CENTER FOR INDEPENDENT EXPERTS (CIE) INDEPENDENT PEER REVIEW OF THE 2017 BENCHMARK STOCK ASSESSMENT FOR THE MAIN HAWAIIAN ISLANDS DEEP7 BOTTOMFISH COMPLEX

Cathy Dichmont Consulting

Center for Independent Experts (CIE) Independent Peer Review of the 2017 Benchmark Stock Assessment for the Main Hawaiian Islands Deep7 Bottomfish Complex

1 CONTENTS

1.2 LIST OF FIGURES

1.3 LIST OF TABLES

2 Executive Summary

The review workshop for the 2017 Benchmark Stock Assessment for the Main Hawaiian Islands Deep7 Bottomfish Complex took place in Honolulu, Hawaii from November 13 ‑ 17, 2017 with review panel members Drs Martell (chair), Dichmont and Sparholt (Center for Independent Experts (CIE) peer reviewers), NOAA scientists and industry members. The review documents and presentations were of a high standard, clearly describing the methodology and uncertainties associated with filtering the data, and undertaking the standardisation and the assessment. Several requests were made to the assessors and these were provided with great professionalism. The level of discussion and input from all parties present at the workshop, including those from industry, greatly enhanced the review process.

The work applied to the Deep7 bottomfish complex (which included the major species opakapaka) and opakapaka only. Unless stated otherwise, all the statements apply to the Deep7 and opakapaka approaches and assessments.

The work presented is of a very high standard, with substantial progress being made, particularly in cleansing and filtering the data. Being able to track fishers over time has greatly enhanced the database, which then allowed the standardisation to include a fisher effect. This factor was highly significant.

The below table summarises the response to each Term of Reference (ToR), which are supported in all cases, except using the base case for management. During the review workshop, an alternative base case was proposed and run. This proposed base case is supported. If this proposed base case is accepted, updating all components of the assessment report using the proposed base case is a short-term priority. In the medium term, further work on the independent survey, including vessel and technology data in the standardisation, investigating spatial effects in the standardisation or assessment and undertaking further work concentrating on multi-day trip definitions in the early few decades of the CPUE data is recommended. It is essential that the survey is undertaken annually and that as much of the CPUE data are available for the next benchmark assessment as is possible, so that the overlap of these two indices is as long as possible.

Table 1: Response and short-term changes summarised for each Terms of Reference

Terms of Reference	Response	Short-term changes
1. Are data filtering methods as decided upon by a series of regional community workshops correctly applied? Is the scientific uncertainty with respect to the input data quality and filtering methods well documented, including its potential effect on results?	Yes	
2. Is the CPUE standardization properly applied and appropriate for this species, fishery, and available data?	Yes	

3.	Are the assessment models used reliable, properly applied, adequate, and appropriate for the species, fisheries, and available data?	Yes	
4.	Is each model appropriately specified and configured?	Yes	
5.	Are decision points and input parameters reasonably chosen?	No	Yes, for proposed base case with new prior on survey catchability
6.	Are assumptions reasonably satisfied?	Yes	Add new prior for survey catchability
7.	Are primary sources of uncertainty documented and presented?	Yes	
8.	Are the final results scientifically sound, including estimated stock status in relation to the selected biological reference points and overfishing limits, and can the results can be used to address management goals stated in the relevant FEP or other documents provided to the review panel?	No	Yes, for the proposed base case with new prior on catchability
9.	Are the methods used to project future population status adequate and appropriately applied for meeting management goals as stated in the relevant FEP?	Yes	

4 BACKGROUND

The Main Hawaiian Island (MHI) Deep7 bottomfish fishery catch, amongst other species, is composed of deep sea snappers and a sea bass using deep-sea handline gear. There is both a commercial and a recreational fishery. The most extensive data are from commercial fisheries from 1948 onwards through the Fisher Reporting System (FRS) and, since 2000, sales reporting from the Dealer Reporting System (DRS). In 2016 an independent survey was completed.

A series of assessments have been applied to bottomfish. In 2011, the first Bayesian surplus production modelling was undertaken. This assessment expanded its exploration of different standardisation approaches when compared to previous assessments, and incorporated estimates of total fishery catch as previously. The assessment was for the Deep7 complex, rather than all species in the bottomfish complex, in MHI and the Northwest Hawaiian Islands (NWHI). In 2014, a similar assessment model was applied with an updated standardised CPUE series including fisher effects. A CIE review found the 2014 methods to be generally appropriate, but questioned the input data. The 2011 assessment with updated data was thus used for management purposes.

The present 2017 assessment has again undertaken a major review of the data, filtering approaches and standardisation. There is greater consideration and justification of model priors and updated life history parameters to produce estimates of natural mortality were included. New data based on bomb radiocarbon ageing improved estimates of bottomfish longevity and thus natural mortality. An independent survey was started in 2016 (after the last available commercial CPUE data) and is included as an absolute index of abundance in the model. In addition to a MHI Deep7 bottomfish assessment, a single species assessment of the most commonly caught species, opakapaka, is included. A Stock Synthesis model was attempted, but was not presented for review. Most key recommendations from the previous CIE review have been undertaken.

6 DESCRIPTION OF THE INDIVIDUAL REVIEWER'S ROLE IN THE REVIEW ACTIVITIES

The review workshop for the MHI Deep7 and opakapaka bottomfish assessment took place in Honolulu from 13 to 17 November 2017. In attendance were review panel members Drs Martell (chair), Dichmont and Sparholt (CIE peer reviewers), NOAA scientists and industry members, amongst others. Several highly informative presentations were provided on the fishery, management, data filtering, survey, standardisation and the assessment that were good summaries of the documents provided before the workshop (Appendix 1). These greatly contributed to our knowledge base. The Statement of Work provided to the review panel is provided in Appendix 2. A list of participants of the workshop are provided in Appendix 3 and the Agenda in Annex 3, Appendix 2.

During the review, several additional model diagnostics of the base case were provided. A few sensitivity tests were also requested and provided. Also, an additional model run with a proper prior for the survey catchability was run. The interaction between the review team and the review attendees was extremely informative and helpful. The reviewed scientists are thanked for their enormous effort in providing information and further runs, even after hours. This was greatly appreciated.

There was also a public session where the interaction with the review team was positive and informative, which was also appreciated.

A draft panel report was verbally provided on the last day of the review week.

Both a panel and individual reviewers report was required, which were required to address the following Terms of Reference (ToR):

1. Are data filtering methods as decided upon by a series of regional community workshops correctly applied? Is the scientific uncertainty with respect to the input data quality and filtering methods well documented, including its potential effect on results?
2. Is the CPUE standardization properly applied and appropriate for this species, fishery, and available data?
3. Are the assessment models used reliable, properly applied, adequate, and appropriate for the species, fisheries, and available data?
4. Is each model appropriately specified and configured?
5. Are decision points and input parameters reasonably chosen?
6. Are assumptions reasonably satisfied?
7. Are primary sources of uncertainty documented and presented?
8. Are the final results scientifically sound, including estimated stock status in relation to the selected biological reference points and overfishing limits, and can the results can be used to address management goals stated in the relevant FEP or other documents provided to the review panel?
9. Are the methods used to project future population status adequate and appropriately applied for meeting management goals as stated in the relevant FEP?

10. If any results of these models should not be applied for management purposes with or without minor short-term further analyses (in other words, if any responses to any parts of questions 1-9 are "no"), indicate:
    a. Which results should not be applied and describe why, and
    b. Which alternative set of existing stock assessment results should be used to inform setting fishery catch limits instead and describe why.
11. 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.
12. Draft a report (individual reports from each of the panel members and a Summary 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: ARE DATA FILTERING METHODS AS DECIDED UPON BY A SERIES OF REGIONAL COMMUNITY WORKSHOPS CORRECTLY APPLIED? IS THE SCIENTIFIC UNCERTAINTY WITH RESPECT TO THE INPUT DATA QUALITY AND FILTERING METHODS WELL DOCUMENTED, INCLUDING ITS POTENTIAL EFFECT ON RESULTS?

7.1.1 Overall response to questions

- Correctly applied? Yes
- Uncertainty documented and its potential effect on results included? Yes

7.1.2 Background

This is a complex fishery which has multiple datasets, each with its own characteristics, that need to be addressed before these can be used. The complexity reflects the fact that the fishery contains multiple fleets (recreational and commercial), multiple species (snappers, grouper in the Deep7 complex), multiple gear (deep sea handline, tuna line) and uses multiple ports over a reasonably large region. Like many fisheries, there have also been key changes to the databases, the most important being the inclusion of a Dealer-Reporting System (DRS) introduced in 2002 to support the Fisher-Reporting (FRS) system. Effort in hours are only reported after October 2000. Given that recreational catch and not all commercial catch has been recorded, there is an unknown unreported catch that needs to be taken into account. There is also historical confusion over whether the effort unit contains single or multiple day trips. Many attempts have been made to cleanse the data and put it in a form that would be applicable to stock assessments. However, these efforts have increased in the past few years in response to a CIE review in 2014 that questioned the utility of the data for management use.

Much progress has been made since the 2014 review, the most important are inputs from a series of workshops that included fisher stakeholders to help understand the data and resolve uncertainty issues. This process has improved the data and these efforts are commended.

7.1.3 Catch

Under-reporting of catch takes several forms; lost/released, under-reporting and misidentification of species. The decision points made in developing a total catch series are supported. Different scenarios are also included as sensitivity tests in the assessment model. The prior used in the model is discussed in ToR 5, including a recommendation to regularly update this information.

7.1.4 Fish weight (for average fish weight or size)

A series of steps were proposed for the calculation of total catch. These seem to be appropriate for the purposes of calculating average fish weight. The pros and cons of using the DRS or FRS database for mean weight is well analysed and discussed. Key to this discussion is the lack of location data in the DRS data, and indications that the NWHI mean weight is larger than the MHI data. Attempts to use the 1-fish records showed that the average of these are higher than the multiple fish records. Removal of zero pounds or fish number data is supported for this purpose. The final decision to use the FRS data based on a) the FRS data are a longer series while b) the DRS data do not add much more information for this purpose, and c) the lack of location data in the DRS data (which means NWHI data would be included), is supported.

