to our written and oral questions and provided additional analyses when requested. We have attempted to chronicle the sequence of events leading up to the assessment, the materials reviewed, the details of the discussions during the meeting, and our primary findings and recommendations. This report has been reviewed by the Assessment Team for factual errors but opinions and conclusions of the Review Panel have not been altered.

Section 3.0 Background Material

Planning for the assessment review began on June 12, 2018 with a conference call between the Assessment Team (MD DNR and UMCES), and Atlantic States Marine Fisheries Commission (ASMFC) staff, and the Review Panel. Lynn Fegley provided an overview of the request for review and gave a general overview of expectations and deliverables. Three weeks prior to the review meeting the Review Panel received the draft report from Assessment Team on August 6. The computer code for the assessment model, data files and various output files were shared with the Review Panel on August 10. On August 14, another conference call for all parties was held to discuss final preparations and to request initial feedback from reviewers to the Assessment Team. Additional supporting documentation was provided to the Review Panel on August 14 and 15. Review Panel members sent their written questions to the Assessment Team between August 15 and 24. A consolidated list of questions from the panel is presented in Appendix 5. These helped the Assessment Team to focus on key concerns of the Review Panel.

The meetings were held in Annapolis MD at the offices of the EPA on Monday (8/27) and Tuesday (8/28) and at the MD DNR offices on Wednesday (8/29). Support from the Assessment Team was excellent providing expertise in oyster biology, survey operations, and fisheries. The Assessment Team was well prepared and gave detailed effective summaries of the data and models. Moreover, they rapidly responded to detailed questions about the report and to conduct additional analyses requested by the Panel. All decisions about the data quality and model structure were open and transparent.

The facilities were highly conducive to a thorough review of the materials, and there were no glitches in computer service or presentations. Onsite refreshments and lunches allowed the Team to effectively work without interruption. Special thanks are given to Laurinda Serafin for her comprehensive service as rapporteur. Finally, the Panel recognizes the exceptional work of Lynn Fegley for her organizational skills before, during and after the assessment review.

Section 4.0 Review of Activities

Proceedings Day 1

The meeting convened at 9:00 am on August 27 with opening remarks by Lynn Fegley of the MD DNR and introductions of the Assessment Team, and the external review panel. The agenda listed in Appendix 1 was modified slightly to allow flexibility in time allotments for various topics. Paul Rago gave a brief overview of the importance of the peer review in stock assessments and an outline of the how presentations and discussions would proceed. Rapporteurs from the MD DNR were Laurinda Serafin, Lynn Fegley, Amy Larimer, and Rachel Pierce.

The meeting was opened with a thorough summary of the data used in the assessment. The primary fishery-dependent data sources included Dealer Buy tickets, Harvester Reports and bushel tax receipts. The Assessment Team applied rigorous data quality standards to ensure that these data could be consistently interpreted over the assessment time series. Comparison of landings by dealer buy tickets and harvester reports in 2009 and 2010 suggested dealer records were underestimated by 10%; overall landings were adjusted by this ratio for the period 1999-2017. When asked about the sensitivity of the overall assessment to this underestimation correction, the Assessment Team noted that sensitivity analyses were performed testing higher and lower corrections, neither of which altered the trends in the assessment greatly. The Review Panel noted the value in this sensitivity analysis and requested that the results be conveyed within the report. The historical records for landings were important determinants of the spatial resolution of the assessment. Although landings are generally taken from known oyster beds, the historical landings are only recorded at the resolution of NOAA Codes. These subareas comprise one or more oyster beds and have varying productivity depending on their salinity and temperature regimes.

Oysters are harvested by multiple gears which vary in magnitude of landings both regionally and temporally. A total of 21 port samples were taken from catch sourced from several different NOAA code and gears. The primary objective of the port sampling was to estimate the number of oysters harvested per bushel as a basis for converting historical landings recorded in bushels to numbers of oysters. The relatively small sample size was probably insufficient to attribute differences in bushel counts to area or gear but they probably reflect the overall range of possible values. The samples were taken at one dealer location on a single event, yet encompassed a wide range of values of oysters per bushel. It was noted that variations observed by area and gear may also occur over time. Based on the port samples, the Assessment Team used the overall average of 227 oysters per bushel. The Assessment Team suggested there was little evidence of high grading in oyster landings, owing to the absence of a price differential for small and market oysters.

Commercial effort was measured in terms of number of trips, number of licenses per trip, and number of hours fished per trip. Multiple licenses can be used on a single trip. The number of hours fished was found to be an unreliable measure of effort because trips are not often reported to the hour, so relative abundance from commercial landings was measured as catch per license per trip.

During the course of the fishing season, CPUE often declined sharply, thereby allowing application of standard Leslie-Davis depletion models. Standard regression methods were modified to account for the truncation of catch rates when trip limits were attained. This occurred early in the season when oysters were more abundant. Simulation studies, reported but not reviewed by the Panel, suggested that the censored regression models were relatively unbiased. Owing to the typical magnitude of within season declines and discussions with harvesters, the Assessment Team felt confident that the declines in CPUE were attributable to true changes in abundance and not regulations or fishing behaviors. Nonetheless about 15% of the depletion estimates had positive slopes, possibly attributable to changes in catchability over the season.

Depletion analyses were used primarily to estimate exploitation rates which were defined as the ratio of total catch over the initial population size. Where two or more gears were used in a NOAA code area, the estimates were summed if the fishing activity was known to be spatially distinct. Otherwise the gear with the most number of years and valid estimates were used for a particular area. A paired t-test of differences exploitation rate estimates between gears was reported but deemed inconclusive since none of estimates could be validated.

The discussion continued with a description of the fishery-independent data. A fall dredge survey using fixed stations began in the 1930s but sampling methodology was not consistent overtime. The most reliable and consistent survey data began in the late 1990s and tow distance was standardized in the mid 2000's. The standard survey sampled an average of 261 bars with an average of 347 samples per year. To account for missing samples over time, relative density estimates were obtained by using a mixed effects GLM model with a negative binomial distribution. In general, the GLM predictions differed little from the raw means.

A patent tong survey began in 2010 in sanctuary areas of the bay. This survey uses a stratified random design based on substrate type. These data were used to inform the assessment in some NOAA code areas.

Discussion moved to a description of the stage-specific population model used. This model includes three size classes of oysters, spat, smalls (submarket adults) and markets (oysters greater than 3"). Oyster age and growth is difficult to determine and few datasets exists. Transition probabilities among size classes were obtained from a Von Bertalanffy growth function collected by the Paynter Lab who measured size at age from oysters reared in a hatchery and planted at a known time/location in sanctuaries and managed reserve areas. The model assumes 15% survival of 2mm spat, and that all spat transition to smalls by the subsequent year, and about 60% of smalls transition to markets in the subsequent year. The transition rate is estimated in the model but is constrained by a penalty function in the likelihood function. The Review Panel notes that these transition probabilities are consistent with those listed in Rothschild et al (1994) for the Chesapeake, and Munroe et al (2017) for the Delaware Bay.

The stage-specific model assumes 15% annual survival of 2mm spat to the "small" oyster stage. This assumption is based on data from hatchery plantings of spat-on-shell at many restoration sites over several years. This estimate of spat survival may be an overestimate for wild spat, and for planted spat in general as Harris Creek is a low salinity region where predation losses and other mortality factors may be low. One natural mortality rate is applied to small oysters, transplanted seed and market oysters in the stage-specific model, regardless of location within the bay, meaning no spatial pattern in mortality is applied across NOAA codes.

An exponential decay of habitat is applied to the stage-specific model. This habitat loss function is based on data provided in Rothschild et al. (1994), and assumes the loss estimated previously continues through today. The rate of habitat loss is estimated in the model at approximately 4% per year from 1980 onwards, but this value is highly informed by a penalty function in the likelihood function. Habitat within a NOAA code is credited in the model when shell or spat-on-shell is planted in that region.

Maryland law requires three reference points in its fishery management plans. These include abundance and exploitation rate limits, as well as a target exploitation rate. The Assessment Team introduced a model designed to estimate the exploitation reference points for each NOAA code in the oyster fishery.

