INDEPENDENT PEER REVIEW OF THE 2017 GUAM REEF FISH STOCK ASSESSMENT FOR THE CENTER OF INDEPENDENT EXPERTS

Dr Cathy Dichmont

Cathy Dichmont Consulting
cathydichmont@gmail.com

Dichmont, C.M. 2018. Review of the 2018 Guam Reef Fish Stock Assessments. Report to the Center of Independent Experts. Cathy Dichmont Consulting.

Contact Details
Name: Dr Catherine Dichmont
Email: cathydichmont@gmail.com

Review of the 2018 Guam Reef Fish Stock Assessments

1 CONTENTS

1.2 List of Figures

1.3 List of Tables

1.4 Table of abbreviations

Abbreviation	**Description**
OFL	Overfishing Limit
ACL	Annual Catch Limit
SPR	Spawning Potential Ratio
F	Fishing mortality
F_{30}, C_{30}	Fishing mortality that resulted in a SPR of 30% and the catch derived from these
MSY	Maximum Sustainable Yield
GTG-LBSPR	Growth-Type-Group Length-Based Spawning Potential Ratio model
MSE	Management Strategy Evaluation
LH	Life History parameters
C75	75 percentile of catch approach to setting the OFL
ToR	Terms of Reference
BAC-MSY	Biomass-Augmented Catch Maximum Sustainable Yield model
HI	Hawaii Island
BC	Base Case

2 Executive Summary

The review workshop for the Guam reef fishes took place in Honolulu, Hawaii on 6 to 9 February 2018 with review panel members Drs Franklin (chair), Dichmont, and Powers, the Guam reef fish assessor, Dr Marc Nadon, and other scientists involved in the stock assessment, data collection and management. The review documents had been placed on a Google Drive folder with several additional sensitivity test outputs during the review. Several very insightful presentations were provided during the review, which often resulted in very open and free flowing discussions. These greatly contributed to the reviewers' knowledge base.

The assessment approach used is rigorous and well implemented. The package elegantly brings together different options of calculating fishing mortality (F), Spawning Potential Ratio (SPR) and overfishing status (F/Fmsy – in this case the F_{30} value is used as a proxy for Fmsy) based on the available data. It also used two approaches for calculating the Overfishing Limit (OFL), using total catch or an assumed total abundance index from an independent survey. The strengths and weaknesses of the approach used is discussed per species. In some cases, sensitivity tests were provided. The package as a whole is cohesive and comprehensive. It also becomes more uncertain as the level of life history knowledge decreases, which is a positive property of the method. Thus, the approach is supported. It should be noted that model applicability needs to be reviewed on a case by case basis.

Although Terms of Reference (ToR) 2 states that the panel should consider "that the data itself and the general approach have been accepted for stock assessment purposes", no evidence of this was provided before or during the review. A review of the creel survey and of the data managed by Pacific Islands Fisheries Science Center (PIFSC) was undertaken in 2012 and 2013 respectively, but these were mostly prior to the data being used in the assessment. Although much has been done to enhance the creel and data collection program, evidence was provided during the review that the creel survey, especially for spearfishing gear, may be biased and not always able to be used to scale the assessment to absolute abundance values. Similarly, the independent survey has depth restrictions (30m) and most of the species in this assessment have maximum depths deeper than 30m, yet the survey was being used as an absolute index of abundance. This meant that these sources of bias and uncertainty could not be ignored for this review. It is therefore the highest medium-term priority that the creel and independent survey is externally reviewed.

Of the 19 species, it was determined that the approach could be used for 12 species to determine overfishing status. However, only five of these were able to be used to generate the OFL and C_{30} values. It should be noted that this is mostly not due to the method itself, but that the data generally could not conform to the assumption that they can be applied as an absolute and unbiased index of abundance.

The findings per species therefore show:

- that the approach is able to provide overfishing status for most species despite uncertainty in life history parameters.
- That the approach relies heavily on the assumption that either catch or the independent survey provides an index of absolute abundance. In most cases, this assumption does not hold true given the available data and issues therein.

Two alternative OFL setting approaches had been used in the past. The first was that of Sabater and Kleiber (2013) which used a Biomass Augmented Catch Maximum Sustainable Yield (BAC-MSY) model. Prior to the BAC-MSY approach, the OFL and ACL was based on the 75th percentile of the catch (here called the C75 approach). Despite 75% of the average catch being used elsewhere in the world, a difference when this approach was applied in Guam was to continually extend the catch series to include new data. The ratcheting effect of continuously increasing the data range is very real, and therefore this approach would not be recommended unless the time period is fixed. It should be noted that lack of detailed documentation on these approaches did hamper the review process.

The ideal would have been that the 2017 assessment used both the biosampling and creel survey data to obtain catch, at least as a test. As is generally found with data limited species, undertaking alternative approaches such as updating the BAC-MSY would also have been beneficial. However, these were not available for review. Therefore, given the large uncertainties in the single species assessment and the ratcheting effect of the C75 approach, for those species where the reviewers' assessment was "No" for OFL setting, the BAC-MSY approach is favoured. Ideally, the approach would work best if life history is considered in the lumping together of remaining species.

Several medium-term model alternatives and data foci recommendations are provided in the report. An important recommendation is to undertake a Management Strategy Evaluation that could more closely investigate the many trade-offs in the decision process of undertaking the Guam assessments, including options of when aggregation of the data would be more beneficial than undertaking a single species assessment.

4 BACKGROUND

This review addresses the data limited assessment of Coral Reef fishes in Guam. The assessment approach is to link a series of data limited methods to: calculate life history (LH) parameters and natural mortality (M) for input into an updated version of the length-based model of Hordyk et al. (2016) to calculate the current fishing mortality rates (F), Spawning Potential Ratio (SPR) and the sustainable fishing mortality rate, defined as the fishing mortality that resulted in a SPR of 30% (F_{30}). After calculating the current biomass either directly through diver surveys or indirectly using the total catch, the catch limit C_{30} is derived.

Several approaches can be used to derive M and LH parameters, depending on the information available per species: a) if local or relevant external LH parameters are available, then use these and used the Ault et al. (1996) approach to calculate M if required; b) if none/some of these are available, then apply the step-wise approach of Nadon and Ault (2016) to obtain these.

The mean length model to calculate SPR in the Hawaii assessment had been replaced by the Growth-Type-Group Length-Based Spawning Potential Ratio model (GTG-LBSPR) model that has been updated to include "Lee's Phenomenon" (which comes from the differential fishing mortality rates across fish of the same age if selectivity is size selective, as is the case in Guam reef fishes). Instead of using mean length, the length composition is used in this model. The SPR and overfishing limit (F_{30} as the proxy for Fmsy) is an output of this model.

To scale fishing mortality up to biomass units, two possible approaches can be used: using diver survey data or using the total catch. The diver survey data is restricted to a 30m depth so species that occur deeper or are not seen in the survey are not well captured by this approach. The alternative is to scale the F_{30} value with the total catch to calculate the corresponding catch, C_{30}. This approach relies on good total catch figures from the creel surveys for the most recent years and is therefore less reliable if the species in not well represented in creel surveys.

6 Description of the Individual Reviewer's Role in the Review Activities

The review workshop for the Guam reef fishes took place in Honolulu on 6 to 9 February 2018 with review panel members Drs Dichmont, Powers and Franklin (chair), the Guam reef fish assessor, Dr Marc Nadon, and other scientists involved in the stock assessment, data collection and management. The key assessment and background documents were provided in a Google Drive folder (Appendix 1). The Statement of Work and Agenda provided to the review panel are included in Appendix 2. Participants of the workshop are provided in Appendix 3.

Several presentations were provided during the review, that resulted in very open and free flowing discussions. These greatly contributed to our understanding of the methods. Past review issues were also presented and discussed.

During the review, some additional work was requested and provided. The interaction between the review team and the review attendees was extremely informative and helpful. Thanks to those that undertook additional tasks during the review.

A panel report was required, and panel views were discussed during the review. An individual reviewers report was provided addressing each Terms of Reference (ToR), being:

1. For each individual species, review the application of the general approach for each of the following calculations. For each calculation, consider decisions points, input parameters, assumptions, and primary sources of uncertainty.
    a. Fishing mortality (F), spawning potential ratio (SPR), and corresponding overfishing limit (F at SPR=30%, aka F_{30}).
    b. Generation of overfishing limit from C_{30} (catch levels corresponding to F_{30}) distribution calculation.
2. Determine whether the results for individual species from question 1 can be used for management purposes under the Magnuson-Stevens Act and relevant Fishery Ecosystem Plan (FEP) 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.
3. 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 (3-5 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.
4. Draft a report (individual report from Chair and review members, and additional consensus report from Chair) addressing the above TOR questions.

7 SUMMARY OF FINDINGS FOR EACH TOR IN WHICH THE WEAKNESSES AND STRENGTHS ARE DESCRIBED

7.1 TOR 1: FOR EACH INDIVIDUAL SPECIES, REVIEW THE APPLICATION OF THE GENERAL APPROACH FOR EACH OF THE FOLLOWING CALCULATIONS. FOR EACH CALCULATION, CONSIDER DECISIONS POINTS, INPUT PARAMETERS, ASSUMPTIONS, AND PRIMARY SOURCES OF UNCERTAINTY: A) FISHING MORTALITY (F), SPAWNING POTENTIAL RATIO (SPR), AND CORRESPONDING OVERFISHING LIMIT (F AT SPR=30%, AKA F30); B) GENERATION OF OVERFISHING LIMIT FROM C30 (CATCH LEVELS CORRESPONDING TO F30) DISTRIBUTION CALCULATION

7.1.1 Overall response

The assessment approach used is rigorous and well implemented. The package elegantly combines different options of calculating fishing mortality (F), Spawning Potential Ratio (SPR) and overfishing status (F/Fmsy – in this case the F_{30} value is used as a proxy for Fmsy) based on the available data. It also uses two approaches for calculating the Overfishing Limit (OFL), using total catch or an assumed total abundance index from an independent survey. The strengths and weaknesses of the approach used are discussed per species. In some cases, sensitivity tests were provided. The package is cohesive and comprehensive. It also becomes more uncertain as the level of life history knowledge decreases, which is a strong positive characteristic of the method. Thus, the approach is supported. It should be noted that model applicability needs to be reviewed on a species by species basis.

Although Terms of Reference 2 state that the panel should consider "that the data itself and the general approach have been accepted for stock assessment purposes", no evidence of this was provided before or during the review. A review of the creel survey and of the data managed by Pacific Islands Fisheries Science Center (PIFSC) was undertaken in 2012 and 2013 respectively, but these were mostly prior to the data being used in the assessment. Although much has been done to enhance the creel and data collection program, evidence was provided during the review that the creel survey, especially for spearfishing gear, may be biased and not always able to be used to scale the assessment to absolute abundance values. Similarly, the independent survey has depth restrictions (30m) and most of the species in this assessment have maximum depth distributions deeper than 30m, yet the survey was being used as an absolute index of abundance. This meant that these sources of bias and uncertainty could not be ignored in this review. It is therefore the highest medium-term priority that the creel and independent survey are externally reviewed.

7.1.2 Background

7.1.2.1 The general approach

The numerical model GTG-LBSPR was applied to calculate F, F_{30} and SPR using input parameters obtained from life history parameters and size composition data, as well as their associated uncertainty. Although the GTG-LBSPR model is fairly recently published, its original version is well known, and the analytical approach is common in using the von Bertallanfy growth function as its basis and length based per recruit theory at its core. The GTG version stands for an approach that addressed Lee's phenomenon where selectivity is size based. For each species, the best available information is used. The GTG-LBSPR model requires specific life history parameters, a size frequency distribution, and their associated uncertainty.

A series of possible approaches to obtain life history parameters and their associated uncertainty are available drawing from a large and published body of work. These methods range from (the most preferred approach) using the raw data from locally obtained life history parameters, to using a family level stepwise approach based on a local maximum length value. The decisions with respect to which parameters and approach to take should be better articulated.

Medium term recommendation 1. Describe in detail in a guideline format the full range of decisions required on, for example, which input parameters to use, when a sensitivity test is required, how associated uncertainty is estimated and when to use the stepwise approach.

Size frequency data are mainly available from biosampling of a commercial co-operative. Creel survey data are also available. Several decisions are required to choose the most appropriate size frequency data and these need to be better articulated, but the approach is supported in general. The utility of each of these data sources differs per species and is discussed below.

The approach is not appropriate if selectivity is dome shaped or if it is the case that the population is not in equilibrium, because of variable recruitments and natural mortality. This meant that only 19 of the 24 species for assessments were used in this application. This original screening of the data is supported, but should be explicitly included as a decision in the flowchart and included in the report.

The approach to calculating F, F_{30} and SPR is supported as a whole.

Short term recommendation 1. Include in the report the reasons for excluding the five species for which the approach was not applied.

