WESTERN PACIFIC STOCK ASSESSMENT REVIEW

Benchmark Review of the 2016 Stock Assessment of the Main Hawaiian Islands Reef-Associated Fish

Consensus Review Panel Report

August 29-September 2, 2016

Prepared for
Pacific Islands Fisheries Science Center, National Marine Fisheries Service, NOAA
Pacific Islands Regional Office, National Marine Fisheries Service, NOAA
Western Pacific Regional Fishery Management Council

Prepared by
Erik C. Franklin (Chair)

Executive Summary

This document provides the consensus report for a Western Pacific Stock Assessment Review (WPSAR) of the "Benchmark Review of the 2016 Stock Assessment of the Main Hawaiian Islands Reef-Associated Fishes of Hawaii, 2016". This WPSAR addresses a set of five (5) Terms of Reference (TOR) for benchmark stock assessments of 28 species of reef-associated fishes in the Main Hawaiian Islands, following guidelines established in the Western Pacific Stock Assessment Review (WPSAR) framework. The review was held August 29, 2016 through September 2, 2016 in Suite 1701, Finance Factors Building, 1164 Bishop Street, Honolulu, Hawaii 96813. This review considers the following 5 TORs with a brief description of their purpose: TOR 1 is the response to the CIE reviews related to the assessment methods; TOR 2 is the appropriateness of the approach applied to 28 coral reef-associated Hawaiian fish stocks; TOR 3 is whether results are scientifically sound; TOR 4 is whether results are useful for management purposes; and TOR 5 provides recommendations for improvements and future research.

The content of this review addresses the five TORs applied to an examination of the length-based stock assessments of 28 coral reef fish species in the Hawaiian Islands. The stock assessment approach incorporated population abundance estimates and length composition data from fishery-dependent catch statistics and fishery-independent diver surveys with life history parameters for longevity, growth, survivorship, and maturity. A stochastic simulation approach (called the "step-wise approach") was used to obtain demographic and life history parameters for species with no species-specific data available. The foundation of the assessment approach was the use of the average length in the exploited phase of the population to estimate total mortality rates and fishing mortality rates for each species. The method assumed equilibrium conditions and a survivorship function that determined the natural mortality rate. A numerical population simulation, incorporating the mortality estimates and demographic parameters, was used to generate probability distributions for metrics on stock status including spawning potential ratio (SPR) and the ratio of fishing mortality to fishing mortality at an SPR of 30% (F/F30). Probability distributions for Catch30 (i.e., catch for SPR of 30%) and OFLs were determined from the population abundance estimates (using catch estimates or diver surveys) and the estimated F30 distribution. These methods were previously reviewed by the Center for Independent Experts (CIE) in 2015, and TOR 1 addressed the response of the stock assessment author to the CIE recommendations. The panel responded to the ten sub-questions of TOR 1 (i.e., TOR 1(a) – TOR(j)) with "yes" or "no" answers and caveats identified when necessary. For TOR 1, the panel replied with a consensus "yes" or "yes with caveats" for 6 sub-questions and "no" or "no with caveats" for 4 sub-questions. Not all panelists replied with an explicit "yes" or "no" response to TOR 1 so these replies were inferred by the chair for this report. Detailed responses to each sub-question of TOR 1 are included in the individual reviews of each panel member.

In general, the panel found that although the approach is a sound method to assess the status of "data poor" coral reef fish stock in Hawaiian waters, it should be carefully applied on a species-by-species basis. Thus, the panel accepted the stock assessment approach in the Tech Memo as feasible and appropriate for species-specific assessments of Hawaiian coral reef fishes as identified in TOR 2, and likely to provide scientifically sound results as identified in TOR 3 for management purposes as identified in TOR 4 but with the caveat that this acceptance is reliant on improvements and recommendations from this review being incorporated for the final stock assessments. The sequentially intertwined nature of the textual definitions for TORs (2), (3), and (4) made it

challenging to comment on them individually for each species. Thus, the panel responded in aggregate with a “yes” or “no” answer for each species to the three TORs. For TORs (2) – (4), the panel replied with a consensus “yes” or “yes with caveats” to thirteen species and a majority “yes” or “yes with caveats” to 15 species. Not all panelists replied with an explicit “yes” or “no” response to TORs (2) – (4) so these replies were inferred by the chair for this report. Detailed responses to TORs (2) – (4) for each species are included in the individual reviews.

To address TOR 5, the panel recommends a number of improvements necessary for the assessments to be considered final. The decisions made regarding data inputs, assumptions, and analysis steps should be identified and documented explicitly for each species in the comment sections. All assessments should be rerun with the survivorship rate of 4.35% or 4% (or lower if appropriate and justified), not the current rate of 5%. The current assessment method can generate negative values of fishing mortality, an unrealistic situation that requires further refinement and justification. A more detailed description of the workflow including software and data files should be included in the methods. Better descriptions of the data sources, particularly of the catch estimates and diver surveys, need to be included as components of the enhanced species comment sections. Sensitivity tests should be conducted, especially for species near thresholds for stock status metrics, to explore the impact of data inputs and assumptions on management-relevant results. These recommendations are summarized in more detail in the individual reports for a range of critical, short-term to suggested, long-term improvements to be made to the assessments.

In conclusion, final assessment results should be well justified and documented in the Tech Memo to provide a clear understanding of their reliability for the status determination and management evaluation of each species. The panel does not regard the draft stock assessments and current report as sufficient to meet advisory or management needs. While the panel generally accepts the assessment method and believe that it can be effectively used for management purposes, there are general and particular species-specific concerns for TORs (2), (3), (4) and recommendations for TOR (5) as detailed in the individual reviews that need to be addressed in the final assessments. Addressing these concerns and recommendations should be possible in the time available and likely would provide sufficiently sound scientific results to support management decisions for these fish stocks.

BACKGROUND

Section 301(a)(2) of the Magnuson-Stevens Fishery Conservation and Management Act (MSA) requires that fishery conservation and management measures be based upon the best scientific information available. MSA § 302(g)(1)(E) provides that the Secretary of Commerce (Secretary) and each regional fishery management council "may establish a peer review process for that Council for scientific information used to advise the Council about the conservation and management of a fishery." Consistent with this provision, the Western Pacific Regional Fishery Management Council (Council), NOAA's National Marine Fisheries Service (NMFS) Pacific Islands Fisheries Science Center (PIFSC) and the Pacific Islands Regional Office (PIRO) have established the WPSAR process. WPSAR is a cooperative effort to improve the quality, timeliness, objectivity, and integrity of stock assessments and other scientific information used in managing fishery resources in the Pacific Islands Region. The WPSAR process may be applied to scientific information used by the Council directly to fulfill its management mandate in the execution of the MSA.

The WPSAR framework document outlines the scope of WPSAR, defines roles and responsibilities, summarizes the various review levels, describes the sequencing and timing of the WPSAR process in coordination with the larger Council process, and provides mechanisms for resolving disputes. This framework is available from the WSPAR website, at: http://www.pifsc.noaa.gov/peer_reviews/wpsar/index.php.

PIFSC scientists are conducting stock assessments on exploited coral reef fish species in the Pacific Islands Region which are listed in the Council's Fishery Ecosystem Plans. These stocks are generally classified as data-poor due to a lack of reliable, long-term, catch and fishing effort data. Historically, the Council has set and NMFS has approved setting of annual catch limits (ACLs) using a percentile of median historical catch levels and more recently, a biomass-augmented catch-MSY method has been applied (Sabater and Kleiber 2014, NOAA 2015).