The reference point model was a modified from the production model developed in Wilberg et al. (2013). The model calculated a carrying capacity that was based on the amount of available habitat. Habitat is increased by production from living oysters, planted oysters (spat-on-shell, or transplanted seed from other areas) and other added substrate. Habitat was decreased by habitat loss, which occurred at a constant rate. The model estimated an intrinsic rate of population increase and habitat production. These were used to generate estimates of the exploitation rate expected to generate maximum sustainable yield and the maximum limit on sustainable exploitation rate.

The Assessment Team used the minimum abundance estimated between 1999 and 2017 as the abundance threshold for each NOAA code. This was based on the assumption that if abundances as low as those observed previously have not so far caused a population crash, they should be sufficient to prevent a crash in the future. The Review Panel noted that the use of 2017 as a terminal year for this time window may be problematic given that 2017 is the year being assessed.

The Panel noted that these reference points represent a substantial advancement in the oyster management for Maryland, where no previous reference points existed. The Panel pointed out that, although it is common practice in fisheries, the separation of the stock assessment model and reference point model is not ideal and can introduce some potential statistical problems. For example, the reference point model uses the outputs of abundance and habitat from the stock assessment model as “data”, which can be problematic (Brooks and Deroba 2015). The Panel recommended that in future assessments, the stock assessment and reference point models should be combined, although the Panel recognized that the available data does not currently support such a model.

The first day’s review finished with several recommendations from the Panel to the Assessment Team to provide an overview of the general principles governing the model development, and some insights into the path of model development. The Panel also requested a presentation of the Bayesian mortality model.

Proceedings Day 2

Per the Reviewers' request, Mike Wilberg opened with a summary of the basic tenets governing the stock assessment model development. The Assessment Team tried to use as much relevant data as possible, including the dredge survey, patent tong survey, estimates of stage based natural mortality rates, and matching of fishery dependent estimates of fishing mortality. Generalized models were applied to 36 separate spatial units with varying degrees of underlying productivity and fishery characteristics. Despite the availability of high resolution information on the locations of oyster beds and knowledge of current and historical management practices the finest resolution possible was the NOAA Code area because of reporting characteristics of the fishery.

The stage-based model used in the assessment was based on earlier work by Wilberg et al. 2011. Additional model features were included to incorporate box count mortality rates, indices of natural mortality, and survey data from relatively unbiased patent tong surveys in some areas. Wilberg provided additional information on how habitat was modeled in the assessment model and its utility in evaluating the effects of management interventions and hatchery supplements to the stock.

The Review Panel noted that the dredge survey improvements since 2005 may ultimately lead to improved use of the fall survey as a measure of swept area abundance. Owing to historical survey practices and recording procedures, the long-term data cannot be used for direct abundance estimation.

Kathryn Doering presented her Bayesian model for estimation of natural mortality. The model incorporates information from the dredge survey on the numbers of live and recently dead (box) oysters in both the small and market size groups. The model allows for use of external information from experiments, and addresses the important time series information for box dynamics.

Dead oysters do not remain as 'boxes' permanently, and the two valves that make the box will disarticulate from one another over time. This rate of disarticulation is not well defined in general but is likely to vary widely with biological and environmental characteristics, and little data from the Chesapeake exists upon which to constrain model parameters. Uncertainty information from the model suggests that durations of boxes estimated from experiments may not be representative of durations observed in fishery operations. Overall the Assessment Team and the Review Panel felt that incorporation of additional information into the natural mortality estimator represented an important advance.

At the Panel's request the Assessment Team reviewed some of the stage model results in more detail focusing on model runs with poor model performance. It was expected that such results would provide insights into model performance when data conflicts were present. For the most part, the NOAA codes in which poor model performance occurred (#5, 82, 129 and 331) were ones where fishing is no longer occurring and data are sparse or non-existent. In general, the

Panel felt these additional investigations demonstrated reliable model performance over a range of data quality conditions.

The definition of overfishing depends on a parameter, *d*, the intrinsic rate of habitat loss, which is based on Wilberg et al. (2013) and is near the value estimated in Powell and Klink (2007). In many cases, this parameter drives the intrinsic rate of population increase (*r)* to its lower limit (zero). As the *d* parameter is fixed rather than estimated, the Panel requested additional sensitivity runs of the model to estimate the effects of a 50% reduction and 100% increase in that parameter.

The Panel considered the Assessment Team's progress in answering TOR 4 and 5 at length. Both of these TOR were considered difficult to answer because of data limitations and uncertainty about key biological processes of reproduction, settlement, post stock growth and mortality, and hydrological processes. For example, larval behavior in the water column interacts with tidal transport and salinity gradients in ways that are poorly understood and difficult to measure. Consequently, quantification of how hatchery plantings affect spawning potential is challenging.

The day concluded with a review of the questions submitted by the Panel in writing prior to the meeting (Appendix 5). In particular, it was important to determine that all questions had been satisfactorily answered. The Panel agreed that all of the questions had been addressed. A number of recommendations regarding the content and formatting of the Assessment Report were made by the Review Panel.

Proceedings Day 3

The Panel met in closed session at the MD DNR offices in Annapolis from 9:00 a.m. to noon to write initial conclusions. At noon, the Assessment Team and MD DNR staff met with the Review Team to summarize the initial conclusions of the Review Panel for each Term of Reference. The meeting adjourned about 1:30 pm with some closing remarks by the Chair on a job well done by the Assessment Team and a review of the timetable for preparing a final report.

Section 5.0 Summary of Terms of Reference

1) Complete a thorough data review: survey data, reported harvest and effort data, studies and data related to population rates (growth, mortality and recruitment), available substrate, shell budgets, and sources of mortality.

   Panel Conclusion ➔ *The Assessment Team met this TOR.* Overall, the Assessment Team did an outstanding job of reviewing the existing data sources. Their thoughtful and scholarly reviews paid close attention to the data collections procedures and how they may have changed over time. Each change was considered with respect to its implications for deriving consistent measures of scale and trend. The uncertainty of observations was also evaluated where possible. Collectively these approaches led to a focus on more recent data (i.e., 1999 onward) rather than the very long time series of catch records. While this approach restricts inferences to a more contemporary history, it preludes inferences that may be less reliable and driven by sharp contrasts in landings and strong, but ultimately unverifiable assumptions about historical productivity. The Review Panel agreed with the decisions of the Assessment Team.

The Review Panel did note that relatively recent time series on disease trend data are provided in the report at a Bay-wide resolution. In some cases, the fall dredge survey collects these data at NOAA code level. It may be useful to examine disease status and trends at NOAA code level in the future.

a) List, review, and evaluate the strengths and weaknesses of all available data sources for completeness and utility for stock assessment analysis, including current and historical fishery-dependent and fishery-independent data.

Panel Conclusions➔ *The Assessment Team met this TOR.*

- As noted above, the Assessment Team did an exemplary job of reviewing the available data sources. Appendix I of their report contains a thorough review of the data sources available for fishery-dependent and fishery-independent data. For each data source, strengths and weaknesses of the data, utility for the assessment and important changes in how data were collected (if any) through each time series are listed.
- The fall MD DNR dredge surveys for abundance, mortality and disease provide important data for the assessment. The MD DNR patent tong survey in the spring also provides useful data including shell heights.
- The fishery-dependent data includes dealer buy tickets, harvester reports and bushel tax receipts. These three time series are generally coherent but differ in terms of the spatial resolution of each. Those strengths and weaknesses are well laid out in Appendix I, and described in the report.
- Appendix II of their report has details time series of oyster seed (both hatchery and wild spat) and shell plantings by source
    - For some plantings, in particular those being done for restoration, other sources may provide additional information content that could be explored. This may be the case if these plantings are being monitored by other groups apart from DNR; however, we note that data sources outside of DNR may have sampling error that may be problematic so should be reviewed cautiously.

b) Identify the relevant spatial and temporal application of data sources.

Panel Conclusions➔ *The Assessment Team met this TOR.* All stock assessments are ultimately limited by the availability of data collected consistently over time and at a relevant spatial scale. This principle is especially relevant to sessile species. For this assessment, the limiting factor was the recording of removals as the NOAA Code level of spatial resolution. Finer scale information at the resolution of oyster beds would be desirable but is not possible. The Review Panel does not endorse imputation methods that might result in finer spatial resolution but would definitely increase the overall uncertainty and ultimately, the utility of the assessment for management.