Medium term recommendation 2. Better describe in a guideline format the reasons for choices of which size frequency data to use, and also when dome shaped selectivity is evident.

The OFL calculation requires that the F_{30}, a relative measure, is scaled to biomass using some form of abundance data that provide information about absolute abundance. Two sources are usually available per species – using the total catch and the biomass derived from the diver survey. For the assessment report, in most cases the biomass and catch related management measures are provided for both the survey and the catch methods rather than using a model averaging approach. As per previous reviews, it is supported that model averaging of the two options is *not* undertaken.

The general approach is described in a series of published papers, a short methods section and a flowchart. This is insufficient for management use given the series of decisions that are actually applied to each of the multiple decision points. This approach therefore needs a major methods section, an updated flowchart, but because the flowchart can never fully describe the full complexity of the decisions, it is recommended that a Guideline on how to apply the modelling approach to a species is produced.

Medium term recommendation 3. Fully describe the methods in the assessment document, update the flowchart to describe more of the steps in detail and add these to a Guideline document that provides guidance on how to apply the modelling approach to a species (High).

In this set of Guidelines there were a few decision point defaults which are recommended to be changed as described below.

Medium term recommendation 4. Change the present defaults in the method to (this is put as a medium-term recommendation even though this change affected which Base Cases were to be finally recommended): a) The maximum age used for a species was generally the maximum of that recorded within a reasonable set of related studies. It is recommended that this be changed to a default of using the Guam life history parameters. This includes the longevity input value, unless the implied maximum length or L_∞ is close to or below the mode of a local size frequency distribution or if the sample size of Guam longevity sample is very small; b) A set recent period of length frequency was used unless the sample size was very small when a longer period was taken. This decision was inconsistently applied on a case by case basis and it would be better to have a consistent rule with stated reasons for exceptions. One of these is to expand the size frequency until more than 100 animals are selected. Automatically expanding to a 10-year period, say, may be more appropriate so that each species does not have a different period. c) Similarly, use a longer series for the L_{99} calculation when the stepwise approach is used; d) Rather than providing standard deviation values for the derived parameters, provide percentiles. Given the skewness of most of the output statistics, useful values would be ones that describe the 90th and 50th percentile confidence intervals; e) In terms of the data, it is recommended that the largest catch measure is always used. For example, when the biosampling catch is greater than the boat-based catch, this should be used and flagged for review. A sensitivity test using the official creel data could also be provided. Similarly, the Fisheries Receipt Book program could also be used as a check on the commercial catch.

Model behaviour was not clearly articulated. For example, the likelihood profile of a single run was not provided for investigation. This means it is unclear how well the model fits the original data (not bootstrapped). The distributions are essentially the estimated value, excluding their uncertainty.

Medium term recommendation 5. Show a likelihood profile of the original data and parameter input values for a series of species, testing the influence of data and input parameter assumptions.

The approach to calculate the minimum length associated with C_{30} is calculated. There is an error in how this is calculated or displayed since there is often a separate and much lower peak in the Lc30 distribution. If this is not an artefact, then it should be explained.

Short term recommendation 2. Investigate the calculation of the Lc30 distribution.

7.1.2.2 Data and associated assumptions of its use in the assessment

Although part of ToR 2, the following statement is best discussed here: "…(considering that the data itself and the general approach have been accepted for stock assessment purposes)". Throughout the review, it was an issue that the panel and this reviewer grappled with. Although it was stated that the data should be taken as a reviewed series of data, this was not possible for the following reasons:

- No evidence was provided (beyond that provided during the review workshop) that the data used in this assessment had been reviewed – the previous reviews provided during the review cited related to data mostly gathered prior to the data period used in the Review.
- A document provided during the review workshop (Bak, 2012) that reviewed the creel survey in 2012 queried the use of the data for stock assessment purposes, yet this is how the data are being used. As a partial response, during the review workshop, written explanations were provided on what actions were taken post the creel survey review.

- Despite these actions, the creel survey derived catch for several species was less than that obtained in the biosampling i.e. the actual animals used for biological sampling obtained from mainly spearfishing gear through a single commercial operation is more than the creel derived catch that should capture recreational/artisanal and commercial catches from all gear types. This means that there was evidence provided that the creel survey catch is at least an underestimate of the spearfishing catch. Yet, these data are used in the stock assessment as an absolute abundance scalar.
- Verbal statements were also provided during the Review that participation rates by users of spearfishing gear is sporadic, and that the expansion to total catch from interview catch rates by species is susceptible to large fluctuations due to occasional small sample size (as shown in many of the species' catch series).
- Discussions during the workshop also drew into question the independent survey at the individual species level. Apart from the *clearly stated bias* in the survey for species that have ranges deeper than the survey's 30m, there was also discussion during the workshop that there may be some issues with the index as sample size and foci (e.g. putting more sample effort in Marine Protected Areas which is not treated as a separate stratum in the analyses) have changed over time and potentially affected the index over time. In the case of several species, there seem to be an indication that there is also a mean-variance relationship.

This means that this reviewer considered the above issues with the data in terms of how these affect the assumptions made within the stock assessment. There are:

- Whether the survey can be treated as an absolute index of abundance. Even for the few species where there is a range overlap with the survey, this assumption was generally not accepted. There are few surveys in the world where this would be true, and this survey was not designed to be an absolute index of abundance. The departure from this assumption increases as the depth range of the species increases or the species is not properly seen within the survey. These are discussed on a case by case basis in the individual species report. On the other hand, for many of the 19 species, the survey may have been a good index of relative abundance, but the assessment approach did not use this index in this way. Several potential approaches open up for use are discussed in ToR 2.
- Whether the catch figure produced is an accurate reflection of the total catch. This is also discussed on a case by case basis, but in general this is less likely for spearfishing dominated species' catch given that often the biosampling total catch is more than the creel survey derived catch.

The above assumptions are not ignored in the assessment report and are openly discussed. The enormous effort that has been put into collecting these data is also acknowledged.

The data and how they are derived are not described in detail. This makes it difficult for the reviewer as this means that the data and the methodology to derive the input data for the model is unclear.

Medium term recommendation 6. Fully describe the data and how these are obtained (High)

It is therefore urgent that the creel survey is independently reviewed. Given that this requires a specific skill set, it may be possible to undertake this review for a few regions if resourcing is an issue.

Medium term recommendation 7. Undertake a review of the creel survey for use in the stock assessment (High)

Similarly, the independent survey may need some filtering and re-analysis, and the design reviewed so that it can be used in the stock assessment. Evidence for this is that the survey sample size has changed over time and that some surveys added more sites to Marine Protected Areas, which are not treated as a separate stratum.

Furthermore, the depth range of the survey is restricted to 30m since they are undertaken by divers. Some attempt at calibrating the survey for the assessed species should be investigated. Furthermore, it may be beneficial to investigate the use of alternative gears, such as stereo cameras, to undertake deeper surveys.

Medium term recommendation 8. Undertake a review of the independent survey to provide a relative index of abundance. The review should consider past indices and whether these need some form of filtering and re-analysis for the indices to be better applicable to stock assessment. It should also investigate the utility and practicality of deeper survey approaches or calibration options. (High)

Globally, obtaining total catch and effort figures from a multi-gear highly dispersed set of fisheries that also contain recreational, artisanal and commercial sectors is very complex. It is appreciated that obtaining catch and effort figures are difficult and gaps are likely to emerge despite large effort in obtaining the data. This seems to be the case in Guam where the creel program is extensive. Allied approaches of obtaining these data from certain sectors may therefore be required. For example, the low participation rate of spearfishermen may be difficult to overcome in the short-term. The existing project to interact with spearfishing clubs is supported. This issue is discussed in Griffiths *et al.* (2010) and this proposes approaches, such as respondent-driven sampling used widely in the immunology and disease fields.

Medium term recommendation 9. The present study of working with spearfishing clubs to obtain more information about the spearfishing catch and effort is important (High)

Medium term recommendation 10. In support of the work with spearfishing clubs, investigate allied methods of obtaining catch and effort for, particularly, the spearfishing component of the fishery. Model based approaches such as Respondent Driven Sampling may be useful here. (High)

In the creel survey, often a large component of the catch fell within the category "Assorted reef fish". In the approach used in this assessment to calculate the standard deviation in the catch per species, this category was included in the species catch frequency distribution. However, technically this means there could be reef fish catch in this category and not included in the catch for that species. One approach in the short term would be to exclude the assorted catch category when the species level catch distribution is being used in the expansion and bootstrapping. This assumes that the assorted species catch distribution is the same as the recalculated catch distribution. Special attention would be required to make sure this assumption only applied to reef fish may be required.

Medium term recommendation 11. When calculating the standard deviation of the catch, recalculate the species catch distribution by excluding the assorted categories such as assorted reef fish. Special attention would be required to make sure this assumption only applied to reef fish may be required.

However, ideally when the creel survey has been reviewed, a better approach would be to bootstrap the entire creel survey expansion to total catch process to obtain the catch distribution by species. The assorted species categories could then be addressed more directly.

Medium term recommendation 12. Bootstrap the entire creel survey expansion to total catch process to obtain the catch distribution by species. The assorted species categories could then be addressed more directly.

7.1.2.3 Life history parameters

Several recent studies have been undertaken in Guam and these have increased the knowledge base of the life history parameters used in the assessment. This effort is commended. However, there are still species for which life history parameters were missing or derived from studies undertaken in several regions. The latter is incongruous, given the known correlations in life history parameters, so using information from multiple regions increases the risk of a mismatch in the parameters. In general, using local life history parameters was preferred as recommended above.

An unpublished study on the effects of temperature on longevity was provided during the review. However, it was unclear whether it fully considered sample size and its influence on longevity. This work is encouraged and, once published, would need to be considered in the approach.

Medium term recommendation 13. An unpublished study on the effects of temperature on longevity was provided during the review. However, it was unclear whether it fully considered sample size and its influence on longevity. This work is encouraged and, once published, would need to be considered in the approach.

In general, using local parameters is likely to be further supported by this study.

Short term recommendation 3. In the short term, update some key life history parameters for *Scarus forsteni*, *S. rubroviolaceus* and *S. schlegeli* as described in the relevant species section below. These could be provided from the biosampling program (High).

Medium term recommendation 14. Continue to fill the gap to obtain local life history parameters, especially longevity and growth.

The Kritzer (2001) study is used to obtain the standard deviation on longevity if this is not available. However, another issue with regard to longevity is whether sample size biases the value obtained. This is an issue, as often local longevity parameters were available, but from a small sample size. This meant that values from other regions were used. A simulation study similar to Kritzer is recommended which investigates bias in longevity with respect to sample size and the resultant F and SPR calculations.

Medium term recommendation 15. Undertaking a simulation study similar to Kritzer is recommended, which investigates bias in longevity with respect to sample size and the resultant F and SPR calculations.

7.1.2.4 Displaying and discussing the results

The assessment results were comprehensive and well laid out. However, a few improvements are suggested:

- Display the size frequency plots to illustrate what the size frequency at the SPR 30% would look like.
- Add the sample size for the size frequency plots.
- Add region in the life history tables together with the provided references.

- Although the number of iterations that fell below 0.3 are provided, stock status is not being determined using this approach. Therefore, the related iterations of $F/F_{30} < 1$ would be useful despite being implied by the former values.
- The descriptions could benefit from a more rigorous set of sub-headings that could be described for each species. Examples of these are in this review document in the species section as well as the Excel sheet created during the workshop.
- Given the skewness of the distributions, the standard deviation is less useful than percentile-based measures such as those that describe the 50th and 80th percent confidence intervals i.e. 10th, 25th, 50th (median provided in report), 75th and 90th percentiles.
- Present size distribution over time unless sample size is too small. In these cases, a lumped early and late distribution could be provided as requested during the review.
- Always provide the full set of key statistics for sensitivity tests.

Medium term recommendation 16. Modify the results section to include the SPR=30% size frequency, size frequency sample size, life history region, number of iterations of F/F_{30} to fall below 1, description sub-headings, 50th and 80th confidence intervals related percentiles, size distributions over time.