Which species to use in the Deep7 mean weight calculation was extensively reported:

a) the concept of a deep7 mean weight is difficult to interpret, given the different life histories included in the complex and that the species composition has changed over time. The use of these data is most beneficial for individual species' assessments.

b) Fishers also reported that size based targeting would influence this calculation.

This analysis was done with care. Combinations of data including opakapaka only, opakapaka and onaga, and other combinations were investigated. Adding species to opakapaka had little impact to the mean weight calculations since 50% of all fish by numbers and 33% of all records were of opakapaka. Five options were discussed, and these were voted on in the workshops given there was not an immediate unanimous position. The final decision was to use opakapaka and onaga. All fish combined is supported although this decision should be regularly revisited when this information becomes more important to the benchmark assessment.

7.1.5 CPUE

The filtering of the data for CPUE standardisation followed a 4-step process:

Step 1 selected data records that were bottomfish records and the filters are all supported. The decision to keep the uku catch in the filtering process was shown to be correct given that this was a significant factor in some of the standardisation models.

Step 2 selects data records that comply with a unit of effort and the filters are supported.

During the workshop, the distance travelled frequency graphs were provided with smaller x-axis ranges and the catch frequencies were also provided. Discussions during the review workshop (especially comments from the industry) and investigation of these frequency plots hinted that there may be a trend over time in the distance travelled with early periods being dominated by slower boats (sampan's) whereas later smaller, faster boats started entering the fishery. The sampans may have travelled for a longer part of the day, and therefore multi-day trips could have been achieved with smaller distances as reflected by the single mode in the distance frequency plot. The second mode at about 25 nm in subsequent years could be the addition of the faster, smaller vessels. There is therefore a case for re-investigating a differential cut-off for the early years. This matters since, if this theory is correct, then the CPUE in the early period will be over-inflated, i.e. it is not a consistent bias over time. This uncertainty is not well reflected in the choice of a single cut off value.

Medium term recommendation 1. Re-investigate a different (shorter nm) cut-off for the distance travelled in the early years. As an initial step, investigate the influence of this as a sensitivity test. This matters since, if this theory is correct, then the CPUE in the early period will be over-inflated, i.e. it is not a consistent bias over time.

The catch frequency plots provided during the review supports the choice not to use the catch as a break point for multi-day trips as has been done in the past, because there is no clear differential modal structure in the plot.

Vessel size is reported to be recorded, but is only available on a database from another Division and was not available to the review. Adding this data may be extremely useful to the standardisation in general, but also may add more light to the filtering of single and multi-day trips.

Medium term recommendation 2. The author supports the recommendation that vessel size data be obtained from the vessel register database as a factor for the standardisation.

Step 3 selects the data that are representative of CPUE trends over time and the steps are supported. Significant work on adding fishers' names to the data is commended. The application of a cumulative experience factor in the standardisation is supported (especially given the standardisation results) and removal of these data would not have been supported by the reviewer as other standardisations have applied experience factors successfully.

The data that were from tagging and other scientific studies were not filtered in these analyses and future work should investigate undertaking this work.

Medium term recommendation 3. Remove tagging and other scientific records from the standardisation dataset

The list of potential factors produced in Step 4 are an informative first step for the standardisation.

7.1.6 Fishery Independent Survey

A fishery independent survey was undertaken in calendar year 2016 to obtain length-structured abundance and total biomass for use in stock assessment models. It was designed as a stratified random survey. The survey domain was mapped to the full bottomfish habitat from 75m to 400m depths. The sampling frame consisted of 500*500m primary sampling units (PSUs). The survey was stratified by substratum composition and substrate complexity. Samples were allocated among strata based on a Neyman scheme. PSUs within a stratum were randomly selected without replacement from a uniform distribution. At each selected PSU, Deep7 numbers and length-composition was obtained. Two main gear types were used – hook-line fishing and stationary stereo-video cameras. Hook-line fishing within a PSU ran for 30 mins using 1 vessel, 2 lines each line with 4 hooks and two bait types. The fork length (FL) by species of all fish was measured. Two randomised 15 min replicate camera video deployments were undertaken per PSU. The MaxN method was used to calculate species level counts and to measure FL.

The survey estimated relative abundance and biomass (using length-weight conversions for the MHI). The hook line method formed the majority of sampling, but there was overlap between some of the video and hook line sites to calibrate the two methods. The hook line estimates were converted to video through a single mean calibration value, ignoring the error in this calculation. During the workshop, the error distribution of this catchability measure was provided for opakapaka and showed that it is an important factor to be considered. The error in this catchability value should therefore be included in the analyses.

Medium term recommendation 4. Include the error in the calibration catchability between the hook line survey and the video. These can be bootstrapped as was shown during the workshop. These analyses should also be updated as new surveys are undertaken and more calibration sites become available.

To turn the now converted line and video densities into an absolute index of abundance, a video sample radius had to be obtained to produce a feasible range of the effective area sampled. Two different radius assumptions are made – a radius of 20.2m or 41.6m. Length-based simulation models were used to test which of these are more likely. These analyses showed that the estimated radius of 20.2m for the effective sampled area was more appropriate. This latter value was used to estimate total biomass.

There is a sizable difference in total biomass for the different radii, yet the fit of model 1 or model 2 to the size frequency (as a test) was not a good fit. There is therefore uncertainty in this catchability scalar. The error in this survey catchability value was not considered in the use of these data in the assessment. The assessment treated the single survey index as an absolute index of abundance with a point prior (i.e. perfect knowledge) in the survey catchability. Only the survey CV was considered in the assessment. Although, the catchability point prior seems to be only a scalar, this scalar affects the assessment results substantially (as was shown in the tests during the review). The use of the survey as an absolute index of abundance, without error in the catchability scalar, is not supported. In the context of a Bayesian assessment, a point prior is not usually accepted if the prior is known to be uncertain. This issue is further discussed in ToR 5.

Medium term recommendation 5. Continuing the surveys are a high priority. (HIGH)

Medium term recommendation 6. Investigating whether the survey can be used as an index of total abundance is a high priority. The present investigations of using 360^{o} camera with the stereo camera is supported. (HIGH)

Medium term recommendation 7. The survey report's recommendation to further investigate alternative approaches to the MaxN method used to estimate density from video footage (the MaxN method can bias density estimates as it is non-linearly related to true-length-structured abundance) is supported.

7.1.7 Summary

- The large effort, especially with support from industry, that has been made to improve the data quality has greatly improved the data and is commended.
- Uncertainty is well investigated and documented in the workshop report (which is well set out).
- The issues highlighted above do not detract from using the data in the assessment in its present form.

- Further research on the catchability of the survey is essential for the continued use the independent survey as an absolute index of abundance (but see Terms of Reference 5 for further discussion on this).
- Over time this survey would be of great value, even as a relative index of abundance and should be continued.

7.2 ToR 2: Is the CPUE standardization properly applied and appropriate for this species, fishery, and available data?

7.2.1 Overall response to question

Yes.

7.2.2 Background

The data are zero inflated - zero catch data are correctly included in the dataset since it is possible to have a zero bottomfish catch despite targeting bottomfish. 17% of the total record are zero catches. A delta-normal model was used, although a Poisson and negative binomial distribution model was also considered, but was not used. The former because the overdispersion parameter was very large, and the latter because it had conversion issues. The Delta-normal is a standard approach to standardizing zero inflated data of this kind and is supported.

The Bernoulli process that modelled the zero catches used a binomial distribution with a logit function. The positive catch process assumed a lognormal process. Predicter variables included categorical (Year, Area, Region, Quarter, Cardinal and Ordinal wind direction) and continuous (sqrt catch of uku, cumulative fisher experience, wind speed) variables. Fisher was added as a random effect – all others were treated as fixed effects. Two interaction terms were added, being Area*Year and Area*Quarter. Because of non-linearity between wind speed and positive CPUE, an additional term of the square of wind speed was added.

The factors tested were appropriate. There are several other factors highlighted in the workshop report that could be included over time, especially the oceanographic and decadal influenced factors (e.g. El Niño). However, these need to be included cautiously to ensure they are "interpreted" appropriately in the standardisation, i.e. as changes to availability or recruitment/biomass, whichever is appropriate.

However, the absence of technological factors (they are of course presently not available) is of importance. For this fishery, it is likely difficult to incorporate this information, but this should be attempted, given the likely significant influence over time this would have on fishing power. The separate database containing vessel information may provide a way forward.

Medium term recommendation 8. Investigate the inclusion of oceanographic factors and other factors mentioned in the workshop in the standardisation. Explain why factors mentioned in the workshop are not included in the standardisation.

Medium term recommendation 9. Investigate the inclusion of technology changes such as GPS, plotters, sounders and electric reel types. Several studies have already started on this. If these are not known per fisher, it may be more achievable per

vessel. Although offsets in the standardisation for these factors may be fairly arbitrary, they may need to be investigated in a sensitivity test. Arbitrary values should not be added outside the standardisation without adequate justification.

A forward selection process was applied using an 0.05% AIC threshold rule. The random effect, fisher, was added first and the fixed terms thereafter. Year was always included in the final model as it is required for the standardisation. The R library *lme4* was used. Model selection was done using maximum likelihood, but for the generalized linear models where a restricted maximum likelihood was applied. This approach is appropriate.

The two CPUE series (before and after 2002) were standardised separately. The decision to switch to two different CPUE sets is appropriate. However, given the sensitivity of the model to the survey index it would have been more appropriate to also test a CPUE index as a single series in days.

The total effort in days of time is in itself a useful measure to check the model against, as it is the only consistent effort series over time. As an example, the check against the annual catchability model sensitivity test undertaken during the review was informative (see Term of Reference 6 for further discussion on this topic).

Medium term recommendation 10. Keep the days effort series over all years for use in checking the assessment model output.

Medium term recommendation 11. Run the model with a sensitivity test with the CPUE series as a single series in day effort units (this may be possible in the short term, which can be used as a check on the influence of the survey on the overall CPUE catchability and observation error as this model is simpler).