- Appendix I contains a review of the data sources available for fishery-dependent and fishery-independent data. For each data source, the spatial and temporal extent of coverage are listed.

c) Document changes in data collection protocols and data quality over time.

Panel Conclusions➔ *The Assessment Team met this TOR.*

- An excellent summary of changes in survey design and methodology is provided in Appendix I.
- The Assessment Team provided detailed information about differences among surveys and their objectives which provide important context to these data sources and how they can be used.
- In the case of the fall dredge survey, an important change in protocol is missing from Appendix I. The survey prior to 2005 did not record swept area, likewise records of overfull dredge hauls show approximately 20% of tows overfilled the dredge. Since 2005, swept area is being tracked and therefore the time series since 2005 may be useful for swept area calculations and abundance estimates.

d) Justify inclusion or elimination of each data source.

Panel Conclusions➔ *The Assessment Team met this TOR.*

- The Panel had minor concerns regarding some decisions for exclusion, e.g. 50 samples per year in CPUE, or the number of years of data when multiple gears present in depletion analyses.
- Swept area estimation from 2005 onward might have been used but the time series was not long enough. Comparisons with the patent tong survey might be useful as patent tong samples index absolute abundance on the bottom. Some caveats and concerns regarding the patent tong survey is the design in terms of sampling repeat sentinel stations versus a random sample design.

2) Develop stock assessment model or index based approach that estimates biological reference points and document status of the stock relative to estimated reference points. To the extent possible, quantify sources of uncertainty within model.

Panel Conclusions ➔ *The Assessment Team met this TOR.*

- The Panel endorsed the reference point model and stock assessment model as useful for estimating biological reference points and determining the status of the stock.

- The reference point model developed by the stock Assessment Team incorporated habitat, which is unusual in fishery stock assessment, but necessary for oysters in Maryland. The Review Panel noted the innovative nature of the reference point model and endorsed it for management, but noted some potential improvements for future stock assessments, particularly with regard to the use of habitat.

- The stock Assessment Team developed a custom stock assessment model that leveraged the available data on oysters in Maryland. Most modern stock assessments are implemented using existing model frameworks, but that was not an option in this case. Oyster stock dynamics, including the dependence on existing habitat for successful recruitment, as well as the peculiarities of the fishery and the data, required a model specifically tailored to Maryland oysters. The Review Panel found this approach to be innovative and well adapted to the available data.

- The separation of the reference point model and the stock assessment model was necessary due to limitations of the data, but introduced some concerns for the Review Panel. These included the use of output from the stage-structured model as data in the reference point model. This can be problematic because there is uncertainty, typically some degree of autocorrelation, and potentially retrospective bias in model output, that may not be accounted for. Incorporating the reference point model into the stage-structured stock assessment model, perhaps implementing a cultch dynamics component, would be an improvement, but would require additional data.

- The Panel expressed some reservations regarding the modeling of habitat in the stock assessment model. The model uses an exponential decay function to model habitat, which results in a predisposition to decline. Likewise, the model does not include dead oysters nor cultch as habitat, although these are measured and estimated elsewhere. The panel noted that while this is currently appropriate given limitations in the data, it may not sufficiently represent future conditions if the stock abundance were to grow to the point that live, dead, and cultch creates habitat that is self-sustaining.

- The stock Assessment Team applied a general model framework across each of the 36 NOAA code regions in the Maryland oyster fishery, rather than tuning individual models to each region. The Review Panel noted that this approach is preferred because it reduces the probability of undetected aberrant model behavior. Tuning the model to individual NOAA codes could produce instabilities that would be inefficient to diagnose. In addition, tuning to each NOAA code would likely require unique modeling choices for each region, which would have to be justified and documented. These alternative approaches would have complicated the stock assessment and review considerably.

- The stock Assessment Team chose to use global penalty functions to constrain the estimation of model parameters, rather than directed tuning in problematic regions.

While the Review Panel agrees that this was the better approach, the implications of relaxing the penalty constraints were not fully explored during the review and may bear further investigation in the future.

- The Panel had some difficulty interpreting the various q (typically referred to as catchability) parameters in the stock assessment model. The dimensionality of these parameters might be determined with some additional work, which would help with interpretation and might lead to some useful diagnostic tools, such as describing a reasonable range for each q parameter given the data.
- Among the few tuning choices made by the stock Assessment Team were the imposition of fixed effective sample sizes on the fishery dependent depletion time series and the mortality time series. The fishery depletion data were given a relatively low weight such that the model was not forced to precisely match them. This decision was justified by evidence that the depletion data were probably less precise than they appeared. The natural mortality time series was given a relatively high weight, such that the stock assessment model was forced to fit them better than it otherwise would have. This decision was justified by comparison between the mortality index and an external Bayesian model designed to estimate natural mortality in Maryland oysters. The Panel was not shown sensitivity runs demonstrating the effect of alternative weighting decisions. These should be explored in the future.

3) Compare estimates of stock status generated by index and model-based approaches. Justify selected approach.

Panel Conclusions➔ *The Assessment Team met this TOR*.

- The Panel notes that regardless of the method used, the conclusions regarding stock status in most NOAA codes are clear. Abundances in many NOAA codes are near time series low values and harvest rates are likely above long term sustainable levels.
- Index methods for estimating natural mortality indicate lower natural mortality than the stock assessment model in some NOAA codes. This may result from mortality due to harvest occurring before the fall dredge survey in these areas (the model assumes that harvest occurs after the survey). This theory was not analytically explored however, and the Panel could not verify the root cause of the discrepancy. The Panel notes that the Assessment Team did attempt to align the model estimates with the index-based estimates of natural mortality by increasing the relative importance of the natural mortality index in the estimation of model parameters.
- An index based abundance trend was presented in Appendix IV, and used to support the choice of spatial scale in the assessment. The index-based trends were however, not directly compared to model estimates of abundance. A direct comparison between

these components of the assessment report might be useful in evaluating this TOR further in future assessments.

- The fishery depletion time series was down-weighted in the estimation of model parameters. This contributed to a misfit between the (index based) depletion estimates of harvest rate and the model estimates of harvest rate in some NOAA codes. In other NOAA codes the fit was quite close. The Panel noted that the decision to down-weight the depletion index, and thus sacrifice fit to the data, was justified by inconsistencies in the data.

4) Include sanctuaries and restoration efforts in sanctuaries in the development of stock assessment approaches.

Panel Conclusions➔ *The Assessment Team partially met this TOR.*

- Sanctuaries are difficult to evaluate since many of the key biological processes are poorly understood. The approaches used by the Assessment Team are innovative and appropriately utilize available data. The Review Panel did not recommend specific alternative approaches.

- In an ideal data world it would be appropriate to partition removals between sanctuary and open areas within NOAA codes, and model consistently within those boundaries. However, data limitations preclude such an approach within the NOAA code areas. Hence the dynamics of oysters within areas that contain sanctuaries represent a mixture of habitats that may be improving due to hatchery plantings, substrate improvements or protection from harvest, and habitats that are subject to exploitation. Strictly speaking, this compromises the utility of the assessment model estimates of exploitation because part of the population is not subject to harvest.

- The Panel notes that sanctuaries are often implemented for the purpose of establishing subpopulations of oysters that are intentionally excluded from fishery access. Because of this distinction, we question the utility of including sanctuaries in the calculation of assessment metrics as they are, by definition, not part of the fishery. The one practical issue with parsing sanctuaries that were traditionally part of the fishery is the problem of retrospectively adjusting the fishery-dependent and fishery-independent time series. For new sanctuaries, that were never part of a fishery, this should be an easier distinction.

5) Examine how hatchery plantings (aquaculture and public fishery) impact spawning potential in the fishery.

Panel Conclusions➔ *The Assessment Team met this TOR to the extent possible. This was an exceptionally difficult problem for reasons described by the Assessment Team. In general it is difficult to track fate of hatchery plantings and several different life stages stocked. A variety of methods were used to stock hatchery plantings which further* *complicates the analyses. For these reasons the Assessment Team heavily qualified their conclusions. The Panel concurs with Assessment Team's concerns.*

- The Panel had difficulty with wording of the TOR, in particular, it is unclear how hatchery-reared oysters can contribute directly to the spawning potential of the animals that are part of the fishery. We limit our discussion here to the 'spawning potential in the fishery' and make a distinction below between the role that may be played by aquaculture oysters (those sourced from hatcheries and placed on privately owned farms) versus the role played by hatchery sourced oysters that are used to supplement the wild fishery.