7.1.3 Review approach

For this review given the large number of species to cover, a general approach was developed so that each species can, *at least initially*, be reviewed with some systematically applied criteria:

1. The first step was undertaken during the review workshop in session. Comments per species or information were recorded, being sample size of lengths to create the length frequency distribution, the life history method (local, external or stepwise), any life history sensitivity runs, concerns with the diver surveys, concerns with the catch data, diver biomass relative to the catch biomass, depth range of species divided by the survey maximum depth of 30m, and additional notes and comments.
2. These and the original results were distilled into a scoring system. **It should be noted that this system was more to consolidate the information and highlight uncertainties than as a system that would be combined into an overall score.** The following categories were used to score the Base Case of each species (scored from 1 (bad) to 3 (good)):
    a. Catch – Score 1. Poor representation of major catch source (e.g. spearfishing is dominant gear; biosampling >creel or sporadic catch); Score 2. Catch more representative but variable; Score 3. Good.
    b. Survey – Score 1. Deeper range, migratory or not commonly seen in survey, or highly variable; Score 2. middle range of 1, less variable; Score 3. Good depth overlap, easily identified and commonly seen in survey.
    c. Life history – Score 1. Stepwise with low basis for L_{99} or family level information; Score 2. Stepwise or life history mostly from other areas; Score 3. Local parameters and well sampled.
    d. Model fit to size frequency data – Score 1. Large residual pattern; Score 2. Moderate residual pattern; Score 3. Well fitted.

e. Model plausibility – Not scored, but the total catch relative to the biomass values, their standard deviations and distributions were described. A question mark over model plausibility was highlighted a) if one or both of the biomass values were smaller than the total catch, and b) if the CVs were greater than 100% and this was not only explained by a very large tail of few individual bootstrap runs in the distribution.

f. Model sensitivity: F and SPR – Score 1. Highly sensitive to plausible tests i.e. no easy way to choose between very different test results; Score 2. Sensitive, but able to choose which test is most applicable; Score 3. Not sensitive (in some cases no tests were done and therefore not scored).

g. Model sensitivity: Biomass and OFL – Score 1. Highly sensitive to plausible tests i.e. no easy way to choose between very different test results; Score 2. Sensitive, but able to choose which test is most applicable; Score 3. Not sensitive (in some cases no tests were done and therefore not scored).

2. A final step to determine whether a species' results could be used for each component of the ToR was to review results that indicate biases or that the results are too uncertain. The following were an indication of bias:

   a. A plausible sensitivity test that changes the overfishing status compared to the BC. The shape of the distribution could modify this point if the distribution is highly skewed.

   b. The absolute biomass estimate is less than the total catch. This rule was applied to each biomass result independently i.e. catch or biomass derived.

   c. The depth range of the species is much deeper than 30m. This pointed to potential bias in the survey derived biomass.

   d. If the biosampling catch was more than the total catch, this was taken as an indication of potential bias in the catch derived biomass result. The distribution of catch derived measures was considered here.

Indications of large uncertainty were:

e. A sensitivity test that is plausible, but that is very different to the BC and increased uncertainty.

f. Highly variable or sporadic catch (for the catch derived measures and equilibrium assumption).

g. Highly uncertain or annually variable survey index (for the survey derived measures).

h. The coefficient of Variation (CV) of the overfishing status is greater than 100% and distribution is broad (i.e. the high median is not because of a narrow but long tail in the distribution.

i. The CV of the OFL and/or the biomass measures are greater than 100% and distribution is broad (i.e. the high median is not because of a narrow but long tail in the distribution).

7.1.4 Per species response

Below is the response per species to the above criteria. Unusually, for clear articulation, the summary response for the question in ToR 2 – whether the results could be used for management – is provided at the end of each species' section so that the full argument is placed together. Whether a species' assessment can be used for management has, just as for ToR 1, been divided into a) whether it can be used for determining whether overfishing status (i.e. overfishing or not overfishing) and b) whether it can be used for setting the OFL and to determine the uncertainty.

7.1.4.1 Naso unicornis

Catch: Large drop in the catch from early decades. Data affected by peaks in net fishing catch which are likely juvenile runs. BC Score 2.

Survey: Diver biomass limited by depth, likely underestimate. BC Score 2.

Life history: Local information used for all parameters except for the longevity input value where Hawaii Island (HI) information is used. However, two factors mean a change to the Base Case is recommended. The first is that the general rule of using the largest longevity is not supported (as discussed in the above general section). The Guam sample size is large and unpublished work shows that there is a potential temperature effect on longevity for this species. An alternative sensitivity run was requested. Initial BC score 2, new BC score 3.

Model fit to size frequency: The length frequency data show two distinct size frequency distributions where the 2012 to 2014 data are being heavily weighted towards larger animals. The large animals were not present in the 2010 and 2011 data and disappear in 2015 and 2016. If the data are accurate, then this implies that the population is not in equilibrium (a basic assumption of the method) as animals entered the population. As an alternative, this could also be a sign that the data are not accurate, which seems more likely given how suddenly the size frequency changes on either end of the 2012-2014 data. A sensitivity run using the 2012 to 2014 data was requested. The model fits the BC and recommended Base Case frequencies well. BC Score 3. Recommended Base Case Score 3.

Model plausibility: BC – Total catch larger than survey biomass. Given the depth range of the species (247m), this is expected. The total catch is less than the biomass derived from the catch, but the CV of these values is very uncertain. If the distributions of catch, C_{30} and biomass are investigated, this is not due to a large tail in the outputs. Doubling the median catch value means this figure falls well within the bulk of the distribution. The above statement is also true for the recommended BC. It should be noted that this large variance is not due to using the stepwise approach as local life history parameters are available.

Model sensitivity (SPR and Overfishing status): Four runs were undertaken. The runs were done using original BC (longevity 50 years and size data from 2014-2017) and combinations of changing one of the longevity assumptions to the Guam value of 23 years or using size data from 2012-2014.

BC score and test about longevity: 2 – although the Overfishing status is changed from the BC, it is possible to choose between the two runs in that the Guam values have a large sample size as discussed above.

Regarding the 2012-2014 size data: Test run score with respect to initial BC is 2. As expected, the model is sensitive to the larger size frequencies, but does not change the F status (i.e. for the same longevity, the overfishing status does not change, although the new recommended BC median value is close to 1).

Short term recommendation 4. **Change the *Naso unicornis* Base Case to use the Guam longevity value.**

Medium term recommendation 17. **Investigate the size frequency changes over time in terms of data accuracy or whether there was a population change.**

Model sensitivity: Biomass and OFL: The model is sensitive to the change in size frequency and less so to the change in longevity. But the findings of model plausibility above remain true for all tests.

Summary for ToR 2: (a) Yes for the 23-year longevity Base Case; (b) No for OFL based on the results from the model plausibility test

7.1.4.2 Carangoides orthogrammus

Catch: Highly variable catch. Mostly line fishing except for peaks in juvenile net fishing. BC Score 2.

Survey: Not applicable.

Life history: Used stepwise approach. L_{99} from local creel samples, although larger HI values could have been used. However, the HI L_{99} is larger than any animals in the size frequency distribution. Mean lengths have been stable since 1985. BC Score 2.

Model fit to size frequency: Hint of a bimodal distribution means the fit can only be moderate given model assumptions. When the 2009-2012 data were compared with 2013-2017, there was no sign in the earlier data that would explain the bimodal distribution as this occurred in the latter temporal part of the data. Sample sizes are small, being 65 versus 80 individuals. The BC used data from 2006-2017. BC Score 2.

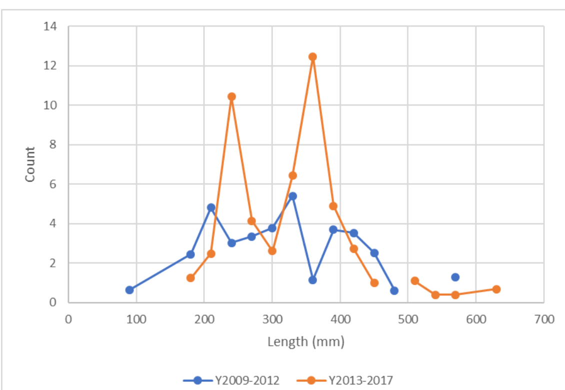

Figure 1. Size frequency distributions comparing 2009-2012 and 2013-2017. Source: Nadon during review

Model plausibility: Total catch smaller than the biomass derived from the catch (no biomass from survey was calculated). However, the standard deviations of the catch derived measures are large. Although the distribution is reasonably narrow with a large tail, a C_{30} value of double that of the median value is still feasible.

Model sensitivity (SPR and overfishing status): None undertaken.

Model sensitivity: Biomass and OFL: None undertaken.

Summary for ToR 2**: (a) Yes for overfishing status; (b) No for OFL setting due to large CVs**

7.1.4.3 Caranx melampygus

Catch: Anomalously high catch in 2013 drives the 4-year average higher. This affects the total catch relative to biomass figures. Given these data are from line fishing and therefore likely higher sample size, this result is less likely to be due to spurious data but is still possible. BC Score 3.

Survey: Mobile species, much deeper range than survey which means the survey is likely an underestimate (ratio species maximum depth to survey maximum depth is 8). BC Score 1.

Life history: BC using mainly life history parameters external to Guam. However, the parameters are taken from several different regions and more likely to be internally inconsistent. BC Score 2.

A stepwise sensitivity test was requested during the workshop. This new test was likely to produce more consistent life history parameters and have higher variance reflecting the mixed region derived uncertainty. This is recommended as the new BC. New BC Score 2.

Short term recommendation 5. Change the *Caranx melampygus* BC to the stepwise approach

Model fit to size frequency: Initially high residuals, but stabilizes with increasing size. BC Score 2. New BC not shown.

Model plausibility: BC – total catch more than biomass derived from the survey as expected; total catch is slightly more than the biomass derived from the catch (but note the one outlier driving this result).

New recommended BC – total catch less than biomass from survey as expected; total catch less than total biomass from catch despite catch outlier in 2013. However, this outlier plus the stepwise uncertainty resulted in a very large standard deviation.

Model sensitivity: SPR and overfishing status: New BC results in same overfishing status as original BC despite moving to stepwise approach.

Model sensitivity: Biomass and OFL: New BC substantially changes the biomass derived output values from the catch (as well as the uncertainty). The new BC CV is unacceptably high.

Summary for ToR 2: (a) Yes for overfishing status but using the recommend new BC; (b) No for OFL setting due to large sensitivity to LH parameters and high CVs

7.1.4.4 Elagatis bipinnulata

Catch: Highly variable. Only one gear (line). BC Score - 2

Survey: Not applicable

Life history: Stepwise approach. Length sample size from 2006-2017 of 214 individuals. L_{99} value drawn from diver surveys in the northern Mariannas so not local. BC Score – 1.

Model fit to size frequency: Large residuals in the middle range of the distribution. BC Score – 1.

Model plausibility: Highly uncertain assessment due to absence of life history parameters, no diver biomass and high CVs in biomass derived from catch. This result is generally well reflected in the output probability distributions.

Model sensitivity (SPR and overfishing status): None undertaken.

Model sensitivity: Biomass and OFL: None undertaken.

Summary for ToR 2: (a) No for overfishing status due to large uncertainty in life history parameters and CV of overfishing status; (b) No for OFL setting due to (a)

7.1.4.5 Myripristis berndti

Catch: Moderately variable in time. Contains low spearfishing period in the data as for all spearfishing data. BC Score – 1.

Survey: Cryptic in day time, deeper water range (ratio of 5). BC Score – 1.

Life history: No local life history and cannot do stepwise approach as not enough life history studies for the family for meta-analyses. Mixture of life history parameters from other regions. BC Score – 1.

Model fit to size frequency: Sample size for length sample is 1,647 individuals. Model fit to residuals are reasonable. BC Score – 2.

Model plausibility: Only 1 LH parameter from HI and none from Guam. No stepwise approach for this species so drawn from a mixture of LH parameters including regions far away from Guam. Total catch greater than both survey and catch derived biomasses. BC Score – 1.

Model sensitivity (SPR and overfishing status): None undertaken.

Model sensitivity: Biomass and OFL: None undertaken.

Summary for ToR 2: (a) No for overfishing status as no family level LH parameters for stepwise approach, many LH parameters from far away and a mixture of regions so uncertainty likely to be under-represented; (b) No for OFL as total catch greater for both survey and catch derived biomass and large uncertainty in LH parameters.

7.1.4.6 Lethrinus erythacanthus

Catch: Highly sporadic catch where there is no catch in some years. BC Score – 1.

Survey: Not applicable.

Life history: Stepwise approach used with L_{99} from 687 individuals so large sample size. BC Score – 2.

Model fit to size frequency: Sample size for size frequency in the BC was very low (49 individuals). A sensitivity test of using data from 2006-2017 expanded the sample size to 154 individuals.

BC model fit to size frequency data is moderate with large expected residuals given the distribution being from a small sample size so reasonably jagged. BC Score – 2.

When the larger sample size is used, the size structure changes substantially with larger animals added and the model fit worsens. It is recommended that the larger sample size test is used as the Base Case despite the worsening fit to the data. This is because the original BC sample size is too small. New BC Score – 1.

Short term recommendation 6. Change the *Lethrinus erythacanthus* BC to the test which used the 2006-2016 data.