Wind speed and wind direction were not available in the full early period, so were not included. The sqrt(uku) could not be used in the Bernoulli model because of the filtering process. Fisher was an issue in the Bernoulli model with memory and convergence issues. Fisher reduced model deviance the most among all the predicters in all the lognormal models. The Year*Area term produced errors in several of the models.

The reason for the Year*Area interaction term producing errors should be investigated. At times the solution could be simply combining some of the marginal areas. On the other hand, this error could be highlighting a spatial expansion/contraction of the fishery that is significant and should be considered either within the standardisation or within the model. Even if there are no solutions, these investigations often highlight spatial expansions and contractions of the fishing area over time, and also the impact of closures such as BRFAs. These changes in the area of the fishery could affect the assumption of a linear and proportional relationship between standardised CPUE and biomass.

Medium term recommendation 12. Investigate the reason(s) for the error in Year*Area interaction term. This is a high priority. (HIGH)

Model diagnostics relied on a) visual comparison of residuals plotted against predictor variables of the response variable and b) against predicter variables. C) Pearson residuals for the lognormal process, d) quantile residuals for the Bernoulli process, and e) quantile plots for the lognormal process were also provided. These are the usual list of performance statistics one would investigate and is appropriately used and interpreted. These show that the model fits are generally good to very good with a few exceptions:

- Some sign of heteroscedasticity in the residuals of the Bernoulli processes of the early (mostly) and later (less obvious) Bernoulli process models,
- Some deviation from normal in the early and late lognormal model, and
- Residual patterns in the sqrt(uku) variable.

None of these issues would discredit the models, especially since the obvious approach of addressing these issues by using a gamma distribution with log link function model was unable to converge with fisher as a random effect. Also, when a fixed effect only model was compared with a gamma distribution model, the residual patterns were not resolved.

For the opakapaka only standardisation comparable results are obtained, except that the Area*Quarter interaction terms are not included in the final Bernoulli process model, sqrt(pounds of uku) for the early CPUE series was available and significant in the early Bernoulli process and sqrt(uku) was not significant in the early lognormal process. The sqrt(uku) showed fairly strong residual patterns. Again, none of these issues would mean rejection of the standardisation.

Index calculation from a delta-lognormal model is more complicated than a standard GLM. The approach taken was appropriate.

To undertake a standalone, single species assessment, a single species effort series needs to be derived from what is essentially the effort unit derived from the whole fishery. This is often rather difficult to do in multi-species fisheries. In the Deep7 bottomfish fishery, fishers do not report targeting, so this has to be inferred from the data. In the case of the opakapaka assessment, the approach of Stephens and McCall (2004) was used, where a logistic regression using the *glm* library in R (R Core Team, 2016) was applied. In this approach, the presence/absence of a non-target species (defined as the species that contribute to 99% of the cumulative catch) was used to predict the probability of catching opakapaka. A backward model selection was undertaken, which resulted in five (5) of the 38 species being insignificant. The critical value was calculated as per Stephens and McCall (2004), so that the predicted probability of catching opakapaka is greater or equal to 0.51, i.e. assumed to be opakapaka targeting. These data were then used to calculate the catch, effort and CPUE series.

The approach of Stephens and McCall (2004) is reasonable. Other approaches have also been attempted and should be investigated. These include TREE models (or its relative random forests). An example of a TREE model being used to split one species from a cluster is used to separate the targeting of banana prawns from tiger prawns in the Northern Prawn Fishery, Australia (www.frdc.com.au/research/Final_Reports/1998-109-DLD.pdf). Although this a simpler case, TREE models are very powerful for data mining in general, and have the property of letting the data "speak for itself". This approach is likely to be informative in general, especially if environmental covariates are added, not only to separate out a species.

In the same fishery, the tiger prawn fishery was further separated using generalised linear models and generalised additive models (GAMs; see Venables and Dichmont, 2004a with further examples in Venables and Dichmont, 2004b). GAMs are especially informative where there are space-time changes in distribution, as is likely the case for bottomfish. GAMs are very flexible (a strength and a weakness), but a powerful tool. The rug of the GAM is also very useful in showing where the data limitations occur.

Finally, it is recommended that these analyses are undertaken, including the presently used approach using a training and test set (i.e. predict the selected model from the training set onto the test set). This allows additional performance statistics to be calculated, including time and spatial trends to the predictions.

Medium term recommendation 13. Consider alternative approaches to targeting opakapaka (and for other new single species assessments) such as TREE (or its relative random forests) and generalised additive models (GAMs). The latter would be useful for investigating space-time changes in targeting.

Medium term recommendation 14. Test the different targeting approaches using a training and test data set. This will allow additional performance statistics to be applied.

7.2.3 Summary

- The standardisation was undertaken with care through attempting several zero inflated approaches and testing several factors.
- Although there are some negative results in the standardisation statistics, these are not of a scale to affect the conclusion that the selected models are the most appropriate.
- As a result, the final standardisation model is properly applied and appropriate for the species complex, fishery and available data.

7.3 TOR 3: ARE THE ASSESSMENT MODELS USED RELIABLE, PROPERLY APPLIED, ADEQUATE, AND APPROPRIATE FOR THE SPECIES, FISHERIES, AND AVAILABLE DATA?

7.3.1 Overall response to the question

Yes.

7.3.2 Background

The assessment model is a Bayesian state-space surplus production model using a modified Pella-Thomlinson form, i.e. it has the additional shape parameter.

This model does not rely on good life history and field data (e.g. age structure over time) as most other models do, and is therefore appropriate given the lack of this kind of data and the complexity of choosing appropriate life history parameters for a complex. Surplus production models are more likely to be useful in long lived species (as applied here), than short-lived species which are highly recruitment-driven and recruitment variation is large. Although other models could be considered, the use of the surplus production model is appropriate.

The choice of the additional shape parameter is usually applied when a more data moderate index of abundance is available, and assumptions of production shape cannot be justified by independent data sources. The inclusion of this model parameter avoids making assumptions about the productivity of the resource. This is therefore accepted as good practice when the data are available to support the estimation of this parameter.

The model is fit to two separate indices of relative abundance from the commercial fishery (CPUE), a single absolute index of survey abundance and (reported and unreported) catches. Using the separate CPUE series is appropriate, but as recommended in previous sections maintaining a single series as a test is recommended.

Both observation and process errors are included in the model. The observation error is related to the discrepancy between the observed and predicted CPUE. The process error is added to the deterministic dynamics of exploitable biomass so that the model can include natural variability from e.g. the environment and/or variation in growth. These are assumed to be random multiplicative events, which is a standard assumption. The use of both observation and process error is appropriate given the data available. It is arguable that the process error needs to be included in all cases, but given the variability in the CPUE index in the early years, its inclusion is important for this assessment.

A key feature of the model is that the dependent data finished in 2015, whereas the independent survey was undertaken in 2016. For the model to run to calculate the proportion of biomass in 2017, the 2016 catch is added, but no (relative) abundance data are available for 2015/6. The year 2015 is used for management purposes. The inclusion of the survey data is a key feature of the model results and are discussed in ToR 5 and 8.

The CPUE standardisation started in 1949, but this year was removed as it is incomplete from the stock assessment point of view. This makes sense as the use of a biological year is appropriate.

The CPUE index is assumed to have lognormal errors where the observation error random variables variance term is weighted by an annual weighting factor derived as a relative CV to the minimum CV from the standardisation. The fishery-independent survey is assumed to be lognormally distributed, and its observation error variance term is determined from the survey CV directly. This implied relative weighting using the data is appropriate.

The model estimates the surplus production parameters (R, K, M, P1 – the latter being the proportion of the first year, 1949, to the carrying capacity K), the CPUE catchability parameters (q_1 and q_2), the variance term (σ_2) of the process error random variables, and CPUE observation error variance ($\tau 1$, τ_2).

The model is able to calculate the MSY-related reference points when all parameters are estimated. The Bayesian model would also provide posteriors of these management quantities.

7.3.3 Summary

- The use of a Bayesian surplus production model is appropriate for the species, fisheries and available data.
- The mathematical set up is properly applied, including the use of the modified Pella-Thomlinson surplus production form, although note later comments on the shape parameter.
- These model types are more reliable for medium to long lived species, as is the case here.

7.4 ToR 4: Is each model appropriately specified and configured?

7.4.1 Overall response

Yes.

7.4.2 Background

The model specification is that of an observation and process error Bayesian surplus production model with two separate standardised CPUE series and an absolute index of abundance. The standardisation is implemented in R and the surplus production model is implemented in R2WinBUGS, which runs the model in WinBUGS. WinBUGS is a well-known model for running Bayesian models and the R2WinBUGS library is all well accepted. These software platforms make sense and, on the face of it, would seem appropriate.

The formulation of the shape parameter of the model assumes that the surplus production shape is not required to go below ~$0.36B_{MSY}/K$ (Figure 1), unless some interesting priors are set up. The mathematical formulation of the production function means that it becomes numerically unstable as the shape function M tends towards zero and undefined at zero (because it is dividing by zero). The present formulation of the prior is such that the model is unable to choose values nearing $0.36B_{MSY}/K$. Given the new life history parameters, the new maximum ages of some of these Deep7 species would make one feel comfortable with this assumption. Furthermore, the posterior has tended to shift the shape parameters to larger values away from this threshold. On the other hand, the chain provided during the review workshop did show that some values tended towards the threshold, so these could not be excluded.

Importantly, a more objective approach (rather than using past assessments) of selecting an M prior would be to use a meta-analysis undertaken by Thorson et al. (2012), which shows that M prior values would be those that satisfy Bmsy/K ratios with a mean of 0.353 and a standard deviation of 0.114 for species in the order Perciformes. This prior could not easily be implemented in the present formulation of the model. The model structures of that within the Thorson et al. (2012) paper have much greater flexibility of possible M values and this approach could be applied to the present model. This paper provides one possible solution for having a viable model that can select any M parameter given the data and priors.