- Aquaculture oysters are livestock (or shellstock in this case) and not part of the public fishery. They should have no bearing on the spawning potential of the oysters in the public fishery. There is no mechanism that we are aware of, by which the presence of a farm, and livestock therein, will increase the fecundity or fertilization success of the oysters in the public fishery.

- In terms of hatchery plantings on public fishing grounds, these spat-on-shell will contribute to future conditions, assuming those plantings survive to reach maturity. These data are tracked in both the habitat and live oyster portions of the model and in the fall dredge survey, so as these animals grow into size classes that can spawn, their contributions would/could be accounted for then along with the other wild set oysters in the fishery.

- In general, the Review Panel believes that larval contributions from hatchery-sourced oysters, whether in sanctuaries or fished grounds, and from farms should not be considered part of the spawning potential for the fishery. These sources cannot be verified as regular and reliable sources of larvae and accounting for them will only overestimate the capacity of the stock Bay-wide.

- Despite the challenges, the Assessment Team made some accounting for standing stock in both aquaculture and hatchery sourced contributions to the fishery and compared those to the standing stock in the fishery. The Panel found the assumptions made in these calculations to be acceptable given the circumstances, and appropriate caveats were made in the report.

SECTION 6.0 SUMMARY CONCLUSIONS

The Review Panel endorses the methodologies used by the Assessment Team to assess the Maryland oyster stock. The modeling approach addresses the essential features of oyster biology and historical data collection procedures. Our endorsement hinges on several unique aspects of the assessment, including:

- Evaluation of existing data sources
    - The Assessment Team conducted a thorough review of all existing fishery independent data for evaluation of trends.

  - Historical information on landings were evaluated with respect to changes in reporting practices over time and spatial resolution.
  - Data sources were integrated into overall assessment where possible. When such integration was not possible, results of analyses apart from the assessment model were compared with model results.
- Biological Processes
  - Atlantic estuarine oyster stocks are strongly influenced by the presence of two lethal diseases, MSX and Dermo. These diseases vary in intensity with both temperature and salinity.
  - Oyster growth and mortality due to predation likewise varies with temperature and salinity.
  - Environmental gradients in the estuary imply that the dynamics of the resource will vary spatially.
  - The result of variable natural mortality can be tracked by monitoring of empty, but articulated shells (known as boxes). Newly articulated boxes (those without fouling) can provide an estimate of recent mortality, whereas old boxes (those still articulated but covered with fouling organisms) are a less reliable estimate of mortality because of uncertainty and variability in the time it takes for a given box to disarticulate.
- Removals
  - The Yates study from nearly a century ago provides a rigorous quantitative description of historical benthic habitats and a basis for defining the desired level of resolution for removals. Unfortunately, data on removals by bar do not exist. The analysts have appropriately used the existing data at the resolution of NOAA code area.
  - The analysts have restricted the assessment time series to a period where data quality issues are minimized.
  - Concerns about the use of commercial CPUE data are well founded since it is difficult to derive a meaningful measure of effort that can be used across all assessment areas and over all time periods.
- Monitoring
  - The assessment benefits from a long time series of fishery-independent data monitoring studies that allow tracking of relative abundance.
  - As these methods have changed and improved over time, the Team has used appropriate measures to restrict the data to a period where consistent inferences are possible.
- Assessment Model
  - The assessment model addresses the key biological processes and removals in a realistic way.

  - The model explicitly accounts for the role of spatial and temporal variation in natural mortality.
  - The stage based model is consistent with the historical data on landings and fishery independent survey monitoring.
  - Assessment model results are compared with index models.
  - The spatial units are all assessed under a consistent but flexible modeling framework. This allows for rapid analyses of overall stock condition while accounting for spatial and temporal variation. While it may be argued that models for individual NOAA codes might be improved with detailed tuning, the Panel feels that this could ultimately lead to overfitting and inconsistencies among spatial units.
- Biological Reference Points
  - A study from nearly a century ago provides a rigorous quantitative description of historical habitats and a basis for potential rebuilding. Any rebuilding will require habitat enhancement, and would benefit from biological shifts such as relaxation of natural disease mortality rates, improved recruitment and continued improvement to water quality.
  - The rationale for excluding biomass rebuilding targets is justified because the known peak abundances likely occurred more than 150 years ago when Chesapeake Bay was a very different ecosystem and diseases were not a dominant factor in the oyster life history.
  - The biological reference point for exploitation appears to be a useful starting point for characterizing the relative magnitude of contemporary fishing mortality. Future modeling refinements may improve this but we agree with the Assessment Team that substantial improvements are not possible in the short term.
  - Parameters that are assumed constant in the current model should be tested regularly and updated as appropriate. In particular, parameters that imply habitat declines consistently over time (in both the assessment and BRP models) with only limited biological contributions to habitat growth (such as the inclusion of dead oyster shell in habitat capacity) should be updated as new information becomes available.
- Effects of sanctuaries, habitat augmentation, and hatchery plantings
  - The Assessment Team did an exceptional job of assessing the efficacy of various management policies implemented by the State.
  - Where data allow, the quantitative impacts of these measures are explicitly incorporated into the model's interpretation of habitat changes, exploitation estimation, and reference point determination.
  - The MD DNR has in place a number of long term studies that may ultimately allow for quantification of the utility of these measures and improvements in

approaches. Rigorous monitoring will be essential. Well designed and monitored management experiments within NOAA code areas may prove useful for improving management interventions.

- Sanctuary and habitat plantings, and aquaculture operations should not be considered a part of the standing stock of the fishery, nor part of the reproductive capacity of the fishery. Doing so will overestimate the spawning potential, and the contributions of sanctuaries, habitat plantings and aquaculture are as yet unclear and likely vary greatly by source.

SECTION 7.0 RESEARCH RECOMMENDATIONS

The Review Panel endorsed the recommendations of the Assessment Team (Appendix 4). In addition, the Panel's recommendations include:

1. An annual dockside monitoring program is recommended to establish the number of small oysters being caught per bushel, and to estimate the size and number of oysters per bushel for each NOAA code over the course of a fishing season. Samples should be collected randomly of the catch coming from a range of bars, NOAA codes, gear types and time of the fishing season as all of these factors may generate differences in the catch composition.
2. The stage based model assumes that recruitment is independent of stock size, whereas the BRP model assumes recruitment is proportional to stock size as a time invariant scalar and reduced by the ratio of abundance to a time varying carrying capacity. Developing an assessment model with the capability of estimating the reference point parameters internally is desirable. The Panel is well aware that this is a significant challenge, unfulfilled in many stock assessments worldwide.
3. Experiments should be performed to estimate of dredge efficiency for the survey. This may be done with coordination of patent tong and dredge surveys at known locations. These data could help in transforming the fall dredge survey data to swept area abundance estimates.
4. In some NOAA codes, relative oyster abundance is estimated independently for more than one gear type. It may be informative to investigate how often do those estimates disagree, and by how much? These comparisons may help in understanding reliability of the estimates that are being generated from each gear type.
5. The BRP model implicitly assumes that Dermo and MSX are here to stay and that future dynamics will be dominated by them. Evidence of trends in these diseases could be important for reference points. Detailed examination of trends from survey based disease incidence and rates of natural mortality should be conducted. Any evidence of disease resistance or changes in virulence should be thoroughly examined.
6. Many assessment models exhibit patterns of over or under estimation of biomass and fishing mortality rates in the terminal year. Reasons for this bias are not

completely known but are often attributed to changes in an underlying rate that are not accounted for in the model. This tendency can often be detected by examining the pattern of terminal year estimates as the time series is progressively shortened one year at a time and by comparing those estimates with the estimates from the entire time series. An analysis similar to this would be useful for oysters and could reveal whether further precaution was warranted when utilizing exploitation rates for evaluation of management measures.