In an effort to use additional available data sources for these stocks, scientists at PIFSC have conducted new coral reef fish assessments using length composition data, abundance data from diver surveys, and certain key population demographic parameters related to growth, maturity, and longevity. PIFSC scientists have been implementing an approach that uses the average length in the exploited phase of the population (L_{bar}) to obtain an estimate of total and fishing mortality rates for coral reef fish stocks (Beverton & Holt 1956; Ehrhardt & Ault 1992). These rates, combined with population demographic parameters, are used in a numerical population model to obtain stock sustainability metrics (e.g., spawning potential ratio, F/F_{MSY}; see Ault et al. 1998, 2008). Overfishing limits can be generated by using recent total catch estimates and/or population size estimates from diver surveys. Furthermore, a novel meta-analytical approach using stochastic simulations was developed at PIFSC to obtain demographic parameter estimates for species with even less data than data-poor species ("data-less" species). These scientific methods recently underwent a rigorous independent review by a panel organized by the Center for Independent Experts, and have now been applied to individual species in the main Hawaiian Islands. There is a need to independently review these species-specific stock assessments prior to submission to a fishery management organization for consideration.

DESCRIPTION OF THE CHAIR'S ROLE IN THE REVIEW ACTIVITIES

This Western Pacific Stock Assessment Review (WPSAR) Benchmark Review consisted of an in-person panel of one review chair (Franklin) who is also a member of the WPRFMC's Scientific and Statistical Committee (SSC), plus 2 additional review members (Choat and Stokes) external to PIFSC, PIRO, and the Council and its affiliated bodies. The review took place in a conference room on the 17th floor of the Finance Factors Building at 1164 Bishop Street, Honolulu, Hawaii 96813 USA from August 29, 2016 through September 2, 2016. An agenda for the meeting and list of attendees is included in Appendix 3.

For the review, I provided independent and impartial scientific expertise and in my role as a reviewer did not represent my respective institutions or affiliations. I followed and complied with my role as chair as outlined in the WPSAR Terms of Reference (TOR; see Appendix 2). I read all required provided documents in advance of the meeting, actively contributed during the in-person panel review, and provided this review report on my scientific opinion addressing all aspects of the TOR (Appendix 2). The review report follows the format outlined in Annex 1 of the TOR (Appendix 2). I provided this individual report to the WPSAR Coordinating Committee point of contact (M. Dunlap) by email after the close of the review.

During the in-person review, I was present for an overview of the WPSAR process (A. Yau), Hawaii state fishing regulations (A. Miyasaka) and the fishery management process to identify annual catch limits (ACLs; M. Dunlap). In addition, I was present for detailed talks from the stock assessment author (M. Nadon) on the general assessment approach, incorporating uncertainty into the stock assessments, data-poor life histories, and recommendations from the CIE review. Following these presentations, Nadon described each stock assessment for the 28 species examined in the report allowing for interactions with the panel.

SUMMARY OF FINDINGS FOR EACH TOR QUESTION

For TOR questions 1-4, I provided a "yes", "yes with caveats", "no" or "no with caveats" for consensus or majority responses from the panel. Only if necessary, caveats were provided to these yes or no answers, and were as specific as possible to provide direction and clarification. These guidelines were difficult to follow with an absolute "yes" or "no" in cases with multiple questions within a single TOR. For TOR 1, the panel replied with a consensus "yes" or "yes with caveats" for 6 sub-questions and "no" or "no with caveats" for 4 sub-questions. Not all panelists replied with an explicit "yes" or "no" response to TOR 1 so these replies were inferred by the chair for this report. Detailed responses to each sub-question of TOR 1 are included in the individual reviews of each panel member. The TOR questions 1(a) through 1(j) are listed below with responses from the reviewer after the questions.

1. Review whether each of the following short-term recommendations from the previous independent peer review were addressed properly for the general (not species-specific) approach, considering that the data sources themselves are not up for review. If they have not been addressed, indicate why not and suggest methods for addressing them.

a. *The development of a clear decision chart to increase transparency in the application of the approach. Clearly articulate the hierarchical nature of the three life history approaches and under which circumstances a method should (or should not) be used.*

Response to 1(a): Consensus YES with caveats. The CIE reviews recommended a decision tree approach that would transparently identify choices made when carrying out the assessment methodology regarding survivorship rates relative to natural mortality, M; the catch and biomass data, the sources of the life history parameters, and parameter uncertainty, The decision tree in Nadon (Fig. 6, 2016) addresses these recommendations in a general sense through the three "Steps" with "Cases" for each step. Unfortunately, this general approach often did not capture some decision steps required for each "case" when confronted with the nuances of individual species assessments. Given the various decision steps for this methods, it is unclear if an objective and complete accounting of all procedures is feasible. I feel that the current status of the decision tree is still quite undeveloped and should be further refined as this approach continues to undergo scrutiny and development.

Steps involving the estimation of total mortality (Z) and natural mortality (M) should be better incorporated into the decision tree. The length-based assessment method relies heavily upon an appropriate specifications of Z and related selectivity parameters determined from the size structure data of the population and M as a key component in the estimation of F (and the proposed F_{MSY} proxy, F_{30}). Both mortality estimates need a clear and transparent justification for their choice.

Although the decision tree is sufficient in outlining the process, there may be a benefit to better integrating Figure (6) with Figure 4 ("Overall approach to obtain OFLs") to facilitate use of the methodological approach by analysts other than the stock assessment author. This is not a high priority but may improve an understanding of the interaction between the processes outlined in both figures.

The panel suggested that a system of ordinal ranking for each "case" in the "steps" could be used to provide an overall quality index for each assessment. Also note that "case" may be a poor choice of terminology for these purposes. Assuming the decision tree will undergo further development, I can't suggest a specific system for the ranks but advise that some approach be adopted to create a relative score for the confidence of the data inputs to the assessment. This is not a short-term priority but could be a tool to help guide decisions related to data sufficiency for individual species assessments with this method.

b. Explore an alternative to calculating mean length across islands and applying that in the remainder of the process by using the sectoral mean lengths (the primary index of exploitation) through to the estimate of fishing mortality, and weight resulting estimates to calculate overall fishing mortality.

Response to 1(b): Consensus YES with caveats. The author presented a comparison of various methods to calculate mean length across islands as suggested by the CIE reviews. The reef area weights by sector were determined from a study (Rohmann et al. Coral Reefs 24(3):370-383) that estimated "potential reef area" as defined by the area shallower than the 10 fathom contour derived from nautical charts (Nadon pers. comm). This approach may not accurately describe the reef areas in each sector. I recommend that the author consider an alternative source for sector area weights from Battista et al. 2007 @ https://coastalscience.noaa.gov/projects/detail?key=208. In general, the chosen approach is reasonable for the purposes of this stock assessment but there may be cases when a different approach is warranted and the use of alternative area weights for sector should be explored.

c. Examine the sensitivity of final results to uncertainty in the value of length at first capture (Lc) used. Ensure the method to calculate Lc is more standardized and repeatable by other assessors.

Response to 1(c): Majority NO with caveats. We were presented with examples of the sensitivity of final results to uncertainty in the value of length at first capture (Lc) for two example species but not for the entire suite of species. The "visual inspection" approach used to determine the selectivity ogives at Lc50% and Lc90% from length frequency data. While Nadon provided details of his process, the subjective nature of the approach does not seem to be repeatable due to potential differences in interpretation of the data. The extent that different assessors may vary in their estimates of Lc may not be significant but could introduce an additional element of subjectivity to the process. The panel also expressed that the CIE reviews recommended that the choice of selectivity parameter values be repeatable and that sensitivity testing of those values be performed for all species. Neither of those recommendations seem to fully be meet for this TOR.

d. When incorporating uncertainty in parameter estimates, evaluate the data underlying the coefficients of variation (CVs) derived from Kritzer et al. (2001), and compare them to those derived for species around the Main Hawaiian Islands and U.S. Pacific territories that can be estimated using e.g. the length-at-age bootstrapping approach for von Bertalanffy parameters: growth rate (K) and asymptotic length at which growth is zero (Linf).