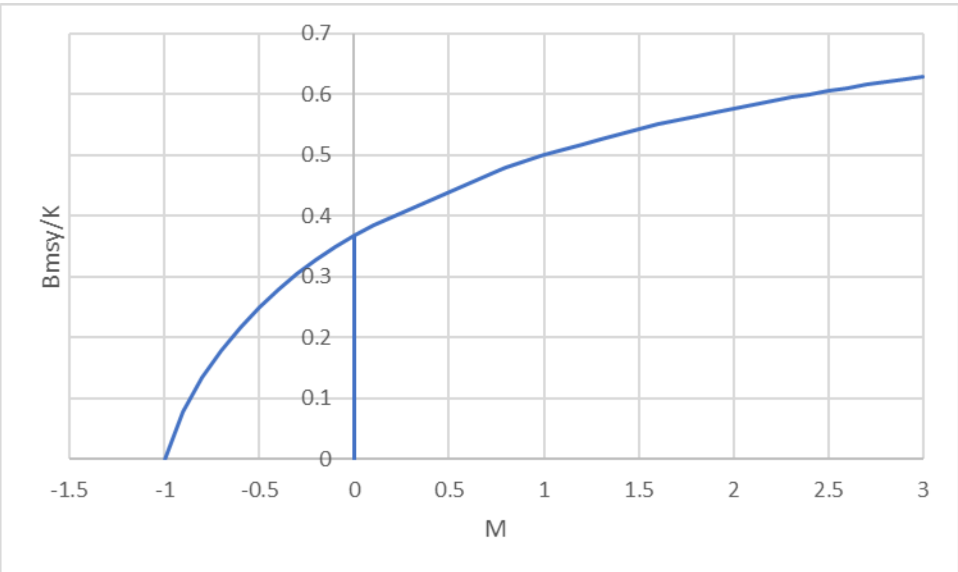

Figure 1. Resultant Bmsy/K value for a specific example formulation showing the effect of changing M. Note an M value is not possible as it is numerically undefined (dividing by zero), so in this plot a very small M was used.

Medium term recommendation 15. Reformulate the surplus production form to that of Thorson et al. (2012) to test whether the implied Bmsy/K 's are similar to those obtained in the present model formulation.

The formulation of the projection code is discussed in ToR 9. However, the inability to project with process error due to WinBUGS not being able to separately simulate projections (without affecting the optimisation) is not appropriate. Several alternative software platforms are suggested in ToR 9.

7.4.3 Summary

- Each model is appropriately specified, except for two issues which are important to address in the medium term
    - The model is unable to project with process error using the WinBUGS platform. This is a good case for updating the model platform to something more modern and appropriate.
    - The model is unable to select a shape parameter that conforms to the suggested priors from a meta-analysis of Perciformes and other orders of fish. A suggested way forward is to port the concepts of the Thorson et al. (2012) paper into this model.

7.5 ToR 5: Are decision points and input parameters reasonably chosen?

7.5.1 Overall response to question

No, for the existing base case under its current formulation.

Yes, for the proposed alternative formulation.

7.5.2 Priors

Prior distributions are determined for R, K, M, P1, q_1, q_2, τ_1, τ_2, σ_2 and C_U. No priors were placed on the survey index variance term which was treated as an absolute index of abundance. Informative priors were applied to the surplus production parameters (R, K, M, P1). The R, K and P1 parameters are drawn from lognormal distributions, whereas the M value is from a gamma function. The R prior distribution was set up based on Musick's (1999) recommendation with a CV of 25%, which produces a 95% confidence interval that approximates (on the log scale) the suggested range of 0.05 to 1.5 yr^{-1} (based on the assumption that the Deep7 are low productivity species as demonstrated by their longevity). The prior of K was set based on the 2011 benchmark assessment and the R range. Assuming that the product of R and K parameters is similar in the 2018 and the 2014 assessments, the mean of the K distribution given the mean R is 0.1 can be calculated. The variance was set up to achieve a CV of 50%. The prior of M was described by a gamma distribution. As described above, this prior should be based on Thorson et al. (2012), but would require recoding of the model. The prior is set up to assume a Schaeffer model with a CV of M of approximately 140%. Care has been taken to find some basis for the informative priors and generally these are accepted.

The prior on the catchability parameters and unreported catch error assumed an uninformative uniform distribution. A uniform, mostly uninformative, prior was also applied to the total catch i.e. the unreported catch error. It was assumed the error in the unreported catch was uniformly distributed about a point estimate with a ±40% error where it is truncated (the latter value comes from the Hawaii Recreational Fishery Survey using data gathered from 2004-2016 and relates to the yearly mean estimates of the percent opakapaka, onaga and ehu that was calculated as not sold). The unreported catch is a significant part of the assessment in that it assumes a very large unknown catch, especially for opakapaka where it has been between 2 and 3 times as much as the reported catch. The approach taken here with including a truncated uniform prior distribution on the total catch and undertaking sensitivity tests where the scale and pattern of the unreported catch ratio changes is appropriate and well executed. Although no update was undertaken on these values since now new information is available, regular updates would be needed. Furthermore, an education program for fishers to log their catch should be progressed, since this is a significant unknown component of the assessment and can also have serious consequences on management if incorrectly calculated.

Medium term recommendation 16. Regularly update the unreported catch ratio. Given no updates were undertaken for this assessment, it is recommended this work to be undertaken before the next assessment.

Medium term recommendation 17. An education program should be implemented that attempts to increase participation in recording catches.

The priors for the process and observation error variance used moderately informative inverse-gamma distributions setting the gamma parameters >0. The central tendency of the observation error variance prior was set up to be tenfold greater than the process error variance prior. Given the uncertainties in the data and the life history of the species, these seem appropriate. There remains a difficulty for the assessment in that the early period of the CPUE data is highly variable. This variability disappears from the data from the latter part of the early CPUE series and from the whole recent CPUE series. There is little information in the data to explain whether this variability relates to changes over time in the catchability or the biomass, which would affect different parts of the observation and process error components of the model. Either way, the surplus production model formulation (even with observation and process error) would struggle to fit these data without additional information, e.g. age or length structure, or other standardisation factors.

Priors for P1 were determined using a 2-step process. Firstly, the prior mean was increased in increments of 0.1 from 0.1 to 1 as per an empirical Bayes approach. The value of 0.4 was shown to minimise the sum of the root mean square error (RMSE) of the fit to the CPUE indices. This result was then used to set up an informative prior for the full estimation stage. A mean from the 0.4 prior run with a CV of 40% (given the flatness of the profile) was used.

The previous assessments used 20% CV, which implied a narrower range of P1 than implied by Step 1, whereas the present results pointed to the need for a bigger CV of 40%. Technically, there is a circularity in the empirical Bayes approach to calculate P1, i.e. using the data to estimate a prior and then applying the same data and resultant prior to estimate the full model posteriors. This approach does not conform to hardcore Bayesian methodology. However, it is unclear how else this prior would be determined and is thus supported. Some discussion during the review with industry as to how much fishing had occurred prior to the data series may indicate higher P1 values, but a conservative value is more precautionary. A CV of 40% is reasonably wide.

7.5.3 Indices of abundance

An independent survey is implemented in the model as an absolute index of abundance with perfectly known catchability. This is not appropriate given the error in both the conversion factor of the hook line sites to video, and the scalar from relative to absolute indices of abundance. In the calculation of the index, neither of these errors were considered, particularly not the scalar. This means that the survey catchability is included as a point prior (i.e. known exactly), and therefore does not conform to standard Bayesian methodology, given the uncertainties. This is an important assumption, in that this scalar is an implicit form of weighting between the CPUE indices and the survey. Sensitivity tests provided showed that the inclusion or exclusion of the survey has a profound effect on the results, meaning that the model is very sensitive to the relative weighting of these two indices. Since there is no real basis for the base case weighting, the base case is not accepted.

It is important that all sources of variance (the conversion from hook to video) and bias (the survey catchability parameter) are fully included in the analyses. To investigate the variance of the hook to video conversion factor, the frequency of the conversion factor was provided showing upper and lower 95% confidence intervals of about 0.23 and 0.53 respectively, with a median of 0.35 for opakapaka (RF-MOUSS GCF Hist (Paka).pdf provided during the review). The mean value was used. The more appropriate approach would be to bootstrap these in the calculation of the video relative index when converting the hook data to video equivalents. Since it was not possible to undertake this full analysis during the workshop as it would have to be undertaken for each species and then summed to the Deep7 species, this issue was not addressed.

Medium term recommendation 18. Include the error in the hook to video conversion factor in the calculation of relative individual species and Deep7 biomass.

The survey catchability was addressed in the review workshop, through the addition of a prior to the survey index. This survey catchability was based on developing a moderately informative prior for the survey radius of effective area, which is the key unknown in the catchability calculation. A maximum (41.6m) and minimum (7.5m) radius is known for the survey, and the previous mean of 20.2m could be used. The prior can then be set up assuming a lognormal distribution and a CV of 0.5. This approach was correctly implemented during the review workshop. The proposed base case results are discussed further in ToR 8.

7.5.4 Summary

- Priors are generally well chosen.
- Although the empirical Bayes approach is strictly not correctly applied, no clear alternatives are available, and this prior does not greatly restrict the final result.
- The unreported ratio was not updated for this assessment, but should be done for the next benchmark assessment as it would be out of date.
- Assuming the survey is an absolute index of abundance is not appropriate, given the uncertainties in the conversion from hook to video indices and in the overall survey catchability.
- The new prior on the survey catchability is a much more appropriate methodology and this model test should be used for management, rather than the present base case.

7.6 ToR 6: Are assumptions reasonably satisfied?

7.6.1 Overall response to question

Yes.

7.6.2 Main model assumptions

Although there are many assumptions in the model, most are discussed in other ToRs (e.g. the priors in ToR 5), but a few key ones are discussed here as well.