7. The recommended reference points may be subject to a shifting baseline bias. In particular, the abundance reference points are predicated on the assumption that time series minima have not previously induced population crashes within a NOAA code, and should therefore be sufficient to maintain abundance over time in the future. This assumption is somewhat problematic for two reasons. The first reason is that some of the NOAA codes reached time series minima in the last or near to last years of their respective time series. The resilience of populations at these abundances has not been demonstrated by continued existence over time. The second issue is a logical circularity induced by the separation of the exploitation rate model and the stock assessment model. The reference point model for exploitation reference points describes abundance as a function of available habitat. The stock assessment model describes habitat as an exponential decay function, such that habitat this year is greater than habitat next year (assuming artificial habitat additions do not swamp the decay rate). Therefore, by inference between the reference point model for exploitation and the stock assessment model for population abundance, the resilience of a NOAA code to historical abundance minima will not be the same as the resilience of that NOAA code in the future, because the habitat in the future will be less than it was. The Panel recommends that merging the stock assessment and reference point models and internally estimating an abundance threshold based directly on model results would be preferable.
8. The assessment model introduces many technical innovations, appropriately considers the information content of existing data sets, and incorporates many external sources of information for model parameterization. The resulting penalized likelihood function is complicated and may induce unexpected variations in model parameters. Further simulation testing of model performance and application of likelihood profile analyses to examine model performance in the vicinity of the optimal values is desirable.
9. Improve the habitat dynamics model, possibly allowing for regeneration of habitat through population growth and replenishment of shells through natural mortality of live oysters.

1. The Bayesian model could be re-run using data from the patent tong survey in key areas if and when sufficient data are available. Doing this would eliminate the need for efficiency correction in that model because patent tongs are assumed 100% efficient. The patent tong survey also has sentinel sites surveyed over time. Results

of this could be compared to the existing estimates of mortality from the Bayesian model that uses the fall dredge survey.

SECTION 8.0 DOCUMENTATION

Referenced Citations

Brooks, Elizabeth N., and Jonathan J. Deroba. When "data" are not data: the pitfalls of post hoc analyses that use stock assessment model output. *Canadian Journal of Fisheries and Aquatic Sciences.* 72(4): 634-641.

Bushek, D., S.E. Ford and I. Burt. 2012. Long-term patterns of an estuarine pathogen along a salinity gradient. J. Mar. Res. 70:225-251. DOI: 10.1357/002224012802851968

Kennedy, V.S., Breitburg, D.L., Christman, M.C., Luckenbach, M.W., Paynter, K., Kramer, J., Sellner, K.G., Dew-Baxter, J., Keller, C. and Mann, R., 2011. Lessons learned from efforts to restore oyster populations in Maryland and Virginia, 1990 to 2007. Journal of Shellfish Research, *30*(3): 719-731.

Kraeuter, J.N., S.E. Ford and M.J. Cummings 2007. Growth measurement in oysters: A comparison of methods. Journal of Shellfish Research, 26:479-491.

Munroe, D**.,** S. Borsetti, K. Ashton-Alcox, and D. Bushek. 2017. Early post-settlement growth in wild Eastern oyster (*Crassostrea virginica* Gemlin 1791) populations. *Estuaries and Coasts*. 40(3): 880-888. DOI: 10.1007/s12237-016-0185-y

Powell, Eric N., and John M. Klinck. Is oyster shell a sustainable estuarine resource? *Journal of Shellfish Research.* 26(1): 181-194.

Rothschild, B.J., Ault, J.S., Goulletquer, P., Heral, M. 1994. Decline of the Chesapeake Bay oyster population: a centure of habitat destruction and overfishing. *Marine Ecology Progress Series.* 111: 29-39.

Wilberg, M. J., Livings, M. E., Barkman, J. S., Morris, B. T., & Robinson, J. M. (2011). Overfishing, disease, habitat loss, and potential extirpation of oysters in upper Chesapeake Bay. *Marine Ecology Progress Series*, *436*, 131-144.

Wilberg, M. J., Wiedenmann, J. R., & Robinson, J. M. (2013). Sustainable exploitation and management of autogenic ecosystem engineers: application to oysters in Chesapeake Bay. *Ecological applications*, *23*(4), 766-776.

Reports

Maryland Department of Natural Resources. 2018. A stock assessment of the eastern oyster, Crassotrea virginica, in Maryland waters of Chesapeake Bay. Draft Report for Peer Review, July 2018. Annapolis MD. 341 pages.

Oyster Metrics Workgroup**.** 2011**.** Restoration Goals, Quantitative Metrics and Assessment Protocols for Evaluating Success on Restored Oyster Reef Sanctuaries. Report to Sustainable Fisheries Goal Implementation Team of the Chesapeake Bay Program. https://www.chesapeakebay.net/channel_files/17932/oyster_restoration_success_metrics_final.pdf

Maryland Department of Natural Resources. 2016. Maryland Oyster Population Status Report. 2016 Fall Survey. http://dnr.maryland.gov/fisheries/Documents/Oyster-Population-2016.pdf

Maryland Department of Natural Resources. 2016. Oyster Management Review 2010-2015. http://dnr.maryland.gov/fisheries/Pages/oysters/5-Year-Oyster-Review-Report.aspx

Appendix 1. Agenda

Draft Agenda

Oyster Stock Assessment Peer Review.
EPA Fish Shack: 410 Severn Ave., Annapolis, Maryland August 27- August 28, 2018
Maryland DNR - August 29, 2018

Panelists: Dr. Daphne Munroe, Rutgers University; Dr. Paul Rago, Woods Hole Massachusetts; Dr. Dan Hennen, Woods Hole Massachusetts

Day 1 - August 27

9:00 - Coffee, continental breakfast

9:15 - Welcome and introductions

The intent for this meeting is to provide opportunity for plenty of discussion and iterative exploration of questions pertaining to the model and its results.

9:30 - Presentation focusing on TOR1 - data. Includes context of assessment time frame, treatment and processing of data for input to assessment, development of priors for model, and experiments of box disarticulation rates.

11:00 - Questions / discussion

12:00- working lunch (provided)

12:30- Presentation of model fit and diagnostics including sensitivity analyses

sensitivity analyses

1:30 - Questions / discussion

5:00 adjourn

Day 2 - August 28

9:00 - coffee, continental breakfast

9:15 - resolution of any issues from previous day

10:00- Presentation of biological reference points focusing on TOR2

11:00 - questions from peer Review Panel

12:00 - working lunch (provided).

2:00 -Overview presentation of index based approaches and comparison of utility of index vs model-based approaches, focusing on TOR 3.

2:30- Questions / discussion

4:00 - Discussion of TORS 4 and 5.

5:00 - Adjourn

Day 3 - August 20

Change of venue *- MD DNR building 580 Taylor Ave. Annapolis, Md. Conference Room C1. Please have photo ID for security desk.*

This will be a closed session for the Panel to begin synthesizing conclusions. The stock Assessment Team will be on call. Team will re-convene with Panel for overview of conclusions. I have left time for lunch, but we can be flexible depending on travel plans. I can provide guidance on quick nearby places to get a bite.

9:00 am - closed session for peer Review Panel to begin drafting report.

stock Assessment Team - on call.

11:30 - Lunch (not provided)

1:00 reconvene with Assessment Team to go over initial conclusions.

2:00 adjourn

Appendix 2. Terms of Reference

The terms of reference for this stock assessment were developed by the stock Assessment Team based on the Sustainable Oyster Population and Fishery Act of 2016 and were reviewed by Maryland's Oyster Advisory Commission:

1) Complete a thorough data review: survey data, reported harvest and effort data, studies and data related to population rates (growth, mortality and recruitment), available substrate, shell budgets, and sources of mortality.

> a) List, review, and evaluate the strengths and weaknesses of all available data sources for completeness and utility for stock assessment analysis, including current and historical fishery-dependent and fishery-independent data.
>
> b) Identify the relevant spatial and temporal application of data sources.
>
> c) Document changes in data collection protocols and data quality over time.
>
> d) Justify inclusion or elimination of each data source.

2) Develop stock assessment model or index based approach that estimates biological reference points and document status of the stock relative to estimated reference points. To the extent possible, quantify sources of uncertainty within model.

3) Compare estimates of stock status generated by index and model-based approaches. Justify selected approach.