Response to 1(d): Majority NO with caveats. The response to the TOR was a comparison CVs for L_{inf} and K among three species derived from Kritzer et al. (2001) with CVs from constrained fits (with a $t_0 = 0$ in the von Bertalanffy growth function), unconstrained fits, or an average of constrained and unconstrained values. The response did not explicitly evaluate the underlying data from Kritzer et al. (2001) but the panel was directed to comment on the comparison of CVs for Linf and K between Kritzer et al. (2001) and Nadon's methods. The panel did not find the averaged Kritzer CVs as a meaningful validation to output from Nadon's method since the justification for the choice was unclear but the intention of the TOR was confusing.

e. Draw maximum length (Lmax) for the data poor life history simulation approach from a distribution rather than using a single point prior to capture this element of uncertainty.

Response to 1(e): Consensus YES. The Lmax values are drawn from a distribution, not a point value.

f. Explore the impact of heavily truncated size data in which the sampled Lmax is not representative of the biological Lmax, what is a safe error in Lmax in terms of biases, false positives and negatives; and relate this to the decision tree.

Response to 1(f): Majority NO with caveats. The TOR response included a presentation of three species examples with a negative bias of 10%, 20%, and 30% in Lmax with associated estimates of Linf, K, Lmat, M, and SPR but did not provide overall guidance on what is a safe error for Lmax in terms of biases, false positives and negatives. There was also no direct relation between this exploration and how it should be integrated into the decision tree. From the CIE review (Dichmont), this recommendation related directly to potential effects of a mis-specification of Lmax as part of a sensitivity analysis step for each species, not as a general exploration.

g. For some Main Hawaiian Island stocks the available Lmax values extended beyond the range of estimates from which the life history parameter relationships were developed for the corresponding family. Therefore, consider the efficacy of estimates and uncertainty developed where input parameters for a species require extrapolation outside the range of data on which the relationships were based.

Response to 1(g): Consensus YES with caveats. For this TOR, Nadon presented an example from family Mullidae of Lmax for species in the assessment that were extrapolated beyond the group Lmax estimate. This situation is not ideal but given the paucity of data his approach to the uncertainty outside the range of the data seemed reasonable. In general, this type of analysis decision should be explicitly included as a component of the general comments section for a species.

h. Research the possibility of using female biomass only for SPR calculations where the male proportion is considered important (e.g. in the case of protogyny; Ault et al. 2008).

Response to 1(h): Consensus YES. Nadon presented research from Brooks et al. 2008 on the topic which could be more fully explored as a long-term stock assessment improvement but is not a priority for this stock assessment.

i. Explore the option of including runs with negative fishing mortality estimates within all calculations and representations.

Response to 1(i): Consensus NO with caveats. The CIE reviews identified this TOR as a major focus for improvement and requested an exploration of alternatives. The current approach presented by Nadon may derive negative F (fishing mortality) values during simulation runs. If a run generates a negative F value, those runs have been discarded from final assessment analysis and visualizations. The rationale for discarding negative F values is that they are infeasible and lead to unrealistic population estimates when deriving data from catch statistics. It was not apparent during the review that any alternative approaches had been explored as stated in the TOR. This is a critical component to the assessment method as the inclusion of negative F values seems to compromise the confidence that one can take in the approach. On the other hand, the discard of negative F values would lead to incorrect estimates of F, F30, F/F30 and SPR.

For the purposes of this assessment, you may consider an alternative method to constrain the life history parameters to a biologically-realistic domain of values. The generation of negative Fs may be a result of simulation draws that early in the assessment method select impossible combinations of values (such as L exceeding Linf). If this is a possibility, you may consider imposing a "biologically realistic" threshold on the set of parameter values generated by the approach during the generation of life history parameters. This process may then eliminate the generation of negative Fs. If this approach is not feasible, you may also consider a censured distribution of Fs that reallocates $F < 0$ to $F = 0$. No matter what final method is chosen, the justification and exploration of the approached should be transparent, documented, and reviewed to ensure confidence in the results.

j. Present OFL distributions arising from all relevant data set combinations separately to ensure levels of uncertainty are understood.

Response to 1(j): Consensus YES. In response to this TOR, Nadon (2016) included Catch30

(~OFL) distributions for each species from catch or diver surveys as well as "combined" in the assessment. The panel discussed the use of OFLs from combined catch and survey data and the attributes and utility of the different data types. In general, the appropriateness of the data source (e.g., catch statistics or diver surveys) should be evaluated on a case by case basis for each species. The panel suggested that the use of a "combined" case for Catch30 may be feasible but would be an unlikely best option for most of the species in this assessment.

The TOR questions 2 through 4 are listed below with responses from the reviewer after the questions organized as general comments and then species-specific comments. These questions have been grouped as they are intertwined at the level of the overall report as well as the individual species assessments.

2. ***Appropriateness of general approach****: Review the appropriateness of the application of the general approach to each individual species being assessed: Determine if decision points and input parameters were reasonably chosen, assumptions reasonably satisfied, and primary sources of uncertainty documented and presented.*

3. ***Scientifically sound final results:*** *Determine whether the final results for each individual species are scientifically sound, including estimated stock status in relation to the selected biological reference points ($SPR_{30\%}$) and overfishing limits.*

4. ***Useful for management purposes:*** *Determine whether the results for individual species from question 3 can be used for management purposes under the Magnuson-Stevens Act and relevant Fishery Ecosystem Plan (FEP), including biological reference points such as MSY-based BMSY, FMSY, and MSY (or their proxies) with no or minor further analyses or changes, considering that the data itself and the general approach have been accepted for stock assessment purposes. If results of this analysis should not be applied for management purposes with or without minor further analyses, indicate which alternative set of existing results should be used to inform setting fishery catch limits instead and describe why.*

General Response to TOR (2), (3), and (4)

The following comments are general to many of the 28 individual stock assessments presented in Nadon (2016). The Tech Memo provides a sufficient overview of the methods and results for each assessment but there are a number of improvements that could be made to clarify the overall approach, points of concern, and robustness of outputs. Following this section are comments for each species-specific response to TOR (2), (3), and (4) that included consideration of life history parameters, input data sources, assumptions, sources of uncertainty, and recommendations.

The review panel has been requested to specifically not comment on issues related to catch statistics and diver survey data. Given the importance of these data inputs to determine population and catch estimates, there is currently a paucity of information in the report on the general quality of data used for the assessments and how these may affect the final results. It would be appropriate for an assessment to include much more detail on the data inputs such as the statistical framework used to generate the catch or population estimates, sampling design and domain, and survey performance statistics especially given the reliance on this data for the current method. A straightforward means to address these deficiencies would be to enhance the descriptive statistics for the data inputs presented for each species as figures or tables. These should include catch statistics, and length distributions through time in a standardized format for each species. They would provide the reader with a better understanding of fishery trends and allow a subjective, visual evaluation of the assumption of equilibrium inherent in the length-based approach.

The current stock assessments were performed with a survivorship rate of 5% for each species. The panel discussed and agreed during the review that the value of 4.35% from Nadon et al. (2015) or a rounded value of 4% should instead be used for all species. In special cases, a lower survivorship rate (and associated estimate of natural mortality, M) could be used if it is appropriate and justified. Each stock assessment should be updated using the recommended survivorship rate.

