Independent review of the stock assessment of King threadfin (*Polydactylus macrochir*) in Queensland, Australia.

Prepared by

Neil Klaer

Prepared for

Queensland Department of Agriculture and Fisheries

Desk-top review
25 February – 1 April 2021

Contents

Executive Summary

This is a desktop review of the 2020 king threadfin (*Polydactylus macrochir*) fishery stock assessment by the Department of Agriculture and Fisheries (DAF) Queensland carried out during the period February 25 to 1 April 2021.

Major uncertainties for the stock assessment (also acknowledged by the authors) relate to total catch (including discards) and catchability change in fishery CPUE. The stock assessment is limited by the available data – particularly length and age compositions only mostly available in recent years. Differences in stock characteristics by area have been accounted for by conducting separate stock assessments for each area. I examined stock assessment settings and diagnostics in detail and agree that the assessment has been competently constructed and is adequate given the available data. The range of results provided for management purposes encompasses most major uncertainties in the stock assessment, and the current base case for each area is the best currently available for the provision of management advice. I have provided advice on further consideration of model uncertainty and improvements to model documentation for scientific purposes.

The assessment document included research recommendations separately as they apply to data, monitoring, management and the stock assessment and I agree with those. Catchability change and effects on the reliability of fishery-dependent CPUE abundance measures is a significant and on-going problem for the assessment of this fishery. Means for further exploration via modelling of alternative mechanisms for that change, as well as the development of fishery-independent abundance measures should be considered.

I support the conclusions that application of the Sustainable Fisheries Strategy to stock assessment results determines that: (a) future harvest levels for the East Coast appear to be sustainable at near equilibrium harvest levels in line with the target reference point and (b) the GoC requires a three-year period of no harvest to allow rebuilding to above 20% of unfished spawning biomass levels.

1 Review Activities

This is a desktop review of the 2020 king threadfin (*Polydactylus macrochir*) fishery stock assessment by the Department of Agriculture and Fisheries (DAF) Queensland (Leigh et al. 2020). The review process was sent out by DAF for competitive tender, and I was contracted to do the review commencing on 25 February 2021. Formal terms of reference for the review were included in the contract. I received the king threadfin stock assessment report and associated model input files on 2 March. During the review, I requested some additional documents listed in the Annex. Having access to the model files greatly assisted the review as I could examine more detailed diagnostics not provided by the assessment report and run my own diagnostic code on the models. It was discovered that I had not received the correct model input files used for the report for areas other than the Gulf, and this was corrected on 10 March. I provided a detailed comparison of model input settings for all area models and my comments on detailed model diagnostics for the Gulf model to the assessment lead author, George Leigh. He provided responses to those comments back to me during the review, which were much appreciated. This exchange ensured that any concerns that I may have had about model implementation were equally understood by myself and the lead author. I completed the review on 24 March and sent my draft report to George Leigh and Sue Helmke. I sought comments on the factual accuracy of statements in my review report and whether the terms of reference had been met. Comments received on 31 March were provided by George Leigh, Sue Helmke, and Alex Campbell. Clarifications were made in the report to statements regarding the acceptability of harvest recommendations based on those comments. The final report was submitted on 1 April. I thank all who I have had contact with for this review which progressed efficiently and professionally.

2 Review of stock assessment of king threadfin

2.1 Objectives of the stock assessment

The stock assessment had the following objectives:

1. Collate the relevant fisheries data.

2. Develop harvest estimates, standardised catch rates and biological data to input to the population model.

3. Estimate stock status with respect to reference points described in the Queensland *Sustainable Fisheries Strategy 2017–2027*.

4. Propose Recommend Biological Catches (RBCs).

5. Provide recommendations for management and monitoring.

2.2 Terms of reference for the review

The Department of Agriculture and Fisheries (DAF) is seeking an independent review of the "Stock assessment of the king threadfin (*Polydactylus macrochir*) in Queensland, Australia". The review is not limited to, but should address the following points:

1. Provide comment model inputs and outputs and adequacy of these data to achieve the objectives of the assessment, including:
    a. Assessing the six management regions in Queensland.
    b. Providing biomass ratio estimates in relation to the fishery reference points.
    c. Assumptions used in the analysis of catch rates.
    d. Appropriateness of harvest sizes.
    e. Confidence in model outputs.
    f. The adequacy of the population dynamic model used in the assessment.
    g. Appropriate recommended biological catch.
2. Provide comment on the accuracy of key statements in the report summary and conclusion. How well are they supported by available data, analysis and literature?
3. Provide comment on recommendations for management and monitoring and inclusion of additional data in future assessments.
4. Any other outputs or graphical figures that the report could have provided.

A formal written report of the findings of the review is to be provided to the lead stock assessment author and a nominated person from Fisheries Queensland. The written review and independent reviewer identification may be released and made publicly available.