4) Include sanctuaries and restoration efforts in sanctuaries in the development of stock assessment approaches.

5) Examine how hatchery plantings (aquaculture and public fishery) impact spawning potential in the fishery.

Appendix 3: List of Participants

Name	***Affiliation***	***Email***	**Dates at Review**		
			8/27	***8/28***	***8/29***
Review Panel					
Paul Rago	Mid Atlantic Fisheries Management Council Scientific and Statistical Committee	Paulrago22@gmail.com	X	X	X
Dan Hennen	NOAA Fisheries, Northeast Fisheries Science Center	Daniel.hennen@noaa.gov	X	X	X
Daphne Munroe	Rutgers University	dmunroe@hsrl.rutgers.edu	X	X	X
Assessment Team					
Mike Wilberg	University of Maryland Center for Environmental Science (UMCES); Professor	Wilberg@umces.edu	X	X	X
Kathryn Doering	UMCES; Graduate Research Assistant	Kdoering@umces.edu	X	X	X
Trey Mace	UMCES/Maryland DNR; Assistant Research Scientist	Marvin.Mace@maryland.gov	X	X	X
Lynn Fegley	Maryland DNR; Director Stock Health, Data Management and Analysis Division	Lynn.Fegley@maryland.gov	X	X	X

Alexei Sharov	Maryland DNR; Stock Assessment Scientist	Alexei.Sharov@maryland.gov	X	X	X
Linda Barker	Maryland DNR; Research Statistician	Linda.Barker@maryland.gov		X	
Mitch Tarnowski	Maryland DNR; Shellfish Biologist - surveys	Mitch.Tarnowski@maryland.go	X	X	X
Frank Marenghi	Maryland DNR; Shellfish Biologist – commercial fisheries	Frank.Marenghi@maryland.gov	X	X	X
Amy Larimer	Maryland DNR; Shellfish Biologist	Amy.Larimer@maryland.gov	X	X	X
Jodi Baxter	Maryland DNR; Deputy Director Shellfish Division	Jodi.Baxter@maryland.gov	X	X	X
Support Staff / Peer Review Panel Note Takers					
Rachel Pierce	Maryland DNR	Rachel.Pierce@Maryland.gov//	X	X	
Laurinda Serafin	Maryland DNR; Shellfish Division	Laurinda.Serafin@Maryland.gov	X	X	

Appendix 4: Research Recommendations of Assessment Review Team

The following research recommendations were developed by the stock Assessment Team (Maryland Department of Natural Resources and University of Maryland Center for Environmental Studies) in the process of completing this stock assessment. They are arranged by category rather than in order of priority.

Data

- Develop mechanisms to improve accuracy and resolution of reported harvest data including bar level data, the number of licensed individuals on a vessel, and the hours spent harvesting.

- Conduct fishery dependent sampling of oyster size distribution to better quantify the number of oysters per bushel and the number of under-sized oysters per bushel.
- Conduct research to better quantify growth rates that can be incorporated into stock assessment models.
- Conduct research to better quantify natural mortality of wild and hatchery -planted spat.
- Develop a means to mark hatchery-reared planted spat so that the proportion of planted versus wild oysters can be determined in subsequent surveys.

Natural Mortality

- Studies to improve estimates of box decay rate. Because box abundance is a critical element in the estimation of annual mortality, understanding how long boxes persist under varying conditions will improve estimates of natural mortality.
- Explore the effects of timing of the harvest relative to when fall survey is occurring to see if explains some of the difference between model-based and box count estimates of natural mortality.
- Research to better define longevity and identify primary sources of natural mortality of oysters.
- Examine resiliency of oyster populations to high natural mortality events.

Exploitation Rates

- A survey conducted just prior to and directly following the fishery would provide a direct means to estimate exploitation within a given year and could provide a snap shot of conditions relative to selected reference points.

Habitat

- Conduct more ground-truthing surveys on unverified current SONAR data so that existing sonar data can be accurately utilized in determining oyster habitat.
- Develop comprehensive maps of current oyster habitat within the Maryland portion of Chesapeake Bay.
- Studies designed to quantify the rate of habitat decay would better inform the assessment and reference point models; and would contribute to development of a shell budget.
- Develop a mechanism to better understand how shell plantings contribute to habitat and how habitat is quantified.
- Conduct research examining how harvest gears impact oyster habitat.

Sanctuaries and Spatial Scale

- The contribution of sanctuaries to oyster population and fishery dynamics within a NOAA code is an important question for management and will require finer scale spatial survey data within and outside of sanctuaries as well as more accurate bar-level harvest data than is currently available.

- Conduct research to help elucidate how individual NOAA codes (as well as sanctuaries and fished areas) contribute to one another's oyster populations. This would allow for a more complete stock assessment model that incorporates feedback among areas rather than the current assessment which treats each NOAA code as though it is an isolated population.

Assessment Model

- Incorporate a shell budget into stage structured assessment in order to allow internal estimation of biological reference points.

- Continue to improve the stock assessment model based on lessons learned from this assessment and as new information becomes available.

- Examine alternative spatial structure for stock assessment.

Biological Reference Points

- Fishing reference points for oysters should account for the accretion and loss of shell since oysters produce their own habitat that is required for population growth. Developing a spawner per-recruit type analysis that instead of egg production represents shell per recruit. Research is needed to determine the ratio of shell per recruit that is suitable for target and threshold reference points.

- Research on target levels of abundance including biological limits of abundance (e.g. necessary conditions for successful fertilization).

Aquaculture

- Developing an aquaculture data base that tracks plantings, standing stock and harvest of diploid and triploid oysters at the NOAA code spatial scale would be improve the model's ability to quantify the contribution of aquaculture plantings to the population dynamics within the NOAA code.

Appendix 5: Composite Questions from Review Panel

Last updated August 25, 2018

Munroe comments noted with italics:
`Hennen comments noted in typewriter font:`
Rago comments noted in regular font:

ADDRESSED 8-27-18

Longevity

1. How important is the lack of longevity estimate (p.2)? Given the high Z's it seems unlikely that MD oysters would have much chance to realize their natural lifespans, but it knowing the maximum age might inform/constrain realized biological reference points. Given information on the incidence and lethalilty of the diseases, could you compute an expected lifespan in the absence of other sources of mortality?
2. *Regarding longevity: the cited longevity (20 years) is likely a pre-disease condition, meaning that in today's oyster world of disease and fishing, we would rarely see oysters older than 7 years. As an example, Harding et al. (2010) saw very few age 3+ oysters, and none older than that, in her Piankatank survey. Oysters are notoriously hard to age as well. I wonder how important this might be here and whether it should be constrained to a lower value for longevity given that oysters exist today in a world that likely constrains the oldest age classes to somewhere closer to 7.*

Growth—surprisingly, none

Disarticulation

1. Based on experimental evidence the box disarticulation rate reveals mu_d=0.51 and sd=0.04, p. 27. This would imply a low probability of exceeding 1 yr of hinge. P(d>~1) would be at 1-cdf_PHI(0.51+12*0.04). The results from the posterior distribution suggest the model fits better if the half life is twice that measured in the experiments (Fig 22). Does that seem reasonable? You mention that boxes may not survive the dredging process. Quick test--If observed box catches were doubled to account for the breakage of half of the shells during dredging, what is the impact on the posterior of the disarticulation rate?
2. *The corrections for boxes persisting longer than one year is a difficult issue. Mortality estimates from observations of boxes is important. Your model gives you an estimate that 67% of boxes disarticulate in one year. We have 4 years of data across the Del Bay stock that shows that the disarticulation rate varies spatially and by season. It appears much more complicated than initially thought. It seems that some boxes disarticulate rapidly, while others persist. How sensitive is your model to the result that about 2 thirds disarticulate within one year?*

Natural Mortality

1. *There are cases where it is assumed natural mortality occurs outside of the fishing season (during summer), however, it is also noted (see page 22, top paragraph) that disease mortality occurs during winter. I think it is true that you see natural mortality year round. How consequential is this to the assumption in the model that natural mortality does not occur in winter when the fishing season occurs.*

Habitat

1. Does the fixed d=0.16 (p. 60) compare favorably to the multiple d estimates by area estimated in stage-based model (p. 38)?
2. What is effect of assumed exponential decline in habitat in the assessment model? (p. 38, second eqn.) Why does habitat appear to have small positive slopes in some cases? Is this the effect of shell planting?