1. Biomass production follows a specified form

The modified form of the Pella-Thomlinson model means that the degree of resource productivity is not assumed outside of the model, as it would be if a Schaeffer or Fox shape was chosen. Some discussion of this shape (M) parameter in terms of the prior and the mathematical formulation can be found in ToR 4. An alternative (more data rich) model choice would be that of an age or length-structured model. The assessment report stated that a Stock Synthesis model was attempted, but could not be completed for the review workshop. Moving to a SS model for the Deep7 assemblage would bring up the issue of which biological information to use. At this stage, at least for the Deep7 group, the choice of a surplus production model makes sense. As a result, the surplus production model with its associated assumptions is supported. Moving to a full age based model, given the degree of unreported catch and the high variability in the early data would need to be undertaken with great care. On the other hand, a delay difference model (Deriso, 1980; Schnute, 1985; Schnute, 1987; Fournier and Doonan, 1987) is a good compromise between a surplus production and a full age model. This model will evolve over time as the survey data are added.

Medium term recommendation 19. Investigate the use of a delay difference model to the Deep7, but particularly the opakapaka stock assessment.

2. Standardised CPUE is proportional to biomass (and therefore catchability is constant)

This assumption is common to almost all stock assessments. These assume that the CPUE standardisation has removed as much of the fishing power bias from the series as possible, and the CPUE index reflects proportional changes in biomass. However, the factors available for the Deep7 and opakapaka standardisation did not contain much information on vessel characteristics and their associated technology changes over time (beyond what could perhaps be bundled into the Fisher effect). For example, the fishery initially consisted of slower sampan's which undertook more multi-day trips and were able to fish in poor weather conditions. These were gradually replaced by modern, faster vessels that were often smaller and undertook mostly shorter, better weather trips. Furthermore, the kind of reels changed to more automatic types which meant a quicker turnaround of the gear at these depths. GPS and plotters were added and were likely to have been influential on fishers' effectiveness, similar to other fisheries. Fisheries and marine closures removed some of the available grounds; and the fishing grounds changed over time in the move from sampans to modern boats. Together these changes in technology and fishing areas are not likely to be included in the present analyses given the terms used (as discussed in ToR 2).

During the review workshop, a scaled plot of the difference between the harvest rate and the effort in days over time were plotted with the catchability values estimated by the random walk catchability scenario. The difference plot is similar to a series of catchability deviations which can then be compared to the random walk catchability values. There are many explanations for this plot, including unaccounted for biomass changes not picked up by the process error components. However, one possible interpretation of these two plots are that a) in the earlier period (1950s and 1960s) multiday trips were defined as single day trips, and b) technology changes were not captured in the standardisation from the mid-1970s. Certainly, there are hints of factors not considered in the standardisation. However, this interpretation cannot be distinguished at this stage from the data provided.

Despite these issues, the standardisation was a major improvement from past assessments given fisher factors could be included and the data themselves are rigorously filtered. Furthermore, a sensitivity test of estimating catchability within the model (above that included in the standardisation) was undertaken and the impact of changing the constant catch assumption could therefore be demonstrated.

3. All the exploitable biomass is mature and available for capture

This is an assumption common to surplus production models. Since the captured animals tend to be fairly large and the age at maturity is only a few years, this assumption is not likely to be breached often.

4. The independent index of abundance that represents total biomass

This issue is extensively discussed in ToR 5 and 8. This assumption is not supported while the survey catchability was treated as a point prior. The proposed base case formulation is supported.

5. Process errors are random and multiplicative events which can be described using the Central Limit Theorem to be independent and lognormally distributed random variables.

Since many surplus production models ignore process error, this assumption is unlikely to be breached and is an improvement on not including process error.

7.6.3 Summary

- The surplus production model has several basic assumptions that are appropriate given the available information.
- The base case form of the survey index is not supported, but the proposed base case is supported.
- The standardisation is a major improvement on previous versions, but there are still indications that the index may not have captured or resolved all issues, especially multi-day trips in the earlier part of the CPUE series, vessel and technology effects and spatial changes. Work on these aspects should be a priority.

7.7 ToR 7: Are primary sources of uncertainty documented and presented?

7.7.1 Overall response to question

Yes.

7.7.2 Uncertainty in data filtering

Primary sources of uncertainty in the data filtering are well presented and documented. Some effort should be made to investigate and/or better explain why certain analyses did not work.

7.7.3 Uncertainty in standardisation

The sources of uncertainty in the standardisation process are well presented and documented. Some effort should be made to investigate and/or better explain why certain analyses did not work. This is particularly important with respect to why the Year*Area term did not converge (as discussed in ToR 2).

7.7.4 Uncertainty in the surveys

Although uncertainty in the survey catchability is not addressed within the model, these are well documented, presented and discussed.

7.7.5 Assessment uncertainty

Uncertainty in the assessment is also well documented and presented, mainly in the form of a large array of base case model fit diagnostics, prior and posterior distribution plots and sensitivity tests. A full table of the sensitivity tests with their impact relative to the base case was provided in Table 16. This was very useful to highlight the relative importance of the different tests. The ratio of B_{2015}/B_{msy} and H_{2015}/H_{msy} would have been useful as these can often be less sensitive than the values in their independent form. However, as the survey is added with a known catchability prior, these management-related outputs are likely to become very sensitive as the model has to "thread the needle" in the last year modelled irrespective of the biomass journey taken to this point.

Short term recommendation 1. Add B_{2015}/B_{msy} and H_{2015}/H_{msy} in the sensitivity test summary Table 16.

During the review, pair plots of key parameters were requested so that correlations between parameters could be investigated. These should be a standard inclusion of future assessments.

Medium term recommendation 20. Include pair plots of key parameters in future assessments.

During the review workshop, the chain samples for Bmsy/K were requested. It is always good as an Appendix or a separate report to provide some visual of the chains.

Medium term recommendation 21. Provide chain figure of key of each sample similar to that requested in the review workshop for key derived parameters, especially those that approach limits.

The empirical Bayes approach of re-calculating the P1 priors was not carried out for the sensitivity tests. Technically this should have been done. A test undertaken during the workshop by changing the R prior showed that the P1 prior value would not have resulted in changing the P1 prior. However, when the P1 prior was re-estimated for the proposed base case, the results indicated that the P1 prior needed to be changed. This shows that redoing the P1 empirical Bayes approach would be required for the sensitivity tests. It is understood that this is a time consuming process, but may be less so within a different programming platform (as recommended in ToR 9).

Medium term recommendation 22. Technically, the empirical Bayes approach of setting the P1 prior should be undertaken for each major sensitivity test. Investigate this option, especially when using a different and faster modelling platform.

Retrospective analyses were provided, but only from the terminal year of 2015 or earlier. These show the sensitivity of the model to new CPUE data. However, the test should have included the base case assessment, since it would highlight the large change between the 2015 CPUE only assessment and the base case.

Short term recommendation 2. Add the proposed base case to the retrospective analyses in Figure 20.2.

7.7.6 Projections

In previous assessments, projections were only undertaken for one year. A multi-year projection is much more appropriate. The results are presented and discussed well, including an open and transparent discussion of the issue of projecting without process error.

7.7.7 Summary

- Data filtering, standardisation and assessment documentation was of a high standard.
- Most primary sources of uncertainty were documented, presented and discussed.
- Only minor additions are suggested in the short and medium term.

7.8 ToR 8: Are the final results scientifically sound, including estimated stock status in relation to the selected biological reference points and overfishing limits, and can the results can be used to address management goals stated in the relevant FEP or other documents provided to the review panel?

7.8.1 Overall response to question

No, using the current base case.

Yes, using the alternate base case.

7.8.2 Base case versus alternative base case biomass and harvest rate

The proposed base case was run during the workshop (as an unchecked draft). The empirical Bayes calculation for P1 was recalculated and showed a different but clearer minimum value and RMSE shape, so a new mean and smaller CV of 20% was chosen. The radius posterior of the survey catchability estimated a larger mean value than that used in the base case. The full results were provided to the review panel. In brief, these showed that the proposed base case:

- Suggested a biomass between that of the base case and the sensitivity test without the survey (although closer to the no survey result) (new Fig 13 and other Figs provided). As expected, the confidence intervals of the proposed base case are wider than the base case.
- A similar trend for the harvest rate is shown, i.e. the results are closer to the no survey test.

The new projections were also undertaken, but with fewer samples given available time. As expected, the management-related results change.

Short term recommendation 3. Update the projections, check the proposed base case input and results, and use this model for management. (HIGH)

7.8.3 Reference points

The minimum stock size threshold (B_{MSST}) was calculated from natural mortality which was derived from maximum age calculations based on growth studies. Past studies in Hawaii and other regions have found a wide range of von Bertalanffy growth parameters when traditional otolith ageing techniques were applied. In these studies, the maximum age of opakapaka could be about 7 or 18 years depending on the study. However, Andrews et al. (2012) and Andrews (pers. Commn as referred to in assessment report) have used bomb radiocarbon dating to re-analyse the Deep7 species' growth parameters. Opakapaka, hapuupuu and onaga maximum age was shown to be 43 years, 53 years and 54 years respectively. The natural mortality estimate using the study of Then et al. (2015) tmax estimator produced a value of 0.156 (used for this assessment), which is much lower than 0.25 and 0.3 values used in previous assessments. These new values are appropriate.

A surplus production model can estimate MSY-related reference points. In this model's Bayesian context, implied prior values are shown as well as the derived posteriors for the B_{msy}, MSY and H_{msy}. As one would expect, the MSY reference values have changed between the base case and the proposed base case. These changes are as one would expect, given the new posteriors.

The stock status for the proposed base case (new Fig 13 and 15) is more optimistic than the base case. This is because the survey is essentially more down-weighted relative to the CPUE data in the proposed base case than when the index is treated as absolute with a point prior. The stock status remains not overfished.

The resource also remains as a harvest rate status of not being overfished (new Fig 14 and 15).

The proposed base case is supported for use in management.

Medium term recommendation 23. Continue the expansion of single spp. assessments for the key Deep7 species.