Model plausibility: New BC fit to size frequency is poor. This species is very data limited and likely not a good candidate for a single species assessment. Catch is sporadic with periods of almost no or zero catch, making the biomass derived from the catch uncertain. BC and new BC distribution has large component in both overfishing and not overfishing categories even when considering the long tail in the distribution. The recommended BC added larger animals which may be an increase in sample size and/or a sign that the equilibrium assumption may be breached.

Model sensitivity (SPR and overfishing status): Model fit sensitive to extending data to earlier periods as these add larger animals. Overfishing status remains the same, but is very uncertain.

Model sensitivity: Biomass and OFL: biomass from catch is similar to the BC and so too the OFL.

Summary for ToR 2: (a) No, as the model fit is poor to the size frequency data; (b) No as the model fit is poor to the size frequency data, the catch is very sporadic and CV is very large.

7.1.4.7 Lethrinus olivaceus

Catch: Highly sporadic, but not as much as *L. erythracunthus*. BC Score – 1.

Survey: Not seen by divers so not applicable.

Life history: Stepwise approach with L_{99} using 696 individuals from biosampling. BC Score – 2.

Model fit to size frequency: Sample size from the length frequency was 73 individuals for the BC. Some poor residuals for a few size classes probably due to low sample size. BC score 2.

A sensitivity test was undertaken extending the data to 2006, thereby increasing the sample size to 157. This decreased the residuals. This run added larger animals which may be due to the increase in sample size and/or a sign that the equilibrium assumption may be breached. New BC Score – 2.

Short term recommendation 7. Change the *Lethrinus olivaceus* BC to the extended size frequency dataset from 2006, despite this worsening the model fit and decreasing model plausibility.

Model plausibility: BC uses a length frequency sample size that is small, so it is questionable that this represents the population. The larger period (2006-2017) adds more and larger animals, but doubles the size frequency sample size. Although this addition of larger animals may mean that the equilibrium assumption could have been broken, the counter is that the size frequency distribution consists of a more reasonable sample size. The BC total catch value is less than the biomass from the catch. However, the CV is large, and the sensitivity test increases the catch derived biomass and OFL by almost three (3) fold.

Model sensitivity (SPR and overfishing status): More data means that larger animals were added. This reduced the current F and decreased selectivity. BC Score – 2.

Model sensitivity (Biomass and OFL): Model is moderately sensitive to added data in terms of medians, but the recommended BC uncertainty is greatly increased, even considering that there is a large tail to the distribution. BC Score – 2.

Summary for ToR 2: (a) No for BC, but Yes for the extended dataset; (b) No for OFL setting given the sensitivity to the sample size and the uncertainty in the biomass derived figures.

7.1.4.8 Lethrinus xanthochilus

Catch: Highly variable, but not as variable or sporadic as *L. erythracunthus* and *L. olivaceous*. Mostly caught by line with variable levels of net and spearfishing catch. BC Score 2.

Survey: Not applicable as the species is rarely encountered in the data. Depth range divided by survey maximum depth is 33. BC Score – 1.

Life history: Uses the stepwise regression from an L_{99} value obtained using Guam biosampling data (580 individuals). BC Score 2.

Model fit to size frequency: Length data from the linefishing gear component of the creel surveys are used, but with low sample size of 60. Given the small sample size, a sensitivity test was requested during the workshop which used length data from 2006-2017. This increased the sample size to 264 individuals. The BC run may have been undertaken with bins that are too small given the sample size. Either way, the distribution is jagged with zero count length bins and the model fit to the length data poor. BC Score – 1.

The expanded dataset and larger bins produced a smoother distribution and a better fit. New BC Score – 2.

Short term recommendation 8. Change the *Lethrinus xanthochilus* BC to the run with the expanded length dataset.

Model plausibility: BC – Despite the use of the stepwise approach, the CV of the F is less than 100%. However, the uncertainty in the catch data, combined with the poor fit to the length data and the stepwise approach, the CV of the biomass derived from the catch is well above 100%, including a very long tail. Despite this tail, the biomass values within the core is still very uncertain. BC comments above hold for the recommended BC.

Model sensitivity (SPR and overfishing status): The extended dataset run decreased the F, as well as the selectivity (as expected) given the addition of larger animals. The test did not change the overfishing status, although the median of the sensitivity test is close to 1.

Model sensitivity (Biomass and OFL): The new data almost doubles the median biomass derived from catch, which flows through to the C_{30} value.

Summary for ToR 2: (a) No for BC, but Yes for the extended dataset; (b) No for OFL setting given the sensitivity to the sample size and the uncertainty in the biomass derived figures.

7.1.4.9 Monotaxis grandoculis

Catch: Highly variable and, at times, sporadic in the early period, but more stable in recent years. Dominated by spearfishing in the earlier period which mostly falls away over time. Sporadic net catch. BC Score – 2.

Survey: Maximum depth range 3 times survey depth so is likely to be an underestimate. 2011 value much higher than prior and subsequent years. BC Score – 2.

Life history: There are no published LH data for this species, so the stepwise approach was used using the L_{99} from biosampling using 437 animals. BC Score – 2.

Model fit to size frequency: Model fit is good with no major residuals patterns (sample size is 661 animals). BC Score - 3.

Model plausibility: Despite the survey being an underestimate of biomass, there is reasonably good agreement between the diver and catch derived biomass estimates. Although the stepwise approach is used (which increases the variability), the overfishing status has a CV lower than 100%, although 59% of the SPR calculations fell below 0.3. Given the variability of the catch, the biomass derived from the catch is uncertain whereas that from the survey is not (in the context of a data limited approach). Given the similarity between the medians of the two approaches, the survey derived catch and OFL values are preferred even though this could be an underestimate of the total biomass.

Model sensitivity (SPR and overfishing status): A sensitivity test using the L_{99} derived from diver surveys in the northern islands, but this L_{99} value is significantly lower than the Guam L_{99} value. However, the northern islands L_{99} is less realistic given the number of animals larger than this length in the size frequency data. Thus, using the test is not recommended.

Model sensitivity: Biomass and OFL: A sensitivity test using the L_{99} derived from diver surveys in the northern islands was undertaken, but this L_{99} value is significantly lower than the Guam L_{99} value and less realistic given the number animals larger than this length in the size frequency data. Thus, using the test is not recommended.

Summary for ToR 2: (a) Yes for BC, but No for the sensitivity run; (b) Yes for OFL setting with the BC using the survey derived biomass and OFL values, No for the catch derived biomass and OFL values for the BC and No for the sensitivity test

7.1.4.10 Aphareus furca

Catch: Highly variable catch with large peaks every few years. BC Score – 2.

Survey: Depth range of species is 5 times that of the survey. Although 2011 index value highest in the series, this is not as much of an outlier compared to other survey data. Survey data therefore seem more reasonable as a relative index of abundance. BC Score – 2.

Life history: The stepwise approach was used with a L_{99} from Guam using biosampling data. BC Score – 2.

Model fit to size frequency: The spearfishing length data consisted of 280 animals, despite the fishery being dominated by line fishing catch (333 animals). The model fit is reasonable, except for some trends in the residuals as the fit could not describe the peak and there is a higher uncertainty for larger animals. BC Score – 2.

Model plausibility: Despite the large tail, a significant proportion of the overfishing status falls in both overfishing and not overfishing categories due to the high CV. The total catch is less than for both the survey and catch derived biomass, however the catch derived biomass is highly uncertain. That from the survey is more reasonable.

Model sensitivity (SPR and overfishing status): Although a sensitivity run was attempted using the line data, there was not enough size data to undertake this analysis.

Model sensitivity (Biomass and OFL): Although a sensitivity run was attempted using the line data, there was enough size data to undertake this analysis.

Summary for ToR 2: (a) No for BC due to high uncertainty; (b) No, as a consequence of not being able to use the model for overfishing status.

7.1.4.11 Lutjanus fulvus

Catch: Catch is low compared to the diver biomass index trend. Yearly catch not as variable as for other species, except in the early time period where the net catch dominated in some years. BC Score – 2.

Survey: Diver biomass limited by depth compared to the maximum depth range of the species (ratio is 4). However, the survey as an index is not as variable compared to other species although there may be a mean-variance relationship on the index. BC Score – 2.

Life history: The stepwise approach was used with a L_{99} from Guam using biosampling data (326 animals). BC Score – 2.

Model fit to size frequency: The line fishing size data started at a lower size compared to the spearfishing data, thus line data was used (93 animals). The model fit shows increased variability in the residuals with larger animals, which is not surprising given that this part of the distribution is sporadic. BC Score 2.

Short term recommendation 9. Extending the range of the line fishing data could be considered to increase the sample size.

Model plausibility: The overfishing status is well above 1 (with a CV less than 100%) and therefore clearly articulated despite the large tail in the distribution. The total catch is more than the biomass from the catch, but below that derived from survey despite it potentially being an underestimate. The uncertainty for the catch derived biomass is also large, whereas that from the diver is better described.

Model sensitivity (SPR and overfishing status): A sensitivity tests using an L_{99} value from the northern Marianas diver survey was run. However, the sample size was small (35 animals). This sample size is too small, and this test is therefore not recommended despite it changing the model output results.

Model sensitivity (Biomass and OFL): A sensitivity tests using an L_{99} value from the northern Marianas diver survey was run. However, the sample size was small (35 animals). This sample size is too small, and this test is therefore not recommended despite it changing the model output results.

Summary for ToR 2: (a) Yes for BC, but No for the sensitivity test (as also recommended in the assessment report); (b) Yes for the BC survey derived biomass and OFL measures, No for the test and the BC catch derived estimates.

7.1.4.12 Lutjanus gibbus

Catch: Catch is variable especially in early years. Mixed gears used, but mostly dominated by line fishing. BC Score – 2.

Survey: Maximum depth 5 times the survey depth so likely an underestimate. Index has large variance per year, but not as highly variable per survey as other species. BC Score – 2.

Life history: Used external studies from a mixture of regions. Males reach larger sizes than females so an *ad hoc*, but logical approach was applied to adjust the size frequency by creating a female only size frequency. This ratio was calculated from an American Samoa unpublished dataset. The application of this ratio does seem to produce internally consistent results when compared to the male and female L_∞'s. However, this approach has no error included, is unpublished and is fairly *ad hoc*, and this aspect should therefore be included in the P* process when scientific uncertainty is considered. BC Score – 1.

The American Samoa growth curve was suggested as a test, so LH regions are more consistent. Test Score – 1.

Model fit to size frequency: Residuals have increasing uncertainty as observed fork length increases. Sensitivity test residuals likely to be worse but residual plot not shown. BC and test Score – 2.

Short term recommendation 10. Consider including the error in the use of the sex ratio in adjusting the size frequency distribution

Medium term recommendation 18. A more preferred approach would be to obtain sex disaggregated sizes data for *L. gibbus* and any other species that has sex disaggregated biology.

Model plausibility: Overfishing status is specified with large CV given that a mix of LH regions are used, however overfishing status remains mostly below 1 with only the tail above 1. Total catch is less than estimated biomass from both approaches. However, the CV of the biomass from the catch is well above 100%, whereas that for the survey is below 100%. Thus, despite the probability that the survey is an underestimate, it is better described for setting the OFL.

Model sensitivity (SPR and overfishing status): The overfishing status remains consistent with the BC when the American Samoa growth data are used.

Model sensitivity (Biomass and OFL): The sensitivity test results are consistent with the BC and the same conclusions would be drawn with very similar C_{30} derived from the survey. Score – 3.

Summary for ToR 2: (a) Yes for both BC and test; (b) Yes for the survey derived biomass and OFL measures, No for the catch derived estimates.

7.1.4.13 Chlorurus microrhinos

Catch: Mainly spearfishing catch. Concern with sudden drop in the catch after 2004, which is likely to be related to under-reporting of spearfishing catches with low participation rates in the survey. BC Score - 1

Survey: Survey covers large portion (ratio is 1.3) of the species' habitat. Some indication of a mean-variance relation with uncertain large index value in 2014, but followed by a large decrease in the index in 2017 influencing the bootstrap runs using the survey index. BC Score – 2.

Life history: Local LH parameters, apart from the maximum age, which is taken from a Great Barrier Reef study (about 157 individuals). BC Score – 2.

A sensitivity run was undertaken using the Guam longevity value, which therefore changed longevity from 14 to 11 years. Despite the sample size for the local longevity value being relatively low at 80 individuals, it was felt that the region where the BC study was undertaken was not clarified in the workshop and this region could be too different to Guam. The results are also fairly similar. There was concern with the long-term decline in the mean length, but it is unclear if the shift from creel to biosampling data may have affected this dataset. Test score – 3.

Model fit to size frequency: Reasonable fit to the size frequency data for BC and sensitivity test (n=620 animals).

Model plausibility: Overfishing status with reasonable CVs for both the BC and the sensitivity test. Total catch is less than biomass from both options. However, as expected given the uncertainties and that the CV of the biomass from the catch is very large, making this measure highly uncertain.