The Tech Memo includes comments sections for each species stock assessment that could be better utilized to provide discussion from the author about the decision process, quality, and uncertainty of a particular assessment. These sections are populated unevenly among species, with some having comments with a good background and interpretation of the reliability of the assessment and others with very little information. A consistent approach should be adopted for the content in the comments that provides the reader with a better understanding of data inputs, assumptions, sensitivity test outcomes, and interpretation of assessment results. These sections should also provide clear guidance on the appropriateness and use of the assessment for management purposes for reliability of F/F30 and Catch30 estimates. For an existing table (probability of overfishing), the range should be limited to 0.50 with increments of 0.01.

The panel discussed and agreed that the credibility of the stock assessments in the Tech Memo would be improved if multiple scientists (i.e., as co-authors) contributed to the selection and vetting of data inputs, analyses, and presentation of results. The prior CIE reviews identified a number of suggestions to improve the assessments which appear to be the sole responsibility of the author of the Tech Memo. Given the large number of species currently under consideration, a group effort to evaluate the inputs, perform the assessments, and examine the model outputs would greatly improve the final results. Although the CIE reviewers and this WPSAR panel can provide fixed recommendations, the best improvements to the process would be attainted through an ongoing, iterative examination of the stock assessments with a collaborating group of experts.

In general, the panel accepted the stock assessment approach in the Tech Memo as feasible and appropriate for species-specific assessments of Hawaiian coral reef fishes as identified in TOR 2, and likely to provide scientifically sound results as identified in TOR 3 for management purposes as identified in TOR 4 but with the caveat that this acceptance is reliant on improvements and recommendations being incorporated for the final stock assessments (i.e., response to TOR 5). These recommendations should be possible in the time available and likely would provide sufficiently sound scientific results to support management decisions. Also, the sequentially intertwined nature of the textual definitions for TORs (2), (3), and (4) made it challenging to comment on them individually for each species. Thus, I provided a “yes”, “yes with caveats”, “no” or “no with caveats” for consensus or majority responses from the panel for TORs (2) – (4) for each species. For TORs (2) – (4), the panel replied with a consensus “yes” or “yes with caveats” to thirteen species and a majority “yes” or “yes with caveats” to 15 species. Not all panelists replied with an explicit “yes” or “no” response to TOR 1 so these replies were inferred by the chair for this report. Detailed responses to TORs (2) – (4) for each species are included in the individual reviews.

Response for Aprion virescens to TOR (2), (3), (4): Consensus YES with caveats. The panel was presented with a revised assessment during the review based on local life history parameters generated by PIFSC scientists. The new unpublished life history data leads to a change in F/F30 from 1.08 to 1.57 (using a survivorship rate of 5%). Given the significant increase between the original and revised assessment over the threshold (of F/F30 = 1), the use of the unpublished data needs to be justified and reviewed. The depth range of this species exceeds the diver surveys. On the other hand, the commercial fishery may not be sampling the entire geographic range of the species. The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30.

Response for Caranx ignobilis to TOR (2), (3), (4): Majority YES with caveats. This species had local life history (i.e, NWHI) data, but not recently collected, so the author may want to explore the stepwise approach as an alternative. Recreational fishing for this species dominates the catch statistics and exhibits high inter-annual variability with a large standard deviation. The catch statistics are problematic for the basis of F/F30 and Catch30 generation. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for Caranx melampygus to TOR (2), (3), (4): Majority YES with caveats. This species had local life history (i.e, NWHI) data, but not recently collected, so the author may want to explore the stepwise approach as an alternative. Recreational fishing for this species dominates the catch statistics and exhibits high inter-annual variability with a large standard deviation. The catch statistics are problematic for the basis of F/F30 and Catch30 generation. Negative F is an issue with this species. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for Carangoides orthogrammus to TOR (2), (3), (4): Majority YES with caveats. This species had life history data generated by the stepwise approach. Recreational fishing for this species dominates the catch statistics but the fishery is small relative to other carangids. The catch statistics are problematic for the basis of F/F30 and Catch30 generation. Negative F is an issue with this species. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Seriola dumerili* to TOR (2), (3), (4): Majority YES with caveats. This species had non-local life history data, so the author may want to explore the stepwise approach as an alternative. Recreational fishing for this species dominates the catch statistics and exhibits high inter-annual variability with a large standard deviation. The catch statistics are problematic for the basis of F/F30 and Catch30 generation. Lbar estimates from catch statistics may be biased toward smaller fish since larger fish may not be kept due to ciguatera concerns. Furthermore, there may be taxonomic misidentifications of this species (with *S. rivoliana*) during the collection of recreational catch data that introduces additional problems. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Mulloidichthys pflueregi* to TOR (2), (3), (4): Majority YES with caveats. This species had life history data generated by the stepwise approach that incorporated the meta-analysis for Mullidae. Recreational fishing for this species dominates the catch statistics but the fishery is small relative to other goatfish. The catch statistics are problematic for the basis of F/F30 and Catch30 generation. Negative F is an issue with this species. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Acanthurus blochii* to TOR (2), (3), (4): Majority YES with caveats. This species had life history data generated by the stepwise approach. Commercial fishing for this species composes the majority of the catch statistics. The major concern for this assessment is that the F/F30 = 1.8 indicating that overfishing is occurring but the recommended catch target for sustainability is Catch30 = 37,000 Kg which is approximately six times greater than the current estimated catch of 6,411 Kg. This discrepancy severely undermines confidence in the result for this species assessment. There may be an issue with taxonomic mis-identification with *A. xanthopterus* in the catch estimates. The catch statistics are problematic for the basis of Catch30 and OFL generation. Negative F is an issue with this species. In general, long-lived species with determinant growth such as those of Family Acanthuridae present challenges for a length-based assessment approach. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Acanthurus dussumieri* to TOR (2), (3), (4): Consensus YES with caveats. This species had non-local life history data, so the author may want to explore the stepwise approach as an alternative. Commercial fishing and recreational fishing for this species contribute similarly to the catch statistics. Negative F is an issue with this species. In general, long-lived species with determinant growth such as those of Family Acanthuridae present challenges for a length-based assessment approach. The diver-survey derived Catch30s estimates are probably more reliable than those derived from catch estimates given survey sampling design and high number of observations although the survey domain does not cover the entire depth range of the species. The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30.

Response for *Naso brevirostris* to TOR (2), (3), (4): Majority YES with caveats. This species had non-local life history data, so the author may want to explore the stepwise approach as an alternative. There are no available recreational fishing or commercial fishing estimates of catch for this species. The calculation of Lc is simply a guestimate since no catch data are available. A declining temporal trend in Lbar suggests the equilibrium assumption of the approach may be violated. Negative F is a minor issue with this species. In general, long-lived species with determinant growth such as those of Family Acanthuridae present challenges for a length-based assessment approach. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, there is no species species alternative available for status determination.