7.8.4 Species assemblages

There are some issues with managing a complex of species within a species group. The most important of which is that minor, but sensitive species in the group can be declining and this could be masked in the overall assessment (e.g. Kleiber and Maunder, 2008). Some agencies have managed using the lowest common denominator, rather than as a group. However, this means that the other species are likely to be underutilised. There are pros and cons to each system. The obvious outlier to the Deep7 group is the grouper, which changes sex and is very slow growing. Trying to undertake a separate assessment of this species, or keeping track of its stock status using more data limited approaches (e.g. mean length approaches) may be worth considering in the interim until the survey has been run for longer.

Longer term recommendation 1. Undertake separate analyses for the grouper (*Hypothodus quernus*), which may be more vulnerable within the Deep7 cluster. This may require additional life history work.

7.8.5 Model fit

Neither the base case nor proposed base case support the view that they are robust. However, the proposed base case increases the uncertainty in the model, because of the inclusion of a prior on survey catchability. This proposed base case is therefore much more reflective of the associated uncertainty in the assessment and can be used for management. The present base case is not appropriate for use in management.

There are key aspects to the assessments that should be considered:

- There are high degrees of correlation among some of the parameters (new extra Fig 1 and extra Table 1). These are not uncommon in surplus production models and the degree to which this occurs in the Deep7 and opakapaka models is as one would expect for a surplus production model.

- The model convergence statistics were only mildly improved using the proposed base case with tests of normality, constant variance, etc., often not passed (as for the base case although the results were different). This highlights that there are still a) difficulties fitting the variability in the CPUE of the early period (1950-1990) and b) that there is conflicting information between the CPUE indices and survey index on the scale of biomass and the recent years' trends (new Figs 6-9 and new Table 12).
- The model is sensitive to several assumptions, most notably how the survey index is weighted.
- There is uncertainty in the unreported catch included in the model. Although these are appropriately dealt with through uninformative priors and sensitivity tests, it is still significant.

Short term recommendation 4. Given the uncertainties in this assessment, it is recommended that an additional 10% buffer be taken into consideration in the ACL setting process.

Given these uncertainties, it is essential for the next benchmark assessment that there is as much overlap between the standardised CPUE series and the survey indices. This means that the survey should be undertaken annually for the immediate future and as much of the CPUE data should be available prior to the next benchmark assessment as is possible.

Medium term recommendation 24. Ensure the longest possible overlap between the survey and the CPUE indices by undertaking the survey annually for the medium term (at least) and standardising as much of the CPUE data prior to the next benchmark assessment.

7.8.6 Summary

- The final results for the base case are not scientifically sound as it over-rates the survey index by treating it as an absolute abundance index. It is therefore not supported for management use.
- The proposed base case is scientifically sound and supported for use in management.
- For both base cases, the resource remains classified as not overfished with no overfishing.
- Selected biological reference and overfishing limits for the proposed base case are appropriate.
- Both models have a reasonable degree of uncertainty associated with them, which should be considered when determining the buffer in the ACL setting process.

7.9 ToR 9: Are the methods used to project future population status adequate and appropriately applied for meeting management goals as stated in the relevant FEP?

7.9.1 Overall response to question

Yes.

7.9.2 Background

The model is written in R2WinBUGS and uses the well-known Bayesian platform WinBUGS to undertake the assessment. However, during the review workshop it became clear that undertaking forward projections while including process errors affected the results of the optimisation and were therefore not included in the projection (the other error forms did not substantially affect the posteriors). This is because WinBUGS is not a package that is well set up to undertake simulations outside of the optimisation module. This is exactly what one would be doing in a projection where one would want to draw from the posteriors without affecting these. The fact that the model is unable to forward project in WinBUGS with process error makes sense in that there is no information to inform the optimisation. The solution to this issue was to undertake the projection without process error. Structurally this therefore understates the projection error and potential bias. This is not ideal in that the options used for the forward projection are being determined by the kind of model platform that is being used, rather than as an active decision by the analyst. Furthermore, it is ironical that the assessors have developed a process and observation error model, which is then not used in the projections.

An alternative to WinBUGS is STAN with NUTS (www.mc-stan.org) or a package such as TMB (https://cran.r-project.org/web/packages/TMB/index.html - which is not strictly Bayesian, but appropriate for mixed effect models such as those used here). There are also packages available that may not be as complex as the present model that could be investigated, such as the Bayesian Surplus production (BSP) model of McAllister (2014). SPICT (Pedersen and Berg, 2017) is another option.

Medium term recommendation 25. Investigate porting the model to other Bayesian platforms such as STAN or TMB, or use another package such as SPICT or BSP.

7.9.3 Base case versus proposed base case projection results

Forward projections are undertaken assuming the same unreported catch level as previously, which is appropriate. Several constant reported catch projections are undertaken iterating from 0 to 1000 (thousands of pounds). Interestingly, the shape of the mean probability of overfishing curve shape is different from the base case to the proposed base case. Not surprisingly, given the resource is above Bmsy and exploitation rate is below the MSY rate, the catch at 50% probability of overfishing is higher for the proposed base case than the base case. This difference explains the greater difference with the ogives over the projection years between the base case (Fig 16) and the proposed base case (new Fig 16). These results are not unexpected. However, what is unclear is what adding process error to the projections would have done.

7.9.4 Summary

- The approach to undertaking projections are appropriate.
- Porting the model to another platform is important so that process error can be added to the projections.
- The proposed base case results are recommended for management use, whereas that for the base case is not (but the reasons are specified in ToR 8).

7.10 ToR 10: If any results of these models should not be applied for management purposes with or without minor short-term further analyses (in other words, if any responses to any parts of questions 1-9 are "no"), indicate: a) Which results should not be applied and describe why, and b) Which alternative set of existing stock assessment results should be used to inform setting fishery catch limits instead and describe why.

7.10.1 Overall response to question

No to base case, yes to alternative base case.

7.10.2 Background

The survey index for the base case and all the sensitivity tests are implemented in these models without consideration of the error propagated through the establishment of the conversion from a relative index of abundance to an absolute one. This means that there is an implied survey catchability prior in the assessments that is treated as a point prior, i.e. know perfectly. This places very high weight on the single survey relative to the CPUE series, and is not appropriate given what is known about the survey at present. For this reason, the presented base case and associated models do not reflect the full uncertainty.

From the base case and sensitivity tests, the choice of whether to use the survey or not is highly influential on the model outputs (Fig 31.1 and 31.2 in the stock assessment report). The base case and other models are required to "thread the needle", i.e. no matter the journey in the estimated biomass takes prior to 2017, it needs to go within the confidence intervals of the 2017 absolute index of abundance. This means that the model is sensitive to several assumptions of priors, especially the production function parameters.

As also stated in the draft paper by Ault et al. (unpublished), the effective sample area is still reasonably uncertain. This dominant parameter in the calculation of effective area is the effective radius. However, an informative prior for the radius could reasonably be produced and was developed during the review workshop (discussed in detail in ToR 8). This was based on using the lower and upper bounds of the radius as described in the draft paper and the most probable mean value as calculated from the associated modelling (also described in the draft paper).

This new version was run during the workshop and the full set of tables and graphics were provided. These show that the stock is not overfished or overfishing occurring as per the previous assessment, but the absolute index of abundance increases – closer to the CPUE only results. This assessment will evolve as more survey data are added, more research is undertaken on the effective area and there is overlap between the survey and CPUE index.

Short term recommendation 5. Complete the full analyses including projections for the Deep7 proposed base case assessment.

Short term recommendation 6. Undertake the full set of sensitivity tests for the proposed base case assessment.

Short term recommendation 7. Complete the full analyses for the opakapaka updated assessment using the proposed base case methodology.

7.10.3 Summary

- The results for the base case are not supported and should not be applied for management purposes.
- The results for the proposed base case are supported and should be applied for management purposes. In the short term, these results should be checked and updated, and more detailed projections should be undertaken.
- There is uncertainty in the assessment and this should be considered when determining the buffer in the ACL setting process.
- These above summary statements apply to both the Deep7 and opakapaka assessments.

7.11 ToR 11: 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.

7.11.1 Short term recommendations

1. Add B2015/Bmsy and H2015/Hmsy in the sensitivity test summary Table 16.

2. Add the proposed base case to the retrospective analyses in Figure 20.2.

3. Update the projections, check the proposed base case input and results, and use this model for management. (HIGH).

4. Given the uncertainties in this assessment, it is recommended that an additional 10% buffer be taken into consideration in the ACL setting process.

5. Complete the full analyses including projections for the Deep7 proposed base case assessment.

6. Undertake the full set of sensitivity tests for the proposed base case assessment.

7. Complete the full analyses for the opakapaka updated assessment using the proposed base case methodology.

7.11.2 Medium term recommendations

1. Re-investigate a different (shorter nm) cut-off for the distance travelled in the early years. As an initial step, investigate the influence of this as a sensitivity test. This matters since, if this theory is correct, then the CPUE in the early period will be over-inflated, i.e. it is not a consistent bias over time.

2. The author supports the recommendation that vessel size data be obtained from the vessel register database as a factor for the standardisation.

3. Remove tagging and other scientific records from the standardisation dataset.

4. Include the error in the calibration catchability between the hook line survey and the video. These can be bootstrapped as was shown during the workshop. These analyses should also be updated as new surveys are undertaken and more calibration sites become available.

5. Continuing the surveys are a high priority. (HIGH).

6. Investigating whether the survey can be used as an index of total abundance is a high priority. The present investigations of using 360^0 camera with the stereo camera is supported. (HIGH).

7. The survey report's recommendation to further investigate alternative approaches to the MaxN method used to estimate density from video footage (the MaxN method can bias density estimates as it is non-linearly related to true-length-structured abundance) is supported.

8. Investigate the inclusion of oceanographic factors and other factors mentioned in the workshop in the standardisation. Explain why factors mentioned in the workshop are not included in the standardisation.

9. Investigate the inclusion of technology changes such as GPS, plotters, sounders and electric reel types. Several studies have already started on this. If these are not known per fisher, it may be more achievable per vessel. Although offsets in the standardisation for these factors may be fairly arbitrary, they may need to be investigated in a sensitivity test. Arbitrary values should not be added outside the standardisation without adequate justification.