3. Is there a conflict between the habitat trajectory on page 38 for the stage based model and the habitat model for BRP on page 60?   In the stage based model, H_y is independent of stock size and it inexorable march to oblivion is only a function of time, a time invariant constant (d), and intermittent replenishment. Shouldn't the potential for self-generated habitat (as function of density) exist in both models?
4. *I would like to have some discussion about the conversion of the calculated oyster habitat to areal habitat. What are the data that inform this conversion? If I understand, you use 20 live oysters/m2 as a max value? This is definitely not a max in terms of ecological carrying capacity. As an example, in Delaware Bay we measured, on average across the entire stock, 56 oysters/m2 in 2017. Our highest density grids were in excess of 380 oysters/m2. Additionally, there would be more than just the live oysters on the bottom providing habitat – dead shell, other cultch etc. Can the data from the tong survey be used to help inform this conversion?*
5. *Shell and habitat are certainly important in oyster dynamics. Having habitat doesn't necessitate catching spat – catching spat is a magical mystery. Nonetheless, I think it is good that you are trying to find a way to include habitat in the assessment model. In the habitat calculation on page 60, I don't see a term for dead oyster shell or shell planted with or without spat? Dead oysters are pretty effective habitat and shell persists. Should be included.*
6. Your insights on habitat will be helpful. The habitat component of the stage based model is pretty much going along for the ride in the Likelihood fcn since it is not coupled to stock dynamics. As you note, the creation of new habitat from natural mortality (ie box density) does not affect the trajectory of H_y.   I'll need to triple check this but that's what it looks like now.
7. *Section 1.2 discusses importance of substrate and the fact that siltation can be a problem. This is most certainly true and is known by baymen who plant shell. There are many anecdotal cases where it is noted that dredging will help to clean cultch, allowing it to catch set more effectively. This should be noted here as it pertains to the idea that silt is a problem.*

Data Issues/Decisions

1. Spatial questions (p. 18)
    a. Do any of the Yates bars span multiple NOAA codes?
    b. Do any of the sanctuary areas span multiple NOAA codes?
    c. Are harvesters free to move about all NOAA code areas?

d. Is there any evidence that the CPUE within a NOAA code follows an Ideal Free Distribution? Highest local concentrations fished first followed by rapid decline in average CPUE as fleet spreads to less profitable areas. This would tend to give overly pessimistic results for depletion (high exploitation).

2. Any relationships between the natural mortality estimates, box densities and the disease incidence estimates from the disease bar survey? (p.35)
3. *An annual dockside monitoring program is probably worthwhile to help you pin down number of small oysters being caught per bushel, and to estimate the size and number of oysters per bushel. Appendix 3 gives you some useful information, but sample size is small and only captures one point in time during the season. We find in our dockside monitoring that count/bushel and size/bushel varies through the season as the fishery targets different beds, and can be pretty different one year to the next.*
4. *Harvest reporting adjustment assumes 10% underreporting. The assumption here is that the first 2 years of harvest reports represent 'true behavior'. I am not convinced that is the case and would argue that the tax values may be a better representation of 'true behavior'. Could the tax be used to represent the harvest timeseries, with harvest allocation proportional by NOAA code based on harvest reports of buyer tags?*
   a. *Either way, the sensitivity analysis relative to the 10% adjustment should be shown.*
5. *You estimate, using the model, differences in efficiency between live oysters and boxes. We have data from multiple gear efficiency experiments that put some hard numbers on this. Observations show that it varies, and that dredge gear will catch boxes more efficiently than live oysters, and cultch more efficiently than both. Here is a table from the most recent assessment document.*

Table 1. Catchability coefficients for oysters, boxes, and cultch by region. The entire time series since 1953 was reconstituted using these catchability coefficients as of 2016 SAW.

	Catchability Coefficient		
Region	**Oyster**	**Box**	**Cultch**
Very Low Mortality	2.41	6.82	9.11
Low Mortality - *Round Island*	2.41	6.82	9.11
Upper Arnolds, Arnolds	8.26	12.69	25.79
Medium Mortality Transplant	8.26	12.69	25.79
Medium Mortality Market	8.26	12.69	25.79
Shell Rock	8.26	12.69	25.79
High Mortality	2.82	5.10	8.46

Mortality Model

Is it a little strange that M has high variance early in the time series and then stabilizes later in all codes? It could be that this model is tracking real and interesting changes in M bay wide, or could this be an artefact of the modeling?

Assessment Model

1. Model building process
   a. Will likely need some description of evolution of model by the assessment team.

i. Nat Mortality model
ii. Assessment model
iii. BRP

b. The current text gives some insights about the assessment teams deliberations. A short description/presentation on your tortuous path might prevent annoying reviewers from asking about things you already did.

2. Quick check. If timing of survey is cause of model conflict and gear efficiency is low, how much effort would be required to create an exploitation rate of 35%? (p. 111)
   a. Simple calculation. Given density estimates from model and estimated efficiency, how much effort would be needed to obtain observed yield? Does this relate well with what is known about fishing activity (e.g., #trips, #person hours, etc.)?
3. Differences in timing of surveys are mentioned as a source of model conflict (p. 41-42). Could the Ricker Type 1 fishery be a source of problem (vs Type 2)?
4. The disparity between the Stock Model estimates of M and the Bayesian natural mortality model (p. 41, Fig 61-66) suggests that the model is reducing the abundance by increasing M. Since M and q are inversely related it would be valuable to look at likelihood profile over a range of fixed q.
5. The stage based model assumes that recruitment is independent of stock size, whereas the BRP model assumes recruitment is proportional to stock size as a time invariant scalar and reduced by the ratio of abundance to a time varying carrying capacity. I see the utility of using coupled logistic models for the BRP but shouldn't some of the rationale be extended to the stage based model?
6. Effect of additive constant (0.001) in p.39 bottom eqn. Such constants often cause mischief in estimation and it may we worthy of a quick test or two.
7. *Section 2.2.2 explains how exploitation and fishing mortality are calculated. I am not sure I agree with two assumptions for this.*
   a. *One is that removals are large enough to cause a decline in abundance, or CPUE. Is this always the case? If I follow the logic here, the assumption is being made that because harvesters catch their max daily rate most often at the early part of the season, that means that CPUE is dropping off. I can think of other reasons that this behavior may be seen that are not related to CPUE. One is market driven. The fishery in Delaware Bay is primarily a summer fishery, and the fishery ends around the time that MD opens. That would mean that there is market incentive to fill the gap left by NJ catches. Also, the weather and conditions would be best early, and would decline as you move into December, Jan/Feb. We see a similar trend in clam fisheries. And I would guess that the fishers who have general permits would move onto fishing other stocks (stripers start in December in MD??) and so effort would shift because other stock dynamics, and not because of declines in CPUE. Have any of these alternatives been explored through conversations with fishers, or other means?*
   b. *The constant catchability assumption is likely invalid. Summer harvest is closed, and that is when oysters grow. Reefs would aggregate during that time and that means catchability declines. So when the fishery starts, that is when catchability would be at its lowest, and as the fishery works beds, the catchability tends to increase.*
8. *You calculate separate efficiency ratios for different life stages. This would be a correct approach for animals that live independently (are not attached to one another), however oysters attach to one another, so what we actually see is that efficiency is the same for all life stages. You are equally efficient at catching a live large oyster as you are a spat because the spat is probably*

attached to a big piece of shell or a live oyster. Can the efficiency for all life stages be constrained to a single efficiency, and if so how does that affect results?