Response for *Naso hexacanthus* to TOR (2), (3), (4): Majority YES with caveats. This species had non-local life history data, so the author may want to explore the stepwise approach as an alternative. Recreational fishing for this species composes the majority of the catch statistics. The trends in catch should be evaluated to determine reliability of the time series. Negative F is a minor issue with this species. In general, long-lived species with determinant growth such as those of Family Acanthuridae present challenges for a length-based assessment approach. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Naso lituratus* to TOR (2), (3), (4): Majority YES with caveats. This species had non-local life history data, so the author may want to explore the stepwise approach as an alternative. Recreational fishing for this species composes the majority of the catch statistics but total estimated catch is small relative to other Acanthurids. The major concern for this assessment is that the F/F30 = 1.4 indicating that overfishing is occurring but the recommended catch target for sustainability is Catch30 (from survey) = 61,700 Kg which is approximately twenty-three times greater than the current estimated catch of 2,693 Kg. This discrepancy severely undermines confidence in the result for this species assessment. There may be an issue with under-reported catch estimates or over-estimated population sizes from the survey methods. The catch statistics are problematic for the basis of Catch30 and OFL generation. Negative F is an issue with this species. In general, long-lived species with determinant growth such as those of Family Acanthuridae present challenges for a length-based assessment approach. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Naso unicornis* to TOR (2), (3), (4): Majority YES with caveats. This species had local life history data. Catch statistics for this species are not available individually (e.g., data presented in the table are aggregate values for three species). The determination of Lc is a guestimate without available catch data. The major concern for this assessment is that the F/F30 = 3.9 indicating that overfishing is occurring but the recommended catch target for sustainability is Catch30 (from survey) = 23,800 Kg which is greater than the current estimated aggregate catch for three surgeonfish species of 22,630 Kg. This discrepancy severely undermines confidence in the result for this species assessment. There may be an issue with under-reported catch estimates. In general, long-lived species with determinant growth such as those of Family Acanthuridae present challenges for a length-based assessment approach. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Monotaxis grandoculis* to TOR (2), (3), (4): Majority YES with caveats. This species had life history data generated by the stepwise approach. Commercial fishing for this species composes the majority of the catch statistics. The major concern for this assessment is that the F/F30 = 1.22 indicating that overfishing is occurring but the recommended catch target for sustainability is Catch30 (from survey) = 21,400 Kg which is approximately eight times greater than the current estimated catch of 2,755 Kg. This discrepancy severely undermines confidence in the result for this species assessment. There may be an issue with under-reported catch estimates. Negative F is an issue with this species. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for Myripristis berndti to TOR (2), (3), (4): Majority YES with caveats. This species had local life history data (but a non-local Lmat). There are no catch statistics available. There may be an issue with correct taxonomic identification during diver surveys and detectability bias since the species is mostly active at night and hides during the day. A declining temporal trend in Lbar suggests the equilibrium assumption of the approach may be violated. Negative F is an issue with this species. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Mulloidichthys flavolineatus* to TOR (2), (3), (4): Majority YES with caveats. This species had local life history data but longevity was estimated. Recreational catch dominates the estimated catch statistics. The major concern for this assessment is that the estimated population size from surveys (= 47,127 Kg) is significantly less than estimated total catch = 60,144 Kg. This discrepancy severely undermines confidence in the result for this species assessment. During the review, it was suggested that the catch of juvenile goatfish (o'ama) contributes to the high levels of estimated recreational catch. Furthermore, the population estimate from diver surveys may be an underestimate since goatfish utilize softbottom habitats adjacent to reefs and may also occur outside of the survey domain (i.e., coral reefs and hardbottom habitats). Negative F is an issue with this species. Conflicting temporal trends in Lbar from surveys versus catch suggests the equilibrium assumption of the approach may be violated. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Mulloidichthys vanicolensis* to TOR (2), (3), (4): Majority YES with caveats. This species had local life history data. The major concern for this assessment is that the estimated population size from surveys (= 38,393 Kg) is less than the Catch30 (from catch) = 51,200 Kg and only 1.5 times larger than estimated total catch = 24,174 Kg. This discrepancy severely undermines confidence in the result for this species assessment. The population estimate from diver surveys may be an underestimate since goatfish utilize softbottom habitats adjacent to reefs and may also occur outside of the survey domain (i.e., coral reefs and hardbottom habitats). Negative F is an issue with this species. Conflicting temporal trends in Lbar from surveys versus catch suggests the equilibrium assumption of the approach may be violated. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Parupeneus cyclostomus* to TOR (2), (3), (4): Majority YES with caveats. This species had life history data generated by the stepwise approach. The major concern for this assessment is that the F/F30 = 1.18 indicating that overfishing is occurring but the recommended catch target for sustainability is Catch30 (from survey) = 23,400 Kg is approximately five times greater than the current estimated catch of 4,392 Kg. This discrepancy severely undermines confidence in the result for this species assessment. Negative F is an issue with this species. A lack of stable temporal trends in Lbar from surveys and catch suggests the equilibrium assumption of the approach may be violated. Given the uncertainty of the data inputs, it is unclear if this approach is reasonable and produces results appropriate for status determination but might with sensitivity testing and justification. If this method is unsuitable for management purposes, the alternative method previously used (Sabater and Kleiber 2013, 2014) incorporates the same catch statistics which introduces similar concerns.

Response for *Parupeneus insularis* to TOR (2), (3), (4): Consensus YES with caveats. This species had life history data generated by the stepwise approach. Recreational catch represents the majority of estimated catch which is small relative to other goatfish species. Negative F is an issue with this species. The diver-survey derived Catch30s estimates are probably more reliable than those derived from catch estimates given survey sampling design and complete coverage of the depth range of the species. The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Parupeneus porphyreus* to TOR (2), (3), (4): Consensus YES with caveats. This species had local life history data. Recreational catch represents the majority of estimated catch which is small relative to other goatfish species. Negative F is a minor issue with this species. The Catch30 estimates from both data sources are in good agreement. The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Scarus rubroviolaceus* to TOR (2), (3), (4): Consensus YES with caveats. This species had local life history data. There are no catch statistics available at a species level. A fluctuating temporal trend in Lbar suggests the equilibrium assumption of the approach may be violated. Negative F is an issue with this species. The diver-survey derived Catch30s estimates are probably reliable given the survey sampling design and high number of observations. The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Scarus psittacus* to TOR (2), (3), (4): Consensus YES with caveats. This species had local life history data. There are no catch statistics available at a species level. Negative F is an issue with this species. The diver-survey derived Catch30s estimates are probably reliable given the survey sampling design but the number of observations is not high (n = 331). The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Scarus dubius* to TOR (2), (3), (4): Consensus YES with caveats.. This species had life history data generated by the stepwise approach. There are no catch statistics available at a species level. Negative F is an issue with this species. Population density is increasing while Lbar is decreasing from diver surveys. The diver-survey derived Catch30s estimates are probably reliable given the survey sampling design and complete overlap of domain with species depth range but the number of observations is barely acceptable (n = 148). The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Chlorurus perspicullatus* to TOR (2), (3), (4): Consensus YES with caveats.. This species had life history data generated by the stepwise approach. There are no catch statistics available at a species level. Negative F is an issue with this species. The diver-survey derived Catch30s estimates are probably reliable given the survey sampling design but the number of observations is barely acceptable (n = 147). The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Chlorurus spilurus* to TOR (2), (3), (4): Consensus YES with caveats. This species had non-local life history data. There are no catch statistics available at a species level. Negative F is an issue with this species. The diver-survey derived Catch30s estimates are probably reliable given the survey sampling design and the high number of observations (n = 757). The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Calotomus carolinus* to TOR (2), (3), (4): Consensus YES with caveats. This species had non-local life history data. There are no catch statistics available at a species level. Negative F is an issue with this species. The diver-survey derived Catch30s estimates are probably reliable given the survey sampling design and domain that includes most of the depth range of the species but the number of observations is barely acceptable (n = 135). The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Cephalopholis argus* to TOR (2), (3), (4): Consensus YES with caveats. This species had local life history data (but Lmat is not local). Recreational catch comprised the majority of the estimated catch. Negative F is an issue with this species. Given the concerns over ciguatera in this species, it is not surprising that catch estimates are low relative to the estimated population sizes. The diver-survey derived Catch30s estimates are probably reliable given the survey sampling design and domain that includes the depth range of the species and the high number of observations. The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Lutjanus fulvus* to TOR (2), (3), (4): Consensus YES with caveats. This species had life history data generated by the stepwise approach. Recreational catch dominated the estimated catch. Negative F is an issue with this species. The diver-survey derived Catch30s estimates are probably reliable given the survey sampling design and domain that includes the depth range of the species and the high number of observations. The application of the approach appears reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30 from survey.