10. Keep the days effort series over all years for use in checking the assessment model output.

11. Run the model with a sensitivity test with the CPUE series as a single series in day effort units (this may be possible in the short term, which can be used as a check on the influence of the survey on the overall CPUE catchability and observation error as this model is simpler).

12. Investigate the reason(s) for the error in Year*Area interaction term. This is a high priority. (HIGH).

13. Consider alternative approaches to targeting opakapaka (and for other new single species assessments) such as TREE (or its relative random forests) and generalised additive models (GAMs). The latter would be useful for investigating space-time changes in targeting.

14. Test the different targeting approaches using a training and test data set. This will allow additional performance statistics to be applied.

15. Reformulate the surplus production form to that of Thorson et al. (2012) to test whether the implied Bmsy/K 's are similar to those obtained in the present model formulation.

16. Regularly update the unreported catch ratio. Given no updates were undertaken for this assessment, it is recommended this work to be undertaken before the next assessment.

17. An education program should be implemented that attempts to increase participation in recording catches.

18. Include the error in the hook to video conversion factor in the calculation of relative individual species and Deep7 biomass.

19. Investigate the use of a delay difference model to the Deep7, but particularly the opakapaka stock assessment.

20. Include pair plots of key parameters in future assessments.

21. Provide chain figure of key of each sample similar to that requested in the review workshop for key derived parameters, especially those that approach limits.

22. Technically, the empirical Bayes approach of setting the P1 prior should be undertaken for each major sensitivity test. Investigate this option especially when using a different and faster modelling platform.

23. Continue the expansion of single spp. assessments for the key Deep7 species.

24. Ensure the longest possible overlap between the survey and the CPUE indices by undertaking the survey annually for the medium term (at least), and standardising as much of the CPUE data prior to the next benchmark assessment.

25. Investigate porting the model to other Bayesian platforms such as STAN or TMB, or use another package such as SPICT or BSP.

7.11.3 Longer term recommendations

1. Undertake separate analyses for the grouper (Hypothodus quernus), which may be more vulnerable within the Deep7 cluster. This may require additional life history work.

7.12 ToR 12: Draft a report (individual reports from each of the panel members and a Summary Report from Chair) addressing the above TOR questions.

A panel report and this individual report has been provided.

8 Conclusions and Recommendations in accordance with the ToRs.

The ToR are all supported, except using the base case for management. During the review workshop, an alternative base case was proposed and run. This proposed base case is supported. In the short-term, updating all components of the assessment report using the proposed base case is a short-term priority. In the medium term, further work on the independent survey, including vessel and technology data in the standardisation, investigating spatial effects in the standardisation or assessment and undertaking further work concentrating on multi-day trip definitions in the early few decades of the CPUE data are all recommended. It is essential that the survey is undertaken annually and that as much of the CPUE data as possible are available for the next benchmark assessment, so that the overlap of these two indices are as long as possible.

The recommendations can be divided into several broad categories:

1) Short-term (1-3 months) recommendations that relate to updating the report for both the Deep7 and opakapaka assessment to reflect a shift to the proposed base case.
2) Medium term recommendations that relate to:
    a) Further investigating the definition of multi-year versus single-year data in the first few decades of the effort data. It is proposed that an additional vessel register database may be extremely useful for these analyses.
    b) Update the unreported catch ratio, as previous studies will be more out of date at the next benchmark assessment. Also include a fisher education program to reduce unreported catch. Ensure that the impact of this education program is also recorded.
    c) Undertake further work on calibrating the survey to:
        i) Better reflect the uncertainty in the scalars and calibration calculations,
        ii) Reduce that uncertainty by including 360^{o} video cameras, and
        iii) Undertake annual surveys which will both increase the number of independent survey biomass indices and the number of hook and video sites for the hook line calibration calculation.
    d) Further enhance the standardisation by:
        i) Including more oceanographic and bottom topographic information in the targeting and standardisation work,
        ii) Including technology factors such as automatic reels and GPS with plotter (either directly through interviews with fishers or indirectly with surrogates such as vessel type and size),
        iii) Maintaining a single effort and CPUE series as a test,
        iv) Undertaking detailed analyses on the space-time changes in the fishery, and
        v) Applying additional techniques to data mining, targeting and standardisation such as TREE and GAM models.
    e) Re-code the assessment model into a faster software framework that can:
        i) Properly undertake projections that include both process and observation errors,
        ii) Calculate the shape (M) parameter of the surplus production form that includes priors such as those considered in Thorson et al. (2012).
    f) Consider alternative approaches for particularly the single species assessments:

i) Consider a grouper assessment, given the species is likely to be more at risk than the other species or at least has a different life history. If a data rich or moderate assessment is not possible, undertake a more data limited approach,
ii) Consider a delay difference model for the opakapaka assessment, and
iii) Continue updating the life history parameters for the Deep7 species.

9 REVIEW PROCESS

The review was undertaken in a very positive and helpful light. The presence of industry at the review was extremely helpful. All requests were undertaken in a positive and collegiate light. The standard of the documentation and presentation was high and greatly contributed to the review process. There was some overlap between the ToR, but this did not affect the review.

10 REFERENCES

Ault, J.S., Smith, S.G., Richards, B.L., Yau, A.J., Langseth, B., Humphries, R., Boggs, C., DiNardo, G.T.. Unpublished. Towards Fishery ‐ Independent Biomass Estimation for Hawaiian Deep7 Bottomfish.

Bedersen, M.W. and Berg, C.W. 2017. A stochastic Surplus Production model in Continuous Time. Fish and Fisheries. 18(2): 226-243.

Deriso, R.B. (1980). Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aquat. Sci. 37: 268-282.

Fournier, D.A., Doonan, I.J. (1987). Can. J. Fish. Aquat. Sci. 44: 422-437.

Kleiber, P. and Maunder, M.N., 2008. Inherent bias in using aggregate CPUE to characterize abundance of fish species assemblages. *Fisheries Research*, *93*(1), pp.140-145.

McAllister, M.K., 2014. A generalized Bayesian surplus production stock assessment software. Collect. Vol. Sci. Pap. ICCAT 70, 1725–1757.

Schnute, J. (1985). A general theory for analysis of catch and effort data. Can. J. Fish. Aquat. Sci. 42: 414-429.

Schnute, J. (1987). A general fishery model for a size-structured fish population. Can. J. Fish. Aquat. Sci. 44: 924-940.

Stephens, A., MacCall, A. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fisheries Research. 70: 299-310.

Venables, W. and Dichmont, C.M. 2004a. A Generalized Linear Model for Catch Allocation: An Example from Australia's Northern Prawn Fishery. Fisheries Research. 70 (2-3): 405-422.

Venables, W. and Dichmont, C.M. 2004b. GLMs, GAMs and GLMMs: an Overview of Theory for Applications in Fisheries Research. Fisheries Research. 70(2-3): 315-333.

11 APPENDIX 1: BIBLIOGRAPHY OF MATERIALS PROVIDED FOR REVIEW

Deep 7 bottomfish benchmark stock assessment for review (not to be distributed beyond reviewers):

- Brian Langseth, John Syslo, Annie Yau, Jon Brodziak, Maia Kapur. 2017. Stock Assessment for the Main Hawaiian Islands Deep7 Bottomfish Complex 2018, With Catch Projections Through 2022. NOAA Technical Memorandum NMFS-PIFSC-XX.

Report from Hawaii Bottomfish Commercial Fishery Data Workshops, 2015-2016. NOAA Admin Report. ~75 pp.

- Annie Yau (editor). 2017. Report from Hawaii Bottomfish Commercial Fishery Data Workshops, 2015-2016. NOAA Technical Memorandum NMFS-PIFSC-XX.

Estimates of Bottomfish Abundance in the MHI in 2016 Based on Fishery-Independent Surveys. NOAA Tech Memo. ~50 p.

- Jerald S. Ault, Steven G. Smith, Benjamin L. Richards, Annie J. Yau, Brian Langseth, Robert Humphries, Christofer Boggs, Gerard T. DiNardo. Unpublished. Towards Fishery - Independent Biomass Estimation for Hawaiian Deep7 Bottomfish.

Previous stock assessment NOT used for management purposes:

- Brodziak, J., A. Yau, J. O'Malley, A. Andrews, R. Humphreys, E. DeMartini, M. Pan, M. Parke, and E. Fletcher. 2014. Stock Assessment Update for the Main Hawaiian Islands Deep7 Bottomfish Complex Through 2013 With Projected Annual Catch Limits Through 2016. U.S. Dep. Commer., NOAA Tech Memo. 59 p.

Independent peer review consensus report for Brodziak et al. 2014 stock assessment:

- Neilson, J. 2014. Stock Assessment Update for the Main Hawaiian Islands Deep7 Bottomfish Complex Through 2013 With Projected Annual Catch Limits Through 2016. Center for Independent Experts, Wellington 6035, New Zealand, 27 p.

Previous stock assessment used for management purposes:

- Brodziak J, Courtney D, Wagatsuma L, O'Malley J, Lee H-H, Walsh W, Andrews A, Humphreys R, DiNardo G. 2011. Stock assessment of the main Hawaiian Islands Deep7 bottomfish complex through 2010. U.S. Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-29, 176 p. + Appendix

Relevant management information (Council FEP and amendment for setting annual catch limits):

- Western Pacific Regional Fishery Management Council. 2009. Fishery Ecosystem Plan of the Hawaii Archipelago. Section 5.3 only, 6 p.
- 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 p.

Reference on unreported to reported catch ratios:

- Courtney, D. and J. Brodziak. 2011. Review of unreported to reported catch ratios for bottomfish resources in the Main Hawaiian Islands. Pacific Islands Fish. Sci. Cent., Natl. Mar. Fish. Ser., NOAA, Honolulu, HI 96822-2396. Pacific Islands Fish. Sci. Cent. Internal Rep. IR-11-017, 45 p.