2.3 Findings according to terms of reference

2.3.1 Provide comment model inputs and outputs and adequacy of these data in order to achieve the objectives of the assessment

2.3.1 a. Assessing the six management regions in Queensland

This assessment assumes that stocks within the six assessment regions (that mostly correspond to management regions) are self-contained with little or no mixing of adult or juvenile fish. Data from tagging suggests that some mixing has occurred (e.g., from AR3 to AR4, from AR4 to AR5). Tagging also shows within-area long-distance movement (e.g., in the Gulf of Carpentaria (GoC) area from Weipa to the Flinders River). Also, there is evidence that fish move potentially hundreds of kilometres from nursery to other areas as they grow. However, there does seem to be supporting evidence that growth is different by assessment region (particularly in asymptotic length) for what could be assumed for assessment purposes to be the resident populations of these large areas. Given the available information, it is reasonable to suggest fairly independent fish populations by assessment region that may have different biological characteristics in growth, natural mortality and stock-recruitment dynamics. It is therefore acceptable to assess those populations independently. A feature of independent area assessments is the ability to borrow population characteristics for the data-poor areas from the more data-rich ones where required, which has been done in this assessment to some extent. I agree with the decision not to assess AR1 due to the lack of supporting data and small catches in that area.

The information given in the assessment report regarding the biological spatial structure of the stock is very brief, and I sought additional information from Moore et al. (2011) and Moore et al. (2017). More detail on this should be given in the report because this spatial structure is crucial for understanding of the spatial structure assumed by the assessment.

2.3.1 b. Providing biomass ratio estimates in relation to the fishery reference points

As each assessment area is treated independently, separate time series of total spawning biomass (spawning output) are estimated for each area as input to QDAF (2020) harvest control rules to reach future catch recommendations for each area. The ratio of $B_{current}/B_0$ can be determined with more accuracy than absolute spawning biomass, and current management is based on a target for that ratio of 0.6 and a limit of 0.2 (QDAF 2020). For any individual stock assessment, management currently relies on the central values of these estimates from a selected base case, and do not specifically take account of stock assessment uncertainty, except indirectly through selection of the target ratio and an uncertainty buffer. Uncertainty in stock assessment results are provided as asymptotic distributional ranges for the base case, and also via results from “high” and “low” sensitivity model scenarios.

2.3.1 c. Assumptions used in the analysis of catch rates

Commercial gillnet logbook data were collated to produce one record per fisher-day and data were filtered to those potentially directed to king threadfin depending on whether king threadfin or known associated species were caught. Remaining records may contain zero catch of king threadfin. This is a common procedure used for multi-species fisheries although the details may differ depending on the application. I agree that such a procedure is required for king threadfin, and how it has been done seems appropriate.

A quasi-negative-binomial generalised linear model was used to standardise annual abundance indices by fleet and Catch Rate Region (sub-area) within assessment areas. Explanatory variables used were also described. I believe that standardisation is something that can be separately documented from the stock assessment as it is a separate modelling process and most often informed or carried out by professional statisticians. To fully evaluate the standardisations, more documentation is required including model diagnostics such as plots to justify distribution assumptions, presentation of the procedure used for choice of optimal models, and presentation and evaluation of results for fitted factors or variables and their significance. This means that I can only comment that the methods used seem appropriate.

A representative standardised catch rate per fleet was chosen from those available as the Catch Rate Region with the greatest total catch of king threadfin. Due to this selection, several CPUE series were not used for the assessments. The AR3 and AR4 series used for the assessment showed a large recent increase which has been used to support justification for an increase in catchability assumed for all assessment areas. For both assessment areas, there are CPUE series that were not selected for use which did not show such a recent increase. It would be an improvement to make use of the series not included in the assessment and ways to implement this need to be considered for the future. This may be via combined standardised indices per assessment area or including all available indices in the assessment models with appropriate weightings. However, I agree that selection of series based on data content and therefore true abundance representativeness is supportable.

2.3.1 d. Appropriateness of harvest sizes

Many assumptions have been made to allow a complete catch history for the fishery by assessment area to be constructed. For east coast commercial harvest these included: assumed zero catch to 1936 and linear increase to the first year of QFB data (mostly 1946), interpolate linearly from mean catch from the last 2 years of QFB (1980 and 1981) to mean catch of first 2 years of CFISH (1988 and 1989). For GoC commercial: three scenarios of catch prior to first logbook records in 1981 constructed using reconstructed historical net fisheries effort estimates and a few historical point-in-time reference years. For recreational harvest: various point estimates from RFish, NRIFS and SWRFS surveys from 1997 to 2019 were interpolated and made comparable using scaling factors, converted to weight from numbers using Boat Ramp Survey length data, and best available information was used to estimate recreational harvest prior to 1997. It was noted in the report that indigenous harvest estimates were available but not used for the assessment and I agree that this should be included in future.

Despite good efforts in historical catch reconstruction, the number of assumptions shows that historical harvest estimates for king threadfin are uncertain and this uncertainty should be evident in assessment results. The calculation of three alternative harvest series for the GoC partly achieves this. Discard underestimation due to the tendency of the species to spoil quickly through flesh necrosis once netted was also highlighted by Welch et al. (2002) as a problem for the accuracy of reported king threadfin catches.

I agree that the current constructed history makes reasonable assumptions and is acceptable.