Response for *Lutjanus kasmira* to TOR (2), (3), (4): Consensus YES with caveats. This species had local life history data. Commercial catch comprised the majority of the estimated catch. Negative F is an issue with this species. The diver-survey derived Catch30s estimates might be reliable given the survey sampling design and the high number of observations but the domain does not include the depth range of the species. Since this species is found from shallow reefs to deep slope habitats, it is unclear how best to integrate the diver survey and catch estimates to generate status determinations and Catch30s for management purposes. The application of the approach may be reasonable for this species given adequate justification and sensitivity testing for status determination (F/F30, SPR30) and Catch30.

5. ***Recommendations for improvement****: As needed, suggest recommendations for future improvements and research priorities. Indicate whether each recommendation should be addressed in the short/immediate term (2 months), mid-term (3-6 years), and long-term (5-10 years). Also indicate whether each recommendation is high priority (likely most affecting results and/or interpretation), mid priority, or low priority.*

Short/immediate term (2 months): The following recommendations from the panel represent actions necessary to provide scientifically sound results useful for management purposes from these stock assessments. I have previously included related comments on these recommendations in the TOR (2), (3), (4) general response section and species-specific responses. These recommendations are critical for the stock assessments to meet the requirements identified in TOR (2), (3), & (4) and should be viewed as high priority.

- Determine and use an appropriate method to handle negative fishing mortality (F) draws in the simulations. For the selected method, add a detailed description to the methods section of the Tech Memo. If the current approach of discarding negative F values for parameter estimation is used, thes discarded values should still be included in distributions presented in figures. For species with negative F as an issue, include discussion in the species assessment comments section.
- Use a survivorship rate of 4.35% or 4% as identified in Nadon et al. (2015) for the stock assessments, not the 5% utilized in the current assessments. If an alternative survivorship is used for a species, the rationale for the choice must be clearly articulated.
- For each species, provide figures or tables for a standardized format of catch time series from commercial and recreational fisheries catch data and diver survey length frequency distributions.
- Add content to the Tech Memo on the quality of the data sources incorporated into the stock assessments including fisheries-dependent commercial and non-commercial catch records and length frequency data, and the fisheries-independent diver survey data used for population abundance and biomass estimation that includes a discussion of potential issues such as detectability bias and mismatches between the sampling domain and species distributions.
- Add a narrative to the methods section that includes a description of the workflow with software programs to perform steps in decision trees. Be sure to include all elements such as programs, spreadsheets, and data files used as inputs to the stock assessments.
- For species with stock status above or near thresholds for management action (~ F/F30 = 1, SPR = 30%), perform sensitivity analyses to evaluate the influence of input life history parameter values on outputs for management decisions.
- Edit the probability of overfishing tables to the range of 0.10 – 0.50 by increments of 0.01.
- Edit terminology for "Cases" in decision table to a different word. Suggestions may include "Alternatives" or "Options".
- Add some representation of error around Lbar means in species-specific Figures.
- In general comments for each species of the Tech Memo, include descriptive text on the evaluation of equilibrium assumption and explicit descriptions of choices made at each Step in decision tree, and a discussion of sensitivity analyses undertaken, where appropriate.

- Don’t show the cumulative probability plots unless they are outputs from simulations
- For each species, don’t show combined biomass estimates and Catch30 from combined biomass estimates unless there is justification and support to combine data sources
- Provide clear advice on the reliability and bias of the biomass estimates (from catch and/or survey) and best alternative to aid decision makers in general comments

Mid-term (3-6 years):
- Identify and publish in peer-reviewed literature, an improved methodology to handle the negative Fs generated by the current method.
- Provide clear advice on the reliability and bias of the biomass estimates (from catch and/or survey) and best alternative to aid decision makers in general comments.
- Develop a data report that describes in detail the fisheries-independent diver survey sampling design, performance statistics, and population estimates (e.g., density, abundance, length composition).
- Determine detectability bias in diver survey method to improve the population estimates derived from the surveys.
- Identify the most appropriate area-based weights for geographic sector-level statistical approximation of population estimates since the current method may not accurately reflect relative amount of coral reef and hardbottom area among sectors.
- Develop and apply a quality rating for data sets used for biomass estimation and use the rating to decide on which data sets are appropriate for stock assessment and management purposes.
- Collect data to generate “local” life history information for fish species that use “non-local” studies or the stepwise approach for the MHI and NWHI.
- Explore methods to improve survey biomass estimates for fisheries-independent data for depth ranges beyond diver depths using advanced technologies such as U/W cameras.

Long-term (5-10 years):
- Collect data to generate “local” life history information for fish species that use “non-local” studies or the stepwise approach for the MHI and NWHI.
- Explore methods to improve survey biomass estimates for fisheries-independent data for depth ranges beyond diver depths using advanced technologies such as U/W cameras

Appendix 1: Bibliography of materials provided for review

Benchmark stock assessment for review (not to be distributed beyond reviewers):

Nadon, M. O. 2016. (draft) Stock assessment of the coral reef fishes of Hawaii, 2016. U.S. Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-XX, XX p.

Previous independent peer review reports:

Dichmont, C. 2015. Center for Independent Experts (CIE) Independent Peer Review of Length-Based Assessment Methods of Coral Reef Fish Stocks in Hawaii and Other U.S. Pacific Territories.

Pilling, G. 2015. Center for Independent Experts (CIE) Independent Peer Review, Report of: Length-based stock assessment methods for coral reef fish stocks in Hawaii and other U.S. Pacific territories.

Stokes, K. 2015. Report on the independent peer review of length-based stock assessment methods for coral reef fish stocks in Hawaii and other U.S. Pacific territories.

Relevant management information:

Western Pacific Regional Fishery Management Council. 2009. Fishery Ecosystem Plan of the Hawaii Archipelago. Sections 4.5 and 5.6 only.

Western Pacific Regional Fishery Management Council. 2011. Amendment 3 to the Fishery Ecosystem Plan of the Hawaii Archipelago.

References:

Kritzer et al. (2001) Characterizing fish populations: effects of sample size and population structure
on the precision of demographic parameter estimates. CJFAS 58: 1557-1568.

Nadon, M. O. et al. 2015. Length-based assessment of coral reef fish populations in the Main and Northwestern Hawaiian Islands. PLoS ONE e0133960.

Nadon, M. O. and Ault, J.S. (2016) A stepwise stochastic simulation approach to estimate life history parameters for data-poor fisheries. CJFAS (in-press).

<u>Supplemental Background Documents:</u>

Previous stock assessments:

National Oceanic and Atmospheric Administration. 2015. Environmental Assessment: Specification of Annual Catch Limits and Accountability Measures for Pacific Island Coral Reef Ecosystem Fisheries in Fishing Years 2015 through 2018. **(**RIN 0648-XD558)

Sabater, M. and Kleiber, P. (2014). Augmented catch-MSY approach to fishery management in coral-associated fisheries. In S.A. Bortone (Ed.), Interrelationships between Corals and Fisheries, CRC Press, Boca Raton, FL (2014), pp. 199–218 321 pgs.