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

Statement of Work for

Center for Independent Experts' Contribution of Reviewers to the Western Pacific Stock Assessment Review of the 2017 Benchmark Stock Assessment for the

Main Hawaiian Islands Deep7 Bottomfish Complex

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

A stock assessment benchmark of the Main Hawaiian Islands (MHI) Deep7 bottomfish complex was conducted through fishing year 2015 by PIFSC scientists. The Deep7 bottomfish fishery is a targeted deep-water bottomfish handline fishery operated off small boats that holds cultural and economic importance for the region. This benchmark assessment incorporated new data in the form of fishery-independent biomass estimates, and also followed data filtering recommendations from a series of 5 community workshops that involved fishers, managers, and scientists on best practices for filtering bottomfish commercial catch and effort data from Hawaii state commercial catch reports for use in stock assessments. As part of that series of workshops, individual fishers' catch reports have been better linked further back in time and this linking is newly applied in this benchmark stock assessment. This assessment used commercial data for the years 1948-2015 and assessed Deep7 bottomfish, by building upon the previous modeling framework from the past three assessments, but with improved data and data filtering as previously described, along with improvements to CPUE standardization, and other modeling approaches. Unreported catch was calculated and included using catch and effort data following methods similar to those applied in previous assessments. After applying best practices from the workshop recommendations for filtering for CPUE calculation, model selection techniques were applied to select the best structural form to standardize CPUE. CPUE in the model was split into two time series (fishing year 1948-2003, and 2003-2015) in order to accommodate new effort reporting from a change in reporting form by the state in October 2002.

Requirements for CIE Reviewers

Two CIE Reviewers are being sought serve as panel members and conduct an impartial and independent peer review in accordance with the SoW and ToRs herein under the Western Pacific Stock Assessment Review (WPSAR) framework (https://www.pifsc.noaa.gov/peer_reviews/wpsar/index.php). CIE reviewers shall have:

- Working knowledge and recent experience in the application of multispecies and single species stock assessment models including statistical catch-at-age and production models sufficient to complete a thorough review in accordance with the SoW tasks and Terms of Reference (ToRs) as specified herein;
- Expertise with measures of model fit, identification, uncertainty, forecasting, and biological reference points;
- Familiarity with federal fisheries science requirements under the Magnuson-Stevens Fishery Conservation and Management Act;
- Understanding of small-scale multispecies fisheries as well as artisanal fisheries and fishing practices;
- Familiarity with hook-and-line fisheries;
- Expertise in the assessment of slow-growing fisheries species, and;
- Excellent oral and written communication skills to facilitate the discussion and communication of results.

Tasks for Reviewers

Pre-review Background Documents

Approximately two weeks prior to the peer review, the CIE reviewers will be provided (via electronic mail or made available at an FTP site) the necessary background information and reports for this peer review. The CIE reviewers shall read all documents in preparation for the peer review including:

- Deep 7 bottomfish benchmark stock assessment for review (not to be distributed beyond reviewers):
  Benchmark Stock Assessment of Deep 7 Bottomfish Using Data through 2015 (draft NOAA Tech Memo). ~150 p.

- Report from Hawaii Bottomfish Commercial Fishery Data Workshops, 2015-2016. NOAA Admin Report. ~75 pp.
- Estimates of Bottomfish Abundance in the MHI in 2016 Based on Fishery-Independent Surveys. NOAA Tech Memo. ~50 p.
- Previous stock assessment NOT used for management purposes:
  Brodziak, J., A. Yau, J. O'Malley, A. Andrews, R. Humphreys, E. DeMartini, M. Pan, M. Parke, and E. Fletcher. 2014. Stock Assessment Update for the Main Hawaiian Islands Deep7 Bottomfish Complex Through 2013 With Projected Annual Catch Limits Through 2016. . U.S. Dep. Commer., NOAA Tech Memo. 59 p.

- Independent peer review consensus report for Brodziak et al. 2014 stock assessment:
  Neilson, J. 2014. Stock Assessment Update for the Main Hawaiian Islands Deep7 Bottomfish Complex Through 2013 With Projected Annual Catch Limits Through 2016. Center for Independent Experts, Wellington 6035, New Zealand, 27 p.
- Previous stock assessment used for management purposes:
  Brodziak J, Courtney D, Wagatsuma L, O'Malley J, Lee H-H, Walsh W, Andrews A, Humphreys R, DiNardo G. 2011. Stock assessment of the main Hawaiian Islands Deep7 bottomfish complex through 2010. U.S. Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-29, 176 p. + Appendix

- Relevant management information (Council FEP and amendment for setting annual catch limits):
  - Western Pacific Regional Fishery Management Council. 2009. Fishery Ecosystem Plan of the Hawaii Archipelago. Section 5.3 only, 6 p.
  - 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 p.
- Reference on unreported to reported catch ratios:
  Courtney, D. and J. Brodziak. 2011. Review of unreported to reported catch ratios for bottomfish resources in the Main Hawaiian Islands. Pacific Islands Fish. Sci. Cent., Natl. Mar. Fish. Ser., NOAA, Honolulu, HI 96822-2396. Pacific Islands Fish. Sci. Cent. Internal Rep. IR-11-017, 45 p.

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 meeting will consist of presentations by NOAA and other scientists to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers.

Contract Deliverables - Independent CIE Peer Review Reports

The CIE reviewers shall complete an independent peer review report in accordance with the requirements specified in this SoW and OMB guidelines. 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**.

Other Tasks – Contribution to Summary Report

The CIE reviewers will assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewers are not required to reach a consensus, and should provide a brief summary of each 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 30 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 December 31, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
November 2017	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 $10,000.

RESTRICTED OR LIMITED USE OF DATA

The contractors may be required to sign and adhere to a non-disclosure agreement.

NOAA Fisheries Project Contact:

Beth Lumsden

NOAA Fisheries

FRMD/PIFSC/NMFS/NOAA

1845 Wasp Boulevard, Bldg. #176

Honolulu, Hawaii 96818

beth.lumsden@noaa.gov

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

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

2. The main body of the reviewer 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.

   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, which shall include suggestions for improvements of both process and products.

   e. The CIE independent 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 summary report. The CIE independent report shall be an independent peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall 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, presenter information, or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

External Independent Peer Review by the Center for Independent Experts under the Western Pacific Stock Assessment Review framework:

2017 Benchmark Stock Assessment for the

Main Hawaiian Islands Deep7 Bottomfish Complex

For questions 1-5 (and each sub-question therein), reviewers shall provide a "yes" or "no" response with explanations to provide clarification and will not provide an answer of "maybe". Only if necessary, caveats may be provided to these yes or no responses, but when provided they must be as specific as possible to provide direction and clarification.

1. Are data filtering methods as decided upon by a series of regional community workshops correctly applied? Is the scientific uncertainty with respect to the input data quality and filtering methods well documented, including its potential effect on results?

2. Is the CPUE standardization properly applied and appropriate for this species, fishery, and available data?

3. Are the assessment models used reliable, properly applied, adequate, and appropriate for the species, fisheries, and available data?

4. Is each model appropriately specified and configured?

5. Are decision points and input parameters reasonably chosen?

6. Are assumptions reasonably satisfied?

7. Are primary sources of uncertainty documented and presented?

8. Are the final results scientifically sound, including estimated stock status in relation to the selected biological reference points and overfishing limits, and can the results can be used to address management goals stated in the relevant FEP or other documents provided to the review panel?

9. Are the methods used to project future population status adequate and appropriately applied for meeting management goals as stated in the relevant FEP?

10. If any results of these models should not be applied for management purposes with or without minor short-term further analyses (in other words, if any responses to any parts of questions 1-9 are "no"), indicate
    - Which results should not be applied and describe why, and
    - Which alternative set of existing stock assessment results should be used to inform setting fishery catch limits instead and describe why.

11. 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.

12. Draft a report (individual reports from each of the panel members and a Summary Report from Chair) addressing the above TOR questions.

Annex 3: Tentative Agenda

2017 Benchmark Stock Assessment for the

Main Hawaiian Islands Deep7 Bottomfish Complex

Honolulu, Hawaii
November 13-17, 2017

Day 1, Monday November 13

1. Welcome and Introductions
2. Background information - Objectives and Terms of Reference
3. Fishery
    - Operation
    - Management
4. History of stock assessments and reviews
5. Data
    - State of Hawaii Fisher and Dealer Reporting Systems
    - Life history information
    - Fishery-independent survey

Day 2, Tuesday November 14

6. Presentation and review of stock assessment

Day 3, Wednesday November 15

7. Continue review of stock assessment

Day 4, Thursday November 16

8. Continue review of stock assessment
9. Public comment period
10. Panel discussions (Closed)

Day 5, Friday November 17

11. Continue panel discussions (Closed; morning)
12. Present results (afternoon)
13. Adjourn

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

13.1 LIST OF PARTICIPANTS

13.1.1.1 WPSAR POCs

1. Benjamin Richards - NOAA PIFSC
2. Marlowe Sabater - Western pacific Regional Fishery Management Council

13.1.1.2 PIFSC assessment authors

3. Annie Yau - NOAA PIFSC
4. Brian Langseth - NOAA PIFSC
5. John Syslo - NOAA PIFSC
6. Jon Brodziak - NOAA PIFSC
7. Maia Kapur - NOAA PIFSC

13.1.1.3 Other PIFSC scientists:

8. Mark Fitchett - NOAA PIFSC
9. Felipe Carvalho - NOAA PIFSC
10. Beth Lumsden - NOAA PIFSC

13.1.1.4 Invited speakers/experts:

11. Joe O'Malley - NOAA PIFSC
12. Kimberlee Harding -Hawaii DAR
13. Kurt Kawamoto - NOAA PIFSC
14. Sarah Ellgen - NOAA PIRO

13.1.1.5 Members of public:

15. Leonard Yamada - Fisher
16. Roy Morioka - Fisher
17. Layne Nakagawa - Fisher
18. Nathan Abe - Fisher
19. Ed Watamura - Fisher

13.2 CIE REVIEWERS

20. Dr. Steve Martell (chair), Seattle, Washington
21. Dr. Cathy Dichmont, Banksia Beach, Australia
22. Dr. Henrik Sparholt, Holte, Denmark