Despite considering the long tail that contributes to the large CV, these data also have issues of being potentially biased (an underestimate) and this bias could be different per year. The biomass from the survey has a more reasonable CV, which is interesting given the large difference in the index in 2014 compared to other years.

Model sensitivity (SPR and overfishing status): Overfishing status changes between the BC and sensitivity run, however using all local parameters (despite the lower sample size) is preferred.

Short term recommendation 11. Change *Chlorurus microrhinos* BC to the sensitivity run with Guam longevity

Model sensitivity (Biomass and OFL): Trends between the BC and the recommended BC are similar.

Summary for ToR 2: (a) No for BC, Yes for Guam longevity run; (b) Yes for the survey derived biomass and OFL measures, No for the catch derived estimates.

7.1.4.14 Hipposcarus longiceps

Catch: Catch dominated by spearfishing gear. Data seem to have systemic under-reporting in recent years, given the sudden drop in catch after 2004. Evidence for this is that the biosampling catch is higher than the expanded creel catches. BC Score – 1.

Survey: Diver surveys overlap well with the maximum depth of this species (ratio of 1.2). BC Score 3.

Life history: Although unpublished (previously published but included new data), local LH information is available for this species. This approach is preferred compared to using the stepwise or mixed region LH parameters. BC Score – 3.

Model fit to size frequency: Mostly spearfishing therefore well represented in the biosampling with relatively large sample size (1,304 individuals). Model fit good. BC Score – 3.

Model plausibility: Model CVs are below 100% and tail is not as large as observed for some other species. Overfishing status well defined. Reasonable agreement between the biomass derived from the catch and surveys (interesting given the potential bias in the catch data). The uncertainty in the catch is reflected by the large median and tail in the management derived measures from the catch, and is not recommended for OFL setting. That for the survey is more reasonable.

Model sensitivity (SPR and overfishing status): None undertaken.

Model sensitivity (Biomass and OFL): None undertaken.

Summary for ToR 2: (a) Yes for BC; (b) Yes for the survey derived biomass and OFL measures, No for the catch derived estimates.

7.1.4.15 Scarus altipinnis

Catch: Dominated by spearfishing where the participation in the creel survey is low. Highly variable and sporadic over the time series. Peak catch in 2014 with much lower catches before and after. This drives the 4-year value (5,000kg versus the usual 300kg). BC Score – 1.

Survey: Highly variable with zero values leading to 2014, a large index value in 2014, followed by very low values thereafter cannot be easily explained unless the survey is biased or the equilibrium assumption is breached. BC Score – 1.

Short term recommendation 12. Add the earlier zero data in the survey index graph as these are not displayed.

Life history: Local LH parameters were used, but the sample size is very low (53 animals). BC Score – 2.

Model fit to size frequency: Fit to size frequency (1,159 individuals) is reasonably good although residual uncertainty increases with size. BC Score – 3.

Model plausibility: Overfishing status calculated with high CVs, but the bulk of distribution is below F/F_{30} of 1.

Model sensitivity (SPR and overfishing status): None undertaken.

Model sensitivity: Biomass and OFL: None undertaken.

Summary for ToR 2: (a) Yes for BC despite large standard deviation; (b) No given the uncertainties in the data and in the F estimates.

7.1.4.16 Scarus forsteni

Catch: Catch dominated by spearfishing gear. Catch practically disappears after 2004 as is the case for several species. It still shows up as about 200kg.year^{-1} in the biosampling surveys so the creel surveys are likely to be an underestimate. BC Score – 1.

Survey: Survey dominated with a peak in 2011 not shown in the years before or afterwards. This is likely due to a single site. The biomass figures in the remaining years are fairly stable. Biomass survey overlap with species is fairly good (ratio of 1.7). BC Score 2.

Life history: The stepwise approach is used because the L_∞ from the local study of 280mm is too small for the diver and biosampling size structure and is derived from a sample size of 12 animals. The L_{99} is calculated from 353 animals using biosampling. This species would benefit from adding larger individuals using the biosampling dataset to update the Guam growth curve.

Short term recommendation 13. Urgent need to update the age, growth and longevity from local data such as the biosampling. Placed as a short-term recommendation if samples are available, else medium term (High)

Model fit to size frequency: Residual patterns in the model fit with increased uncertainty with larger sizes. BC Score – 2.

Model plausibility: F measures highly uncertain with overfishing status also uncertain even considering long tail. However, using updated local LH is likely to reduce the uncertainty where overfishing status is clearly in one category, which means the overfishing status could be used. Biomass derived from catch is correctly not provided given the uncertainty in the catch. Biomass from the diver survey has a CV less than 100%. Large 2011 value was not included in the calculation.

Model sensitivity (SPR and overfishing status): None undertaken.

Model sensitivity: Biomass and OFL: None undertaken.

Summary for ToR 2: (a) No for BC given large standard deviation; (b) No given the uncertainties and No assessment from (a). NB: Using updated local LH is likely to reduce the uncertainty and is likely to place the species as Yes as overfishing status is clear. If the F can be used, then the survey derived management measures could be used for OFL setting.

7.1.4.17 Scarus rubroviolaceus

Catch: Catch dominated by spearfishing gear. Catch practically disappears after 2004 as occurs for several species. It still shows up as about 120kg.year^{-1} in the biosampling surveys so the creel surveys are an underestimate. BC Score – 1.

Survey: Some concern with the peak in biomass in 2011 which is likely to be due to a single large sighting event. The biomass data in the remaining years are fairly stable. BC Score – 2.

Life history: Local LH parameters with low sample size (55 animals), while BC uses HI longevity based on 182 animals. The HI longevity of 22 is much longer than the longevity of 6 years from Guam. Although using Guam LH parameters is the preferred approach, the age gap between HI and Guam is large and the Guam sample size is small making this value highly uncertain. Using the HI longevity is therefore more precautionary. Updating the longevity from the local biosampling data is therefore a priority. BC Score - 2

Short term recommendation 14. Increase sample size for age and growth LH parameters using the biosampling data, examine larger fish for potential higher longevity estimate from local life history data. Placed as a short-term recommendation if samples are available, else medium term (High)

Model fit to size frequency: Fit to the size frequency is reasonable (sample size of 467 from biosampling) with a few exceptions in the residuals and increased residual variance at largest sizes. BC Score – 3.

Model plausibility: Overfishing status with high CVs, but still below 100%. Total catch less than biomass derived measures. Given uncertainty in the creel survey derived catch for uncertainty, this is captured in the CV of the biomass derived from the catch being much greater than 100% (an order of magnitude bigger standard deviation compared to the median). The biomass derived from the survey is less uncertain, but the mean is still slightly more than 100%.

Model sensitivity (SPR and overfishing status): None undertaken.

Model sensitivity: Biomass and OFL: None undertaken.

Summary for ToR 2: (a) Yes for overfishing status (but note the high CV implies that this status is still fairly uncertain, but passes the criterion used); (b) No for OFL setting given high uncertainties.

7.1.4.18 Scarus schlegeli

Catch: Catch dominated by spearfishing gear, although there is the occasional net gear catch in the series that can be significant. Catch practically disappeared after 2000 as occurs for several species dominated by spearfishing catch. BC Score – 1.

Short term recommendation 15. Add the descriptor “other gear (black)” in the legend – and anywhere else where this occurs.

Survey: Diver survey biomasses are variable, but within a consistent range i.e. no sign of a major outlier. Max depth ratio is 3. BC Score – 3.

Life history: Most LH from Guam whereas the growth parameters come from Australia. A sensitivity run was undertaken using the stepwise approach using local L_{99} values which changed the L_{∞} value from 200 mm in the BC to a median of 265 mm in the sensitivity test. This latter value is closer to the Taylor (unpublished) value of 252 mm which provides some evidence that a local growth curve could change the results to some extent. This mixture of regions increases the uncertainty. BC Score – 2.

Short term recommendation 16. Urgent need to update the age, growth and longevity LH from local data such as the biosampling. Placed as a short-term recommendation if samples are available, else medium term (High)

Model fit to size frequency: Large uncertainty in the residual at the mode of the distribution with the residuals better at small and larger sizes. BC score – 2.

Model plausibility: Although the overfishing status is very uncertain, and the CV is greater than 100%, the median value and the bulk of the distribution is well below 1. Total catch is less than the biomass derived from catch and the survey. However, the catch derived biomass CV is much greater than 100%, reflecting the variability in the catch. The survey biomass median is below 100%, but it is noted that the C_{30} derived from the survey is above the recorded catch history maximum in the earlier years when the creel survey may have been more accurate.

Model sensitivity (SPR and overfishing status): The sensitivity test using the stepwise approach increases overfishing status median value to 0.95 with a higher F and F_{30} value adding some uncertainty to stock status particularly since this test seems to reflect local Guam unpublished information better.

Model sensitivity (Biomass and OFL): These values were not provided.

Summary for ToR 2: (a) No for overfishing status, but urgently undertake local growth curve analyses to address the overfishing status uncertainty given that the BC is more certain in terms of overfishing status and the “No” result is driven by the uncertainty created by the sensitivity test that produces an overfishing status closer to 1; (b) No for OFL setting given the No assessment for overfishing status setting.

7.1.4.19 Variola albimarginata

Catch: Catch has been low for most of the series although higher for the last two decades. This catch is dominated by line fishing. This small catch is variable. BC Score – 2.

Survey: Not applicable as never encountered by survey divers.

Life history: The stepwise approach was used with the L_{99} from biosampling (n=435). BC Score – 2.

Model fit to size frequency: Year to year sample size is very low. Size distribution of line and spearfishing gear are similar – the BC uses line fishing size data (n=96). Given the small sample size, the residual pattern shows large uncertainty due to the few larger and smaller sized animals.

A sensitivity test was undertaken using spearfishing size frequency data.

An additional test during the review was undertaken to use the line fishing data from 2006-2017, which changed the sample size from 96 to 172 individuals. This run should be the new Base Case if used for management. Score -2.

Model plausibility: Total catch is more than the biomass from the catch. The CV for these biomass derived values are much higher than 100% and are highly uncertain.

Model sensitivity (SPR and overfishing status): The sensitivity test using spearfishing size data changed the result very little compared to the BC. However, this was not the case for the longer size frequency data period where comparatively more of the smaller and larger animals were added. Although the BC overfishing status CV is greater than 100%, few iterations of the SPR < 0.3 statistic fell below 1 and thus most of the overfishing status values were > 1. However, this is not the case for the extended data test, where the overfishing status CV is large and SPR < 0.3 iterations are 51%. BC Score – 2.

Model sensitivity (Biomass and OFL): The sensitivity test using spearfishing size data changed the result very little compared to the BC. However, the biomass from the catch was doubled, bearing in mind the small dataset.

Summary for ToR2: (a) No considering the uncertainty created by adding a few extra length data; (b) No for OFL setting given high uncertainty, biomass derived catch less than total catch and No assessment for overfishing status.

7.2 ToR 2: Determine whether the results for individual species from question 1 can be used for management purposes under the Magnuson-Stevens Act and relevant Fishery Ecosystem Plan (FEP) 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.

7.2.1 Overall response to question

It was determined that the approach can be used for management purposes for a) 12 of the 19 species reviewed for calculating the overfishing limit i.e. determining overfishing status; but b) only 5 of these were able to be used to generate the OFL. It should be noted that this is mostly not due to any issue with the method itself, but that either (or both) the life history information was too uncertain or that the catch and/or survey data generally could not conform to the assumption that they can be applied as an absolute and unbiased index of abundance.

For those species where the reviewers' assessment was "No" for OFL setting, the Biomass Augmented Catch – MSY (BAC-MSY) approach is favoured. Ideally, this approach would work best if life history is considered in the combination of species data, rather than using family as before. However, this approach is only recommended in the short-term.

7.2.2 Background

The overall 2017 Guam assessment includes a set of alternative approaches depending on the available information. This is an impressive set of alternatives allowing for each species to be assessed with the best available data and approaches. The methods themselves are sound, but rely heavily in the second part of the model on the assumption that catch or survey data are an index of absolute abundance when scaling up the F values to calculate OFL and other catch-based management measures. As described in ToR 1 and summarised below, this assumption was found not to be true in most cases and is often clearly demonstrated in inconsistencies among the model outputs, such as the estimated biomass being less than (in some cases already downward biased) total catch.

As an alternative based on the previous assessments, the Sabater and Kleiber (2013) approach was based on the Catch-MSY approach of Martell and Froese (2012), but with the addition of also including survey biomass indices. This model was used to set the Annual Catch Limit. The Catch-MSY approach was not deemed able to determine overfishing or stock status (Yau, presentation during workshop). The mathematics of the approach is described in some detail and "preliminary" results are provided for American Samoa (Sabater and Kleiber, 2013). However, the Guam model outputs only show as final catch MSY and risk of overfishing tables in the document's Appendices.