9. *Section 4.2, model outputs. Many of these exploitation rates seem very high. Is this because of the small scale that it is being estimated on (the NOAA code region). How might these look if exploitation was estimated regionally, or stock-wide? Do fishers move among beds or NOAA code areas within a season, or from one year to the next? For example, would they hit one area very heavily, then move to another place in a subsequent year leaving that heavily fished area fallow?*
    a. *Overall this gets me wondering whether it is worth exploring ways to group NOAA codes that better capture broader fishing activity and resource use. Maybe regionally?*
10. *The assessment model estimates higher mortality than the box counts or the mortality model. This is interesting, and it is something that we also see in the Delaware Bay assessment. You would see this result if there are unaccounted sources of mortality out there. This may happen with small oysters as small boxes may pass through the dredge, or as you point out for boxes that disarticulate quickly. In our experiments, we have noted up to 20-40% disarticulation within the first 60 days after death, so definitely could be a source of this discrepancy. You might also see this discrepancy if you have growth that means some oysters skip a stage, or stay too long in a stage. This could happen if growth is slower than expected (small oysters remaining small), or faster (spat jumping all the way up to market size). Thoughts on whether that may be the case?*
    a. *Growth is a very difficult thing to get at with oysters. In research recommendations you propose to study this to get a better idea of realized growth rates in your system. This will be highly variable by bed and year. It is worth doing, but it takes a good deal of work and needs multiple years of experiments.*
11. *You calculate an exploitation rate that accounts for planted oysters (pg. 62). In it, you assume that mortality is the same for wild and planted oysters. I think that would be ok to assume if plantings are seed that were wild collected elsewhere, but not if they are hatchery oysters (spat-on-shell). Likewise, using wild estimates of growth and mortality for aquacultured oysters may not be valid. Farmers may use selected lines to produce spat that would grow/survive differently, or they may tend their plants such that growth is different.*
    a. *I suggest adding this as a caveat in the list on page 68.*
12. I would like to see a little more on the sensitivities that were done. That section of the report is sparse and a little confusing. How much were parameters changed? Why choose to change those parameters? Why choose the performance metrics that were chosen? Are the histograms reflecting some sort of time series average, terminal year point estimates, something else? Model stability is hard to judge from this section.

13. As Paul notes the fit to the CPUE data is not great in many NOAA codes and it appears that is on purpose, as there is mention of sensitivity testing being used to down weight the CPUE trend data in the likelihood. I am a little surprised that the fit to CPUE makes any difference given it's tiny contribution to the total likelihood. Am I missing a scaling component somewhere? Perhaps I am looking at the wrong likelihood component?

14. I am unclear on how habitat is used in the model. It is the denominator in the density index, but what data is informing the model? There must be something as d is estimated (and appears to be fitting to something).

15. I expect we will hear a little more about the q's? I am confused about how to interpret them. Should there, for example, be a tighter correspondence between the R in the mortality model (describing the ratio of catchability between lives and boxes) and the ratio of q_sm,mk to q_B? Particularly since q_sm,mk and q_B penalized if they diverge very much? Probably this isn't important, as the quantities appear to serve different purposes, but I would appreciate some help in wrapping my brain around the differences!

16. Is recruitment constrained? Other parameters not discussed in the report? The implementation section doesn't talk much about the set up. I will attempt to dig into the TMB code on Sunday, but I am not an expert.
17. Is it a little weird that the patent tong survey densities are fit better than the dredge survey? I worry some about the fixed survey design coupled with changing habitat.

18. There don't seem to be any results from the fall dredge standardization model presented.
19. The model also creates a disparity between the experimental results for disarticulation and the posterior dist estimated by the model. This probably occurs since the model needs to match the observed box counts.

Index Methods

1. The natural mortality model and assessment model are state-of-the-art state-space models but the depletion analyses use simple linear regression estimators (p.70). While these are good starting points, the better MLE estimators (Gould and Pollock 1997, Seber 1973) are not used. Wholescale revision is not recommended but it might be worth checking a few examples and seeing if it makes a difference. If exploitation is low and observation error is high then neither method will work very well. Theoretically the MLE methods are better, but in practice there may not be too much difference!
    a. Could devise MLE with common N_o and multiple q_i for each gear.
    b. In the DeLury, Ricker and Leslie-Davis models, N_o and q_i have strong negative sampling covariance. This occurs since the intercept is the product of the two parameters of interest.
    c. Error structure is not IID because variance of binomial varies as removal continues. Multinomial, or multinomial with inflated variance may be more appropriate.
    d. Gould, W. R. and K. H. Pollock. 1997. Catch-effort maximum likelihood estimation of important population parameters. Can. J. Fish. Aquat. Sci 54: 890-897.
    e. R code = https://rdrr.io/cran/fishmethods/man/deplet.html I haven't worked with this code but Gary Nelson wrote it so it is probably pretty good.
    f. Could you provide a little more detail on your 50 report threshold for estimation feasibility (p. 71)
2. *An important piece of data that is missing is an estimate of dredge efficiency for the survey. Do you have any data for this? In our work in Delaware Bay, we have found that this is critical to*

estimate, and it can vary spatially and (importantly) with oyster density making low density estimates problematic. I have provided Jason Morson's paper about this issue and would like to discuss this in terms of the abundance estimates being made here for low density samples.

3. *In some cases, where possible, for a given NOAA code, abundance is estimated independently for gear types. How often do those estimates disagree, and by how much?*
4. The disparity between the model estimates of exploitation and the depletion analyses suggests some concerns with model driving down abundance by inflating M during the "M-only" part of the year.

5. 2. If exploitation is as high as suggested, then how much effort would be required to reduce overall abundance by that amount given expected dredge efficiency?

6. The authors provide the depletion estimates (ie slopes) in the .csv file "DepEst..." but the units of the slope are not specified. These have a phenomenological interpretation (ie known dimensions) so we should pursue this. See above.

Reference Points Model

1. The BRP model implicitly assumes that Dermo and MSX are here to stay and that future dynamics will be dominated by them. I agree with this decision. Is there any evidence of trends in these diseases that might be important for the reference points?
2. How often did model parameters hit penalty bounds? How are these runs treated? It looks like the q parameter in the BRP model (p.60) approaches the boundary often (Table 13, p 104-5). High values of q would tend to increase the dynamic K=H_y in second eqn on p. 60. This in turn would lead to higher rates of population growth because of the N_y/H_y term gets smaller. Taken together, wouldn't this tend to result in higher u_MSY and u_crash values (p. 62)?
3. I am a little concerned about shifting baselines. I don't have a better solution, but some of the abundance limits are being set in the terminal year (in codes experiencing a one way trip to the toilet), which doesn't seem like a great way to do business. If I am interpreting correctly shifting baselines may also be a problem for the exploitation rate reference points depending on where r and q (production rates of oysters and habitat) come from.
4. Should there be a linkage between M_y and habitat production?

Hatchery Contributions

1. *On page 69 you note that it is likely that based on overall abundance, because wild oysters are much more abundant they will contribute more to spawning capacity of the stock and lease oysters would be negligible. This would be valid if the density on the bottom is the same across both types. But your wild oyster density seems very low (you mention a max of 20/m2). A leaseholder would likely plant at a higher density on a lease so that it is easier to tend etc. The higher density may mean that fewer oysters overall could be more effective spawners because in closer proximity fertilization efficiency goes up. Do you have data on this?*

Picky notation/format Issues

1. Picky notation/format Issues
    a. Multiple uses of n
        i. Use n as an index for observation on page 24
        ii. Use n as number of planted market oysters (p. 62)
    b. Multiple uses of q
        i. Slope of DeLury depletion estimator= catchability (P. 14)
        ii. Catchability estimates in Assessment model (p. 40-41, p. 100).
        iii. Penalty function for catchability
        iv. Intrinsic rate of habitat production (p. 60)
        v. Use q_h = 1/20 as conversion factor for Yates estimate (p. 60)
        vi.
    c. Use H_hat_t in second equation on p. 60 but use H_hat_y in next line.
    d. It would be useful to have units of parameters. E.g., Table 12 for q, d etc. Some of these look like there may be some scaling factors eg x 100, etc.
    e. Figures 47-60 seem to have a formatting problem. Can the labels be aligned better?

- *Perhaps Mitch Tarnowski could attend the meeting at least for a period during which we will discuss data in case there are questions or discussion that he could help with.*
- *In section 2.2.1, there is reference to depletion analysis in 6.1, but I think it is actually found in 2.2.2.*
- *In figure 5, could the NOAA codes on the x-axis be grouped by region and have the region names listed below?*
- *In figure 7, why are there 2 colors being used for the bars?*

2. At the risk of making complicated graphs more complicated, it seems like the time series of annual landings could be added to Figures 25-60. (You left a space for it on the lower right corner!)