Supplemental references:

Ault, J. S., J. A. Bohnsack, and G. A. Meester. 1998. A retrospective (1979-1996) multispecies assessment of coral reef fish stocks in the Florida Keys. Fishery Bulletin 96:395–414.

Ault, J. S., S. G. Smith, and J. A. Bohnsack. 2005. Evaluation of average length as an estimator of exploitation status for the Florida coral-reef fish community. ICES Journal of Marine Science 62:417–423.

Ault, J. S., S. G. Smith, J. Luo, M. E. Monaco, and R. S. Appeldoorn. 2008. Length-based assessment of sustainability benchmarks for coral reef fishes in Puerto Rico. Environmental

Conservation 35:221–231.
Beverton, R. J. H., and S. J. Holt. 1956. A review of methods for estimating mortality rates in exploited fish populations, with special reference to sources of bias in catch sampling. Rapports et proces-verbaux des reunions du Conseil International pour l'Exploration de la Mer 140:67–83.
Ehrhardt, N. M., and J. S. Ault. 1992. Analysis of two length-based mortality models applied to bounded catch length frequencies. Transactions of the American Fisheries Society 121:115–122.
Gedamke, T., and J. M. Hoenig. 2006. Estimating mortality from mean length data in nonequilibrium situations, with application to the assessment of goosefish. Transactions of the American Fisheries Society 135:476–487.

Appendix 2: Western Pacific Stock Assessment Review Benchmark Review Terms of Reference

This document serves as a Terms of Reference (TOR) for the Benchmark Review of the 2016 benchmark stock assessment of 28 species of reef-associated fish in the Main Hawaiian Islands, following guidelines established in the Western Pacific Stock Assessment Review (WPSAR) framework.

BACKGROUND

Section 301(a)(2) of the Magnuson-Stevens Fishery Conservation and Management Act (MSA) requires that fishery conservation and management measures be based upon the best scientific information available. MSA § 302(g)(1)(E) provides that the Secretary of Commerce (Secretary) and each regional fishery management council "may establish a peer review process for that Council for scientific information used to advise the Council about the conservation and management of a fishery." Consistent with this provision, the Western Pacific Regional Fishery Management Council (Council), NOAA's National Marine Fisheries Service (NMFS) Pacific Islands Fisheries Science Center (PIFSC) and the Pacific Islands Regional Office (PIRO) have established the WPSAR process. WPSAR is a cooperative effort to improve the quality, timeliness, objectivity, and integrity of stock assessments and other scientific information used in managing fishery resources in the Pacific Islands Region. The WPSAR process may be applied to scientific information used by the Council directly to fulfill its management mandate in the execution of the MSA.

The WPSAR framework document outlines the scope of WPSAR, defines roles and responsibilities, summarizes the various review levels, describes the sequencing and timing of the WPSAR process in coordination with the larger Council process, and provides mechanisms for resolving disputes. This framework is available from the WSPAR website, at: http://www.pifsc.noaa.gov/peer_reviews/wpsar/index.php.

PROJECT DESCRIPTION

PIFSC scientists are conducting stock assessments on exploited coral reef fish species in the Pacific Islands Region which are listed in the Council's Fishery Ecosystem Plans. These stocks are generally classified as data-poor due to a lack of reliable, long-term, catch and fishing effort data. Historically, the Council has set and NMFS has approved setting of annual catch limits (ACLs) using a percentile of median historical catch levels and more recently, a biomass- augmented catch-MSY method has been applied (Sabater and Kleiber 2014, NOAA 2015).

In an effort to use additional available data sources for these stocks, scientists at PIFSC have conducted new coral reef fish assessments using length composition data, abundance data from diver surveys, and certain key population demographic parameters related to growth, maturity, and longevity. PIFSC scientists have been implementing an approach that uses the average length in the exploited phase of the population (L_{bar}) to obtain an estimate of total and fishing mortality rates for coral reef fish stocks (Beverton & Holt 1956; Ehrhardt & Ault 1992). These rates, combined with population demographic parameters, are used in a numerical population model to obtain stock sustainability metrics (e.g., spawning potential ratio, F/F_{MSY}; see Ault et al. 1998, 2008). Overfishing limits can be generated by using recent total catch estimates and/or population size estimates from diver surveys. Furthermore, a novel meta-analytical approach using stochastic simulations was developed at PIFSC to obtain demographic parameter estimates for species with even less data than data-poor species ("data-less" species). These scientific methods recently underwent a rigorous independent review by a panel organized by the Center for Independent Experts, and have now been applied to individual species in the main Hawaiian Islands. There is a need to independently review these species-specific stock assessments prior to submission to a fishery management organization for consideration.

The format of reviewer-produced reports is attached in **Annex 1**. The Terms of Reference (TOR) questions for this peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

REVIEWER ROLES AND REQUIREMENTS

This Benchmark Review consists of an in-person panel of one review chair who is also a member of the Council's Scientific and Statistical Committee (SSC), plus 2 additional review members external to PIFSC, PIRO, and the Council and its affiliated bodies. The chair and review members shall have scientific expertise in data-poor stock assessment models and general fishery stock assessment methods. They will also have familiarity with requirements of fishery stock assessments under the Magnuson-Stevens Fishery Conservation and Management Act, and preferably will have familiarity with reef fish fisheries and/or life history.

The chair and review members have been asked to serve as independent and impartial scientific experts, and in their roles as reviewers they are not representing their respective institutions or affiliations. The chair and review members are expected to fulfill and comply with all elements specified in this TOR. The chair and review members are expected to review all required provided documents in advance of the meeting, actively contribute during the meeting and review further provided documents as needed, offer solutions with constructive criticism, and conduct themselves respectfully and professionally.

Review chair: The review chair shall facilitate the review to accomplish the stated goals and objectives articulated within this TOR. At the conclusion of the review, the chair will produce a report outlining **consensus** opinions from the review members addressing all aspects of this TOR especially as outlined in **Annex 2**, according to the review report format outlined in **Annex 1**. The chair will also present the consensus results of the review in-person to the Council's SSC after finalization of the reviewed benchmark stock assessment document. In cases where consensus cannot be reached on an individual TOR question, the review chair will describe the majority view and label the view as majority and not consensus. The review chair will also produce a second, independent review report indicating his or her scientific opinions addressing all aspects of this TOR especially as outlined in **Annex 2**, according to the review report format outlined in **Annex 1.**

Review members: Each review member will produce an independent review report indicating his or her scientific opinions addressing all aspects of this TOR especially as outlined in **Annex 2**, according to the review report format outlined in **Annex 1**.

The chair and review members will provide their respective consensus report and individual reports to the WPSAR Coordinating Committee point of contact after the close of the review, when the Coordinating Committee will check that reports satisfy the TOR and subsequently disseminate the reports. The reports will address all aspects of this TOR especially **Annex 2**, and follow the format as specified in **Annex 1**. The chair's consensus report, individual review member reports, as well as the reviewed final stock assessment document will be made available to the public on the WPSAR website shortly after they are finalized.

LOGISTICS

The WPSAR Coordinating Committee is responsible for setting up logistics of this review, including but not limited to travel arrangements, facility reservation and setup, security clearance in cases where reviews are held in federal facilities and/or where a reviewer is a foreign national, providing documents ahead of the review, and receiving and posting final review reports. The WPSAR Coordinating Committee point of contact for this review is [*insert name and email of CC lead for this review*].

This TOR may be modified by the WPSAR Coordinating Committee up to 1 month prior to the start of the review, but shall not be changed once the review has begun.