Prior to the BAC-MSY approach (2012-2014), the ACL was based on the 75th percentile of the catch (hereafter called the C75 approach) for the entire catch series from 1985. Each year, the new year's data were added. The issue with this approach is that, if the ACL was decreased, the catch would decrease, and this would continually ratchet the ACL downwards (or upwards if the ACL is set in the opposite direction). Given that the same ratcheting effect would also increase the ACL, this approach was seen as creating a disincentive to remain within the ACL. The details of the C75 approach were only made available upon request during the review and mainly consisted of a set of tables that described the ACL's set over time for Guam.

7.2.3 Species summary table

Leading from ToR 1 where the model plausibility and sensitivity in terms of overfishing status and OFL setting is described was then summarised in terms of ToR 2 for clarity, this section provides a summary of the responses to ToR 2 by species below.

Table 1. Per species response to determining overfishing status and OFL setting leading from ToR 1

Species	Verdict regarding determining overfishing status ("F status")	Verdict regarding OFL setting advice
Naso unicornis	**Yes,** for the 23-year longevity Base Case	**No,** based on the results from the model plausibility test.
Carangoides orthogrammus	**Yes**	**No,** due to large CVs.
Caranx melampygus	**Yes,** but recommend new BC using the stepwise approach	**No,** due to large sensitivity to LH parameters and high CVs.
Elagatis bipinnulata	**No,** due to large uncertainty in life history parameters and CV of overfishing status	**No,** due large uncertainties and (a).
Myripristis berndti	**No,** as no family level LH parameters for the stepwise approach, many LH parameters from far away and a mixture of regions, so uncertainty likely to be under-represented	**No,** as total catch greater for both survey and catch derived biomass and large uncertainty in LH parameters.
Lethrinus erythacanthus	**No**, as the model fit is poor to the size frequency data	**No,** as the model fit is poor to the size frequency data and the catch is very sporadic and CV is very large.
Lethrinus olivaceus	**No** for BC, **Yes** for the extended dataset	**No,** given the sensitivity to the sample size and the uncertainty in the biomass derived figures.
Lethrinus xanthochilus	**No** for BC, **Yes** for the extended dataset	**No,** given the sensitivity to the sample size and the uncertainty in the biomass derived figures.
Monotaxis grandoculis	**Yes** for BC	**Yes** with the BC using the survey derived biomass and OFL values, No for the catch derived biomass and OFL values for the BC, and No for the sensitivity test.
Aphareus furca	**No** for BC due to high uncertainty	**No**, as a consequence of not being able to use the model for overfishing status.
Lutjanus fulvus	**Yes** for BC	**Yes** for the BC survey derived biomass and OFL measures, No for the test and the BC catch derived estimates.

Lutjanus gibbus	**Yes** for BC (and test)	**Yes** for the survey derived biomass and OFL measures, No for the catch derived estimates.
Chlorurus microrhinos	**No** for BC; **Yes** using the Guam longevity value	**Yes** for the survey derived biomass and OFL measures, No for the catch derived estimates.
Hipposcarus longiceps	**Yes** for BC	**Yes** for the survey derived biomass and OFL measures, No for the catch derived estimates.
Scarus altipinnis	**Yes** for BC despite the large standard deviation	**No,** given the uncertainties in the data and in the F estimates.
Scarus forsteni	**No** for BC given large standard deviation	**No,** given the uncertainties No for F values. NB: Using updated local LH is likely to reduce the uncertainty and is likely to place the species as Yes as overfishing status is clear. If the F can be used, then the survey derived could be used for OFL setting.
Scarus rubroviolaceus	**Yes,** but note the high CV implies that this status is still fairly uncertain, yet passes the criterion used	**No,** given high uncertainties.
Scarus schlegeli	**No**, but urgently undertake local growth curve analyses to address the overfishing status uncertainty given that the BC is more certain in terms of overfishing status and the "No" result is driven by the uncertainty created by the sensitivity test that produces an overfishing status closer to 1	**No,** given the No assessment for overfishing status setting.
Variola albimarginata	**No,** given the uncertainty created by adding a few extra length data.	**No** for OFL setting given high uncertainty, biomass derived less than total catch and No assessment for overfishing status.

7.2.4 Short-term alternatives to the GTG-LBSPR

Given the lack of stock assessment standard documents focusing on Guam, it was difficult to fully assess the efficacy of the BAC-MSY and C75 approaches relative to the 2017 assessment. Based on information provided during the workshop, the BAC-MSY approach was reviewed, whereas the C75 approach was not. The reviewer obtained the 2014 CIE reviews of the BAC-MSY approach from the NOAA website. Importantly, similar to this review, the quality of the data was queried: "It is recommended that a detailed review be undertaken of the data collection and processing to prepare the data for analysis in the assessment method. Criteria need to be developed for deciding when a dataset or data point is atypical, not representative of the fishery, or constitutes an outlier. In this way ad hoc decision during the analysis can be avoided and the outcomes would become more defensible." (Haddon, 2014).

Since many of the issues that affect the catch data in 2013 and 2017 still exist today, all the past approaches used for Guam would be affected to some degree. Thus, it is worth considering some of the major differences between the approaches in terms of the data.

Both the BAC-MSY and the C75 approaches were applied at the family level, whereas the 2017 assessment was undertaken at the species level. There are clear trade-offs between a combined species assessment and one at the species level. The obvious benefit of undertaking a species level assessment like the GTG-LBSPR is that species level life history parameters could be used. Lumping at the family level would place very different life histories together, where potentially longer-lived, more vulnerable species may be neglected. This issue was also highlighted by one of the CIE reviewers in 2014 (Haddon, 2014).

Uncertainties in the GTG-LBSPR may increase at the species level if LH parameters are not known at the species level, and the catch or survey is highly variable or biased as one disaggregates the data. In the case of the 2017 assessment, only five species had life history parameters entirely derived from Guam, three from external LH studies and 11 used the stepwise approach that draws from family level information.

The catch for many of the species was highly sporadic and/or variable. The catch data of eight species were scored as 1 (poor representation of the major catch source) and 10 were scored at 2 (catch more representative but variable). Only one species was scored at a 3 (good). Arguably aggregating the catch data for those species that scored a 2 to some higher level would most likely reduce some of this variability.

On the other hand, aggregating would not simply resolve the issue of bias, as is the case for the spearfishing catch dominated species (which mostly scored a 1). Sabater and Kleiber (2013) tried to address this latter issue of under-reporting in the creel survey data by combining the creel survey and commercial receipt-book data. This may already go part-way to addressing this issue of bias. Their assumption that double counting would be at a minimum seems appropriate. Additionally, the biosampling data could also be used to fill in the gap, but this step was not done by Sabater and Kleiber (2013), nor was it in the 2017 assessment.

The 2017 assessment has an advantage over the BAC-MSY in that uncertainty in the catch data was included. The 2017 assessment assumed that this uncertainty is described by the inter-annual variability of the most recent catches. This would constitute an advanced step compared to not including any uncertainty as per the BAC-MSY (as pointed out by the 2014 CIE reviewers – Jones, Haddon, Cook). However, this again does not reflect the true uncertainty (which can only be done by bootstrapping the whole expansion process i.e. including annual uncertainty as well as inter-annual variability). However, even this approach does not address bias.

Despite 75% of the average catch being used elsewhere in the world, a difference when this approach was applied in Guam was to continually extend the catch series to include new data. The ratcheting effect of the data is very real and therefore this approach would not be recommended unless the time period is fixed.

The ideal would have been that the 2017 assessment used both the biosampling and creel survey data (as already recommended in this review) in its approach. As a test, the 2017 assessment should have run the approach lumping species of similar biology together. As is generally found with data limited species, undertaking alternative approaches such as updating the BAC-MSY would also have been beneficial. However, these were not available for review. Therefore, given the large uncertainties in the single species assessment and the ratcheting effect of the C75 approach, for those species where the reviewers' assessment was "No" for OFL setting, the BAC-MSY approach is favoured. Ideally, the approach would work best if life history is considered in the lumping together of remaining species.

However, this is only recommended in the short-term.

Short term recommendation 17. For those species where the reviewers' assessment was "No" for OFL setting, the BAC-MSY approach is favoured. Ideally, the approach would work best if life history is considered in the lumping of species. However, this is only recommended in the short-term.

7.2.5 Medium term approaches

In the medium-term it is still unlikely that a good catch series would be available that is long enough to apply a dynamic model (e.g. a biomass dynamic model). This is despite the recommended review of the data and steps being taken to collect a more comprehensive dataset.

Although Coefficient of Variations of output parameters are not necessarily a reflection of the uncertainty given the distributions of the parameters are usually very skewed, these CVs do provide some hint in terms of where to emphasise future work. In the 2017 assessment, the F_{30} estimate often had lower coefficient of variation compared to the related F/F_{30} estimate due to large uncertainty in the current F estimation (Figure 2). The plot also shows that the CV increases with an increase in fishing mortality relative the F_{30} i.e. becomes less certain when fishing mortality is high. This CV in the F_{30} is then amplified by the biomass scalar uncertainty. There are three ways to reduce C_{30} CV, which is to a) move more species away from the stepwise approach (Figure 3), b) increase the smoothness of the size frequency plots by increasing the sample size and c) to reduce the uncertainty surrounding the biomass scalar e.g. catch (either through higher sample size in the creel survey or more targeted catch sampling approaches, smoothing the catch data series or through aggregating the catch data). The greatest amplification in the CVs were seen when the C_{30} is calculated from the catch i (Figure 4).

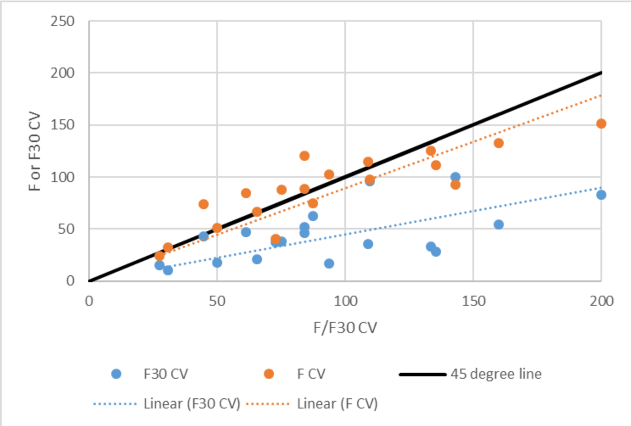

Figure 2. Coefficient of Variation (CV) (%) of the current F, F_{30} and F/F_{30} statistics for the 19 species

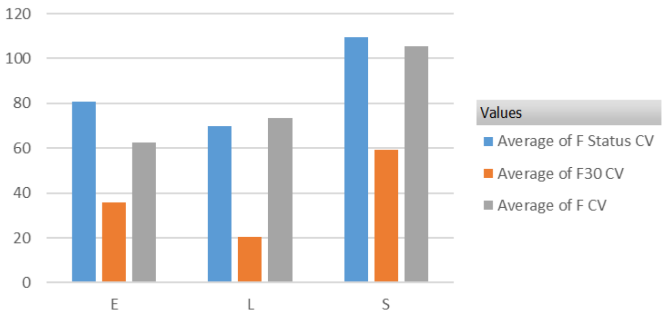

Figure 3. Average Coefficient of Variation (%) for F/F_{30} (F status), F_{30} and F by LH method (E=External LH parameters, L=Local, S=Stepwise approach).

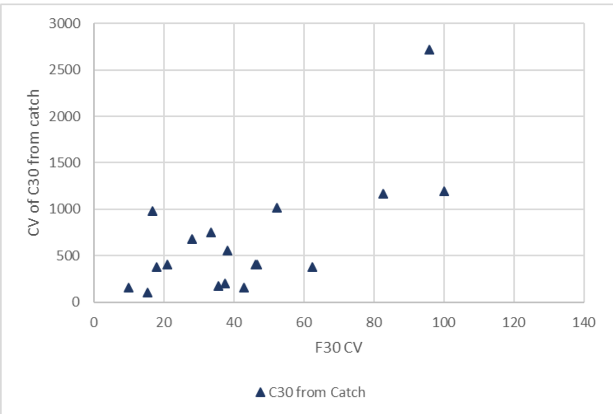

Figure 4. CV (%) of the F_{30} versus C_{30} from Catch CV

The *ideal* medium-term strategy would be to use the GTG-LBSPR model for as many species as possible. This would mean a) increasing the number of species for which local LH parameters are available, b) obtaining a recent biomass scalar (catch or biomass) and c) increasing the sample size of length data for species not captured by the biosampling.

However, it should be noted that this view would need to be moderated post review of the creel and independent surveys, which may find that one or both creel and independent surveys could not be used for assessment purposes. In these cases, see the section on alternative approaches to the GTG-LBSPR approach section below.

7.2.5.1 Expanding the local LH parameters

A list of species is already provided in ToR 1 for short-term LH parameter updating. With the samples collected in the biosampling program, this list should be expanded in the medium-term. Species such as *C. melampygus* and *M. berndti* could be the initial focus (after the species mentioned as short-term recommendations).

Medium term recommendation 19. Obtain or enhance local life history parameters for *C. melampygus* and *M. berndti*.

7.2.5.2 Obtaining a recent absolute biomass scalar

The most applicable absolute biomass scalar, given how it is used in the GTG-LBSPR model, is most likely to be total catch. This is because the biomass survey is unlikely to be scientifically treated as an absolute index of abundance, unless (or probably even if) some calibration is undertaken (but see below on its potential use as a relative index of abundance). The strength of the GTG-LBSPR model is that it uses the most recent catch or survey data, rather than the whole catch time series. This means that efforts undertaken to better determine catch into the future would reflect better on the use of the GTG-LBSPR model in the medium term than, for example the BAC-MSY as long as the equilibrium assumption is not breached. Here, the emphasis for enhanced catch sampling would be on species where spearfishing dominates (as already recommended in ToR 1).

7.2.5.3 Increasing the sample size of the length data

In general, increasing the size of the length dataset would mostly apply to species that were not well sampled in the biosampling program (assuming the biosampling program continues in its present or in an expanded format). Examples of the species where the fit to the length data and/or small sample size affected the utility of the 2017 assessment were *L. erythocanthus*, *L. olivaceous*, *L. xanthochilus* and *V. albimarginata*. In addition, *N. unicornis* length frequency was highly uncertain in certain years and should be a focus for future investigation.

Medium term recommendation 20. Obtain increased length dataset for L. erythocanthus, L. olivaceous, L. xanthochilus and V. albimarginata as an initial focus.

Medium term recommendation 21. Investigate the reasons for the uncertainty in the *N. unicornis* length frequency.

7.2.5.4 Single species versus groups

Given the management system and the fishery characteristics in Guam, managing a species complex as a norm may be much more practical (from a management and scientific credibility point of view). Yet, aggregating species to the presently used family level may compromise some species too much. A study to evaluate the best balance between lumping or disaggregating different species using factors such as data quality (variability and bias), life history, vulnerability, catch size and value etc. should be undertaken. This could be a key input to a Management Strategy Evaluation (see section below) where aggregated and single species assessment approaches could be tested.

Medium term recommendation 22. Undertake a study to evaluate the best balance between lumping or disaggregating different species using factors such as data quality (variability and bias), life history, vulnerability, catch size and value, etc.

7.2.5.5 Alternative approaches to GTG-LBSPR

Although the GTG-LBSPR has strengths, one aspect of potential difficulty will remain the ability to scale to absolute catch if the data cannot be more accurately collected. As stated in ToR 1, there are options for scaling the survey but, in reality, highly reliable data are required anywhere in the world for a survey to be reliably scalable to absolute catch. The difficulty in Guam is that the total catch data are uncertain and using data moderate dynamic methods and data limited methods such as the catch only or biomass dynamic approaches (and variants thereof) will remain scientifically questioned unless the catch data can be corrected for at least the aggregated data over most of the time series.

However, in the case of many species, the survey seemed to be a reliable relative index of abundance, but not of absolute abundance. For example, there were several cases where the depth range of the species was much deeper than the maximum survey depth, yet the species is easily surveyed (i.e. not cryptic). In these cases, the finding is usually that the survey could not be used for OFL setting using the 2017 assessment approach since this index does not conform to the assumption that it is an absolute index of abundance. On the other hand, the ACL can be changed in proportion to the change of the survey index over the period of the data used, i.e. the survey is used as a relative index. This approach would mean the survey would not require calibration (although it does assume the depth range of the species does not substantially change). Here simple data limited harvest strategies could be applied that simply apply Fmsy or Bmsy proxies to change the ACL relative to the change in relative biomass from the survey. This approach is likely to be more robust to data issues with the catch than any approach that tries to scale up using a poor index of absolute abundance. This would require updating the survey index to better reflect relative biomass as recommended in ToR 1.

7.2.5.6 Long-term recommendations

If effort on the survey and catch monitoring is continued, biomass dynamic approaches and related measures would be useful. The value of the survey would increase with time, if the issues of survey design and depth are resolved. Thus, the actions recommended in the medium term, if mostly undertaken, will lead to an increased ability to use more data intensive methods such as biomass dynamic models.

In addition, the range of approaches would further increase if age data are collected for the more common species such as *N. unicornis*. Even the age frequency equivalent approaches to the GTG-LBSPR would then be available as an alternative and test.

Long-term recommendation 1. Start investigating the utility of ageing certain key species such as *N. unicornis*.

Long-term recommendation 2. If the total catch figures have been updated to a level that there is a time series useful for stock assessment, undertake a biomass dynamic assessment with an open shape form (as opposed to assuming a Schaeffer shape).

7.2.5.7 Management Strategy Evaluation

In data limited fisheries, there is great value in undertaking simulation testing. The approach that also investigates the impact of sequential management decisions over time (i.e. the management feedback loop) is Management Strategy Evaluation (MSE) (Punt et al., 2016). Key trade-offs that should be investigated (and would inform where to focus monitoring, life history and assessment attention) are:

- The trade-off between single species and aggregated species lists;
- The most appropriate form of aggregation of species;
- Different assessment and harvest strategy options;

- Trade-off between increasing length frequency sample size through extending the number of years in the sample versus breaching the equilibrium assumption;

- Using smoothing functions (e.g. *loess*) on the catch data to reduce catch variability versus allowing the ACL magnitude to change dramatically for highly catch variable species;

Medium term recommendation 23. To address many of the trade-offs highlighted by the assessment approach and the difficulties with the 2017 assessment approach using Guam data, a Management Strategy Evaluation (Punt *et al*., 2016) or simulation test should be undertaken. This would address issues such as articulating the trade-off between single species and aggregated species lists, the most appropriate approach to aggregating species; different assessment and harvest strategy options; increasing the length frequency sample size by including more years of data versus breaching the equilibrium assumption as more years are included, smoothing the catch data to decrease catch variability versus reducing ACL responsiveness etc. Furthermore, this approach could be tested against running the SPR model component and then applying a harvest strategy that uses the survey data as a relative index of abundance (High).

7.3 TOR 3: 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 (3-5 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.

7.3.1 Short term recommendations

7.3.2 Medium term recommendations

7.4 ToR 4: Draft a report (individual report from Chair and review members, and additional consensus report from Chair) addressing the above TOR questions.

This report and the chair's report achieves ToR 4.

8 CONCLUSIONS AND RECOMMENDATIONS IN ACCORDANCE WITH THE TORS.

Unfortunately, Dichmont et al. (2016) has shown that when the Australian Commonwealth (Federal) set of harvest strategies and associated assessments are tested in a Management Strategy Evaluation that "On average, more data lead to improved management in terms of risk of being overfished and not reaching a target". As a caveat, Fulton et al. (2016) found that the USA buffer system used in similar species to the study performed better at achieving risk equivalency across tiers. Thus, the ideal suggestion of collecting more accurate data is preferred, but it may not be realistic given the difficulties of a greatly dispersed fishery using multiple gear types, landing locations and including recreational as well as commercial fisheries across all these gear types. Thus, a hierarchical approach is suggested:

- In the short-term, for species that the GTG-LBSPR could not be used on, the BAC-MSY approach is recommended (but not at the single species level). It is suggested that similar species are grouped together for this approach, rather than the default of family.
- Some of the species can be assessed using the present GTG-LBSPR. Increasing this list through more LH work and/or (more importantly) improving the quality of the catch data would be required.
- Other species may need to remain as species complexes and assessed using the GTG-LBSPR.
- In the medium to long term, several additional data collection programs are suggested, including expanding the life history parameters for a select group of species, more accurate data collection programs especially for the spearfishing fishers and increasing the length frequency data collection program for the line fishing group, amongst others.
- A Management Strategy Evaluation would be seen as a significant enhancement in the medium term where complex trade-offs can be tested, such aggregation of species, expanding time periods of data and using the survey as a relative index of abundance.

9 REVIEW PROCESS

The review was undertaken in a very positive and helpful light.

Two issues that affected the review should be noted:

1. It was required in ToR 2 that the data were considered to have been accepted for stock assessment purposes despite no evidence of the data having been reviewed. After much deliberation, it was not possible to assume as a default that the data was without error, given information provided during the review. For this clause to be accepted, much more detailed documentation on the data, methodologies used and approach to obtaining data fit for assessment used should have been provided and the data was peer reviewed.
2. The previous assessment (Sabater and Kleiber, 2013) was provided. Although there is reasonable presentation on the mathematics of the approach, the results focused on American Samoa and was clearly stated in the text as preliminary. The Guam information was only provided as two tables in the Appendices without diagnostics and tests. The approach using the 75th percentile of the catch was only described and provided to the panel during the review. This hampered a detailed investigation of the alternative approaches if an assessment of “No” for management use was provided.

10 References

Cook, R. 2014. Report on the Review of the Biomass Augmented Catch-MSY Model for Pacific Island Coral Reef Ecosystem Resources. (https://www.st.nmfs.noaa.gov/Assets/Quality-Assurance/documents/peer-review-reports/2014/2014_08_04%20Cook%20PIFSC%20ACL-MSY%20model%20review%20report.pdf)

Dichmont, C.M., Fulton, E., Gorton, R., Sporcic, M., Dowling, N., Little, R.L., Punt, A.E., Haddon, M., Smith, D.S. (2016). From data rich to data-limited harvest strategies – does more data mean better management? ICES J. Mar. Sci. ICES Journal of Marine Science, doi:10.1093/icesjms/fsw199

Fulton, E.A., Punt, A.E., Dichmont, C.M., Gorton, R., Sprocic, M., Dowling, N., little, L.R. Haddon, M., Klaer, N., Smith, D.C. (2016). Developing risk equivalent data-rich and data-limited harvest strategies. Fisheries Research. 183: 574-587. 10.1016/j.fishres.2016.07.004.

Griffiths, S. P., Pollock, K. H., Lyle, J. M., Pepperell, J. G., Tonks, M. L. and Sawynok, W. (2010), Following the chain to elusive anglers. Fish and Fisheries, 11: 220–228. doi:10.1111/j.1467-2979.2010.00354.x

Haddon, M. 2014. Center for Independent Experts (CIE) Peer Review of the Biomass Augmented Catch-MSY Model for Pacific Island Coral Reef Ecosystem resources. (https://www.st.nmfs.noaa.gov/Assets/Quality-Assurance/documents/peer-review-reports/2014/2014_08_04%20Haddon%20PIFSC%20ACL-MSY%20model%20review%20report.pdf)

Jones, C. 2014. BiomassAugmented Catch-MSY Model for Pacific Island Coral Reef Ecosystem Resources. (https://www.st.nmfs.noaa.gov/Assets/Quality-Assurance/documents/peer-review-reports/2014/2014_08_04%20Jones%20PIFSC%20ACL-MSY%20model%20review%20report.pdf)

Martell, S. and Froese, R. (2013), A simple method for estimating MSY from catch and resilience. Fish Fish, 14: 504–514. doi:10.1111/j.1467-2979.2012.00485.x

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A. A. and Haddon, M. (2016), Management strategy evaluation: best practices. Fish Fish, 17: 303–334. doi:10.1111/faf.12104

11 APPENDIX 1: BIBLIOGRAPHY OF MATERIALS PROVIDED FOR REVIEW

11.1 REVIEW PROCEDURE

1. Statement of Work: Center for Independent Experts' Contribution of Reviewers to the Western Pacific Stock Assessment Review (WPSAR) of the 2018 Guam Reef Fish Stock Assessments

2. Tentative Agenda.

11.2 DOCUMENTS (TO BE REVIEWED ARE IN BOLD)

11.2.1 Benchmark assessment document

- **Nadon, M. O. 2016. (draft) Stock assessment of the coral reef fishes of Guam, 2017. U.S. Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-XX, ~200 p.**

11.2.2 Relevant management information

- Western Pacific Regional Fishery Management Council. 2009. Fishery Ecosystem Plan of the Mariana Archipelago. Sections 4.4.2 and 5.6 only.
- Western Pacific Regional Fishery Management Council. 2011. Omnibus Amendment for the Western Pacific Region to Establish a Process for Specifying Annual Catch Limits and Accountability Measures. Section 3.1 only.