Timeline

This general timeline follows timeframes as outlined in the WPSAR framework.

Timeframe & date(s)	Description
2 weeks before review August 12, 2016	Documents distributed to chair and review members (generally via email)
Review August 29-Sep 2, 2016	In-person panel review
2 weeks after review Sep 19, 2016	Chair consensus report and individual review member reports submitted to WPSAR Coordinating Committee point of contact for a check on satisfaction of TOR. Coordinating Committee will then distribute and post accordingly
Following Council SSC meeting	Chair presents consensus opinions from review

ANNEX 1: Format of Chair's Consensus Report and Individual Reports

Reports should be in pdf format.

1. Each report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations addressing Annex 2 Terms of Reference questions.

2. The main body of the report shall consist of a Background, Description of the Chair's Role or Individual Reviewer's Role in the Review Activities, Summary of Findings for each TOR question (Annex 2) in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the TOR.

a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

b. *Review chair* should describe in a report the **consensus** views from the review members for each TOR question, and should not provide any non-consensus views which can be expressed in individual reports. In cases where consensus cannot be reached on an individual TOR question, the review chair will describe the majority view and label the view as majority and not consensus.

c. *Review chair and review members* should each describe in an individual report, his or her independent views on each TOR question even if these were consistent with those of other panelists, and especially where there were divergent views. The review chair will thus provide two separate reports, a consensus report and an individual report.

d. Each report shall be a stand-alone independent peer review report for others to understand the responses to TOR questions, and weaknesses and strengths of the science reviewed.

3. The reviewer report shall include the following appendices:

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

ANNEX 2: Terms of Reference Questions for Benchmark Review of Reef Fish in the Main Hawaiian Islands

For questions 1-4, reviewers shall provide a "yes" or "no" answer and will not provide an answer of "maybe". Only if necessary, caveats may be provided to these yes or no answers, but when provided they must be as specific as possible to provide direction and clarification. Examples for specific caveats include specific species names, life history types as defined by specific parameter values, and data or method decision points.

1. Review whether each of the following short-term recommendations from the previous independent peer review were addressed properly for the general (not species-specific) approach, considering that the data sources themselves are not up for review. If they have not been addressed, indicate why not and suggest methods for addressing them.
    a. The development of a clear decision chart to increase transparency in the application of the approach. Clearly articulate the hierarchical nature of the three life history approaches and under which circumstances a method should (or should not) be used.
    b. Explore an alternative to calculating mean length across islands and applying that in the remainder of the process by using the sectoral mean lengths (the primary index of exploitation) through to the estimate of fishing mortality, and weight resulting estimates to calculate overall fishing mortality.
    c. Examine the sensitivity of final results to uncertainty in the value of length at first capture (Lc) used. Ensure the method to calculate Lc is more standardized and repeatable by other assessors.
    d. When incorporating uncertainty in parameter estimates, evaluate the data underlying the coefficients of variation (CVs) derived from Kritzer et al. (2001), and compare them to those derived for species around the Main Hawaiian Islands and U.S. Pacific territories that can be estimated using e.g. the length-at-age bootstrapping approach for von Bertalanffy parameters: growth rate (K) and asymptotic length at which growth is zero (Linf).
    e. Draw maximum length (Lmax) for the data poor life history simulation approach from a distribution rather than using a single point prior to capture this element of uncertainty.
    f. Explore the impact of heavily truncated size data in which the sampled Lmax is not representative of the biological Lmax, what is a safe error in Lmax in terms of biases, false positives and negatives; and relate this to the decision tree.
    g. For some Main Hawaiian Island stocks the available Lmax values extended beyond the range of estimates from which the life history parameter relationships were developed for the corresponding family. Therefore, consider the efficacy of estimates and uncertainty developed where input parameters for a species require extrapolation outside the range of data on which the relationships were based.
    h. Research the possibility of using female biomass only for SPR calculations where the male proportion is considered important (e.g. in the case of protogyny; Ault et al. 2008).
    i. Explore the option of including runs with negative fishing mortality estimates within all calculations and representations.
    j. Present OFL distributions arising from all relevant data set combinations separately to ensure levels of uncertainty are understood.

2. Review the appropriateness of the application of the general approach to each individual species being assessed: Determine if decision points and input parameters were reasonably chosen, assumptions reasonably satisfied, and primary sources of uncertainty documented and presented.
3. Determine whether the final results for each individual species are scientifically sound, including estimated stock status in relation to the selected biological reference points ($SPR_{30\%}$) and overfishing limits.
4. Determine whether the results for individual species from question 3 can be used for management purposes under the Magnuson-Stevens Act and relevant Fishery Ecosystem Plan (FEP), including biological reference points such as MSY-based BMSY, FMSY, and MSY (or their proxies) with no or minor further analyses or changes, considering that the data itself and the general approach have been accepted for stock assessment purposes. If results of this analysis should not be applied for management purposes with or without minor further analyses, indicate which alternative set of existing results should be used to inform setting fishery catch limits instead and describe why.
5. As needed, suggest recommendations for future improvements and research priorities. Indicate whether each recommendation should be addressed in the short/immediate term (2 months), mid-term (3-6 years), and long-term (5-10 years). Also indicate whether each recommendation is high priority (likely most affecting results and/or interpretation), mid priority, or low priority.
6. Draft a report (individual report from Chair and review members, an additional consensus report from Chair) addressing the above TOR questions.

APPENDIX 3: Agenda, panel membership, presenter information, or other pertinent information from the panel review meeting.

Agenda:
The meeting will be held from August 29 - September 2 in Suite 1701 of the Finance Factors Building, 1164 Bishop St., Honolulu, Hawaii 96813.

The meeting schedule and agenda are as follows (8:30 am to 5:00 pm each day):

Monday, August 29, 2016
• Introductions (Dunlap, Yau)
• Background information (Yau)
• Objectives and Terms of Reference (Yau)
• Fishery Operation (Nadon)
• Management (Miyasaka; Dunlap)
• Presentation of stock assessments (Nadon)
Attendees: Choat, Franklin, Stokes, Nadon, Yau, Sabater, Dunlap, Lumsden, Boggs, Richards

Tuesday, August 30, 2016
• Presentation and review of stock assessments (Nadon and Panel)
Attendees: Choat, Franklin, Stokes, Nadon, Yau, Sabater, Dunlap, Lumsden, Boggs

Wednesday, August 31, 2016
• Continue review of stock assessments (Nadon and Panel)
Attendees: Choat, Franklin, Stokes, Nadon, Yau, Sabater, Dunlap, Lumsden, Boggs

Thursday, September 1, 2016
• Continue review of stock assessments (Nadon and Panel)
• Public comment period
• Panel discussion (closed)
Attendees: Choat, Franklin, Stokes, Nadon, Yau, Sabater, Dunlap, Lumsden, Boggs, Richards

Friday, September 2, 2016
• Panel discussions (morning, closed)
• Present results of review and recommendations (afternoon, open)
• Adjourn
Attendees: Choat, Franklin, Stokes, Nadon, Yau, Sabater, Dunlap, Lumsden, Boggs, Richards, Seki

---

Panel membership: Erik Franklin (Chair), Howard Choat, Kevin Stokes

Attendees: Marc Nadon (PIFSC)**,** Beth Lumsden (PIFSC)**,** Annie Yau (PIFSC)**,** Chris Boggs (PIFSC)**,** Ben Richards (PIFSC)**,** Mike Seki (PIFSC)**,** Alton Miyasaka (HDAR)**,** Matt Dunlap (NMFS), Marlowe Sabater (WPFMC)