11.2.3 Background documents

- Hordyk, A.R., Ono, K., Prince, J.D., and Walters, C.J. (2016). A simple length-structured model based on life history ratios and incorporating size-dependent selectivity: application to spawning potential ratios for data-poor stocks. Can. J. Fish. Aquat. Sci. *73*, 1787–1799.
- Kritzer, J.P., Davies, C.R., and Mapstone, B.D. (2001). Characterizing fish populations: effects of sample size and population structure on the precision of demographic parameter estimates. Can. J. Fish. Aquat. Sci. *58*, 1557–1568.
- Nadon, M.O., and Ault, J.S. (2016). A stepwise stochastic simulation approach to estimate life history parameters for data-poor fisheries. Can. J. Fish. Aquat. Sci. *73*, 1874–1884.

11.2.4 Previous stock assessment

- Sabater, M, and Kleiber, P. 2013. Improving Specification of Acceptable Biological Catches of Data-Poor Reef Fish Stocks Using a Biomass-Augmented Catch-MSY Approach. Report of the Western Pacific Regional Fishery Management Council.

11.2.5 Supplemental Background Documents

- *Hawaii assessment and independent peer review report:* Choat, JH, Franklin, EC, and Stokes, K. 2016. Benchmark review of the 2016 stock assessment
  of the main Hawaiian Islands reef-associated fish. Consensus panel report prepared by Erik C. Franklin.

- Nadon, M.O. (2017). *Stock* assessment of the coral reef fishes of Hawaii, 2016. PIFSC Tech Memo 60. 200p.

- *References:* Nadon, M.O., Ault, J.S., Williams, I.D., Smith, S.G., and DiNardo, G.T. (2015). Length-based assessment of coral reef fish populations in the Main and Northwestern Hawaiian Islands. PLoS ONE *10*, e0133960.

11.2.6 Additional *documents* provided during the review

- Tribble, G. 2013. Review of Information for Fishery Stock Assessments. Pacific Islands Fishery Science Center. Honolulu Hawaii.
- Anon, 2017. P* Working Group Meeting minutes, Western Pacific Regional Fisheries Management Council.
- Department of Commerce. 2015. Specification of Annual Catch Limits and accountability measures for the Pacific Island Coral Reef Ecosystem Fisheries in Fishing year 2015.
- Department of Commerce. 2015. Pacific Island Fisheries; 2015 Annual Catch Limits and Accountability Measures. Federal Register. 80(168): 52415-18
- Department of Commerce. 2016. Pacific Island Fisheries; 2016 Annual Catch Limits and Accountability Measures. Federal Register. 82(76): 18716-20
- Bak, S. 2012. Evaluation of Creel Survey Program in the Western Pacific Region (Guam, CNMI, and American Samoa). A Report to the Western Pacific Regional Fishery Management Council.
- Anon. 2014. 115th Meeting of the Scientific and Statistical Committee Council Office March 11-13, 2014 minutes, Western Pacific Regional Fisheries Management Council.
- Simons, K. Jan 10, 2018. Letter to Western Pacific Regional Fisheries Management Council on Council actions on the 127th SSC meeting recommendations.
- NOAA Fisheries. 2013. Background and PIFSC Response Concerning Review Panel Reports from The External Review of Information For Fishery Stock Assessments.
- NOAA Fisheries. 2014. Background and PIFSC Response: Panel reports of the Stock Assessments Program Review.
- Murawski, S. 2014. Program review of stock assessment process Pacific Island Fisheries Science Center.
- Nanami, A., Kurihara, T., Kurita, Y., Aonuma, Y., Suzuki, N., Yamada, H. 2010. Age, growth and reproduction of the humpback red snapper *Lutjanus gibbus* off Ishigaki Island, Okinawa. Ichthyol. Res. 57: 240-244.
- Creel vs. Biosamp differences. Xlxs.gsheet
- Anon. Current State of Guam Creel Surveys and Suggestions for the Future. “Guam expansion summary.docs”
- Lengthsamplesize.xlsx.gsheet
- LH Par.xlsx
- Melampygus sensitivity run.xlsx
- Microrhinos sensitivity run.xlsx
- New base cases.docs
- Orthogammus length comparison.xlsx
- Re-runs.docx
- Toby Mathews response to Creel inquiry.gdoc
- WPSAR 2017 Guam Reef Fish – Meeting Notes.gdoc

12 APPENDIX 2: A COPY OF THE CIE STATEMENT OF WORK

Statement of Work
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

Center for Independent Experts' Contribution of Reviewers to the Western Pacific Stock Assessment Review (WPSAR) of the 2018 Guam Reef Fish Stock Assessments

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf). Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

Pacific Islands Fisheries Science Center (PIFSC) scientists are conducting stock assessments on exploited coral reef fish species in the Pacific Islands Region which are listed in the Western Pacific Regional Fishery Management Council (Council) 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 size structure data 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/FMSY; 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 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 passed a rigorous independent review by a panel organized by the Center for Independent Experts in 2015, were recently (2017) applied to individual species in the main Hawaiian Islands, and now this general approach will be used to assess 20 species from the U.S. territory of Guam. Per WPSAR, there is a need to independently review these species-specific stock assessments prior to submission to a fishery management organization for consideration.

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 Council, PIFSC, and the Pacific Islands Regional Office (PIRO) have established the WPSAR process in an 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. CIE reviewers are being sought to participate in a peer review under this WPSAR framework: https://www.pifsc.noaa.gov/peer_reviews/wpsar/index.php. The specified format and contents of the individual peer review reports are found in **Annex 1**. The Terms of Reference (ToRs) of this peer review are listed in **Annex 2**. Lastly, the tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements

Two CIE reviewers are requested to serve as panel members (with a third, non-CIE reviewer serving as chair of the WPSAR panel) and conduct an impartial and independent peer review in accordance with the Statement of Work (SoW) and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of data-poor stock assessment models (preferably length-based assessment models) and general fishery stock assessment methods. They will also have familiarity with requirements of fishery stock assessments under the MSA, and will have familiarity with reef fish fisheries.

Tasks for Reviewers

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 two additional reviewers in accordance with the CIE conflict of interest policy. The panelists shall serve as independent and impartial scientific experts, and in their roles as reviewers they are not representing their respective institutions or affiliations. The panelists are expected to fulfill and comply with all elements specified in the ToRs. The panelists 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.

Prior to the Peer Review: Review the following background materials and reports prior to the review meeting. Two weeks before the peer review, the NMFS Project Contact will send by electronic mail or make available at an FTP site to the CIE reviewers all necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE on where to send documents. The CIE reviewers shall read all documents in preparation for the peer review, for example:

Benchmark stock assessment for review (not to be distributed beyond reviewers):
Nadon, M. O. 2016. (draft) Stock assessment of the coral reef fishes of Guam, 2017. U.S. Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-XX, ~200 p.

Relevant management information:
Western Pacific Regional Fishery Management Council. 2009. Fishery Ecosystem Plan of the Mariana Archipelago. Sections 4.4.2 and 5.6 only.
Western Pacific Regional Fishery Management Council. 2011. Omnibus Amendment for the Western Pacific Region to Establish a Process for Specifying Annual Catch Limits and Accountability Measures. Section 3.1 only.

References:
Hordyk, A.R., Ono, K., Prince, J.D., and Walters, C.J. (2016). A simple length-structured model based on life history ratios and incorporating size-dependent selectivity: application to spawning potential ratios for data-poor stocks. Can. J. Fish. Aquat. Sci. *73*, 1787–1799.
Kritzer, J.P., Davies, C.R., and Mapstone, B.D. (2001). Characterizing fish populations: effects of sample size and population structure on the precision of demographic parameter estimates. Can. J. Fish. Aquat. Sci. *58*, 1557–1568.
Nadon, M.O., and Ault, J.S. (2016). A stepwise stochastic simulation approach to estimate life history parameters for data-poor fisheries. Can. J. Fish. Aquat. Sci. *73*, 1874–1884.

Previous stock assessment:
Sabater, M, and Kleiber, P. 2013. Improving Specification of Acceptable Biological Catches of Data-Poor Reef Fish Stocks Using a Biomass-Augmented Catch-MSY Approach. Report of the Western Pacific Regional Fishery Management Council.

Supplemental Background Documents:

Hawaii assessment and independent peer review report:
Choat, JH, Franklin, EC, and Stokes, K. 2016. Benchmark review of the 2016 stock assessment of the main Hawaiian Islands reef-associated fish. Consensus panel report prepared by Erik C. Franklin.
Nadon, M.O. (2017). Stock assessment of the coral reef fishes of Hawaii, 2016. PIFSC Tech Memo 60. 200p.

References:
Nadon, M.O., Ault, J.S., Williams, I.D., Smith, S.G., and DiNardo, G.T. (2015). Length-based assessment of coral reef fish populations in the Main and Northwestern Hawaiian Islands. PLoS ONE *10*, e0133960.

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with this SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Contribution to the Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the ToRs of this review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 40 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor's facilities, and in Honolulu, HI.

Period of Performance

The period of performance shall be from the time of award through March 31, 2018. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	The NMFS Project Contact in consultation with the CIE provides the pre-review documents to the reviewers
February 5-9, 2018	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports

Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards
The acceptance of the contract deliverables shall be based on three performance standards:
(1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $8,000.

13 RESTRICTED OR LIMITED USE OF DATA

14 THE CONTRACTORS MAY BE REQUIRED TO SIGN AND ADHERE TO A NON-DISCLOSURE AGREEMENT.

NMFS Project Contact:
Beth Lumsden
Beth.Lumsden@noaa.gov
FRMD/PIFSC/NMFS/NOAA
1845 Wasp Boulevard., Bldg. #176
Honolulu, Hawaii 96818
808.725.5330

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. Each CIE independent report (and the consensus and individual reports of the Panel Chair) shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations.
2. The main body of each report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.
    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. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views. The exception is the Panel Chair's consensus report, which shall provide only consensus views or in cases where consensus cannot be reached, can provide majority views.
    c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.
    d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
    e. Each individual report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the consensus report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the consensus report.
3. The individual and consensus reports shall each include the following appendices:

Appendix 1: Bibliography of materials provided for review
Appendix 2: A copy of the CIE Statement of Work
Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Western Pacific Stock Assessment Review
Benchmark Review of Guam Reef Fish Assessment
Conducted in part using CIE reviewers

For questions 1-2 and their subcomponents, 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. For each individual species, review the application of the general approach for each of the following calculations. For each calculation, consider decisions points, input parameters, assumptions, and primary sources of uncertainty.
    a. Fishing mortality (F), spawning potential ratio (SPR), and corresponding overfishing limit (F at SPR=30%, aka F_{30}).
    b. Generation of overfishing limit from C_{30} (catch levels corresponding to F_{30}) distribution calculation.
2. Determine whether the results for individual species from question 1 can be used for management purposes under the Magnuson-Stevens Act and relevant Fishery Ecosystem Plan (FEP) 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.
3. 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 (3-5 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.
4. Draft a report (individual report from Chair and review members, and additional consensus report from Chair) addressing the above TOR questions.

Annex 3: Tentative Agenda

Tentative Agenda for Benchmark Review:
Assessment of Guam Coral Reef Fishes, 2017
Honolulu, HI 96813
February 5-9, 2018

Day 1 Monday February 5, 2018

1. Welcome and Introductions
2. Background information - Objectives and Terms of Reference
3. Fishery Operation and Management
4. History of stock assessments and reviews
5. Data
    a. Guam: Division of Aquatic and Wildlife Resources data collection
    b. Commercial Fisheries Biosampling Program
    c. Coral Reef Ecosystem Division surveys
    d. Biological data
    e. Other data
6. Presentation and review of stock assessment

Day 2 Tuesday February 6, 2018

7. Continue review of stock assessment

Day 3 Wednesday February 7, 2018

8. Continue review of stock assessment

Day 4 Thursday February 8, 2018

9. Continue review of stock assessment
10. Public comment period
11. Panel Discussions (Closed)

Day 5 Friday February 9, 2018

12. Panel Discussions (Closed)
13. Present Panel Recommendations (afternoon)
14. Adjourn

15 APPENDIX 3: PANEL MEMBERSHIP OR OTHER PERTINENT INFORMATION FROM THE PANEL REVIEW MEETING.

15.1 PRESENT (NAME, AFFILIATION)

1. Ivor Williams (NOAA PIFSC FSD)
2. Brett Taylor (NOAA PIFSC/JIMAR)
3. John Syslo (NOAA PIFSC FRMD)
4. Sarah Ellgen (NOAA PIRO)
5. Jarad Makalao (NMFS PIRO FSD)
6. Joseph O'Malley (NMFS PIFSZ LHP)
7. Benjamin Richards (NMFS PIFSC SAP)
8. Marc Nadon (NMFS PIFSC SAP)
9. Annie Yau (NMFS PIFSC SAP)
10. Marlowe Sabater (WPRFMC)
11. Chris Boggs (NOAA PIFSC FRMD)

15.2 REMOTE

None

15.3 REVIEW PANEL

12. Cathy Dichmont, Cathy Dichmont Consulting, Australia
13. Eric Franklin (chair)
14. Joseph Powers (CIE)