2.3.1 e. Confidence in model outputs

Confidence in model outputs derives from the correct use of an appropriate assessment model, while making full use of input data and estimating properties specific to the stock to allow total population estimation for management.

A previous attempt at stock assessment for Gulf of Carpentaria king threadfin used simple biomass dynamic models (Welch et al. 2002). This was found to be unsatisfactory because “the predicted index of cpue was unreliable with the resulting estimates of biological parameters being unrealistic” and “sustainable yield (MSY) and effort corresponding to MSY (Emsy) were also highly unrealistic and this is likely to be due to failure of the data rather than the models”. The CPUE was seen to be unreliable and subject to changing catchability “due to the efficiency of targeting individual schools (hyperstability)” and effort creep particularly since 1998 due to the introduction of power net reels.

Since that time and particularly in very recent years a considerable amount of length and age composition data has become available that could be of potential use for the assessment of the species. The current assessment has been developed using Stock Synthesis (SS) (Methot and Wetzel 2013) that has many advantages including use of input data of most types even if incomplete, verification via simulation of the basic dynamics and many assessment options, fitting of growth within the assessment, appropriate procedures for estimation of parameter uncertainty, wide use throughout the world with many previous applications, and automated methods for production and display of model diagnostics. There are also disadvantages of SS including a steep learning curve and potential risk of inappropriately using it and its many options, but I believe that the stock assessment team have undertaken appropriate formal SS training. I agree with the choice and appropriateness of the stock assessment framework and also acknowledge that there are perhaps equally capable alternatives available such as CASAL.

On model settings, I provided the lead author with a table of settings differences among area models and some associated comments, and notes on settings that I found unusual compared with many other stock assessments. I had specific questions about differences in age binning across areas, lowest minimum length bin, time blocking, hermaphroditism, number of estimated growth parameters, fixed sigma R in eastern models, years to estimate recruitment deviations, implementation of catchability increase, age selectivity for some fisheries, fixed selectivity retention and cryptic spawning output. These were resolved to my satisfaction although I discuss time-varying selectivity in detail below. I have an interest in cryptic spawning biomass which may be large for some stock assessments due to domed selectivity, but not in this case (with a small amount in the GoC models).

Time-varying catchability is well accepted as a common phenomenon that affects both fishery-dependent and fishery-independent abundance series. It is more often seen as an attribute of fishery-dependent abundance series and is often the main reason to attempt to move to fishery-independent measures where possible. For king threadfin there are no fishery-independent abundance (survey) measures available.

Common procedures used to deal with time-varying catchability in stock assessments are:

1. Account for known or proposed mechanisms in CPUE standardisation
2. Ignore or downweight an index if catchability change is expected
3. Model catchability as a function of density or an environmental variable
4. Model catchability as a random walk process
5. Model catchability as a function of time (either as a constant change or stepped process)
6. Use state-space models that allow catchability to change over time

Most often seen in stock assessment is (1) where fisher knowledge and other sources are used to attempt to quantify catchability change over time due to change in biological processes, regulations, environmental influences, gear efficiency, targeting methods or vessel fishing power. This has been successfully done for some assessments by DAF. However, a similar process for king threadfin is made difficult due to differences among assessment regions and perhaps lack of readily available detailed information on fishery practices. Option (2) is not feasible, as there is only a single fishery-dependent CPUE series available for king threadfin in each area by fleet. Options (3) and (4) may be viable ones that have not been explored for the current assessment. Option (4) requires the support of more informative data – particularly length and age data which is unavailable for king threadfin until very recently. This is a considerable challenge for this assessment as any model-based fitting of catchability scenarios is only informed by composition data in very recent years. Option (6) is one that could be explored for this stock, but is an alternative assessment method to SS. The current SS king threadfin assessment uses option (5).

Assuming a constant change in catchability is a strong assumption for a stock assessment because it essentially says that the overall trend shown by the available abundance indices is unreliable. Modelled absolute biomass in particular, but also biomass trend information normally comes most directly from abundance indices (plus catches) and is not reliably provided by composition data. Maunder and Piner (2015) say “there has been a trend to deemphasize the abundance content of age- and size-composition data (Francis, 2011). This is because relatively minor model misspecification (e.g., a too inflexible selectivity curve) can have a large impact on the information about absolute abundance contained in the composition data (Lee et al. 2014)”. For king threadfin, composition data are only available mostly within the most recent 10 years by area except for AR4, which could not be expected to inform the overall pattern of catchability change for the greater number of years covered by abundance indices.

This is the first SS assessment from the hundreds that I have seen that assumes constant catchability change over time (constant change in ln(q) called q_{inc}). I have seen many accepted stock assessments that used stepped changes as an implementation of option 5 above, and one that used density-dependent option 3 (US Pacific Fisheries Management Council: petrale sole, 2015). This is not a reason to reject the procedure but indicates that a degree of caution is required. I understand why it has been done but a strong supporting case should have been presented in the assessment document – i.e. a demonstration that CPUE trends for AR3 (including those not used) are implausible and that models using other assumptions about catchability change (including none) produce implausible results. This is done to some extent with sensitivity analysis results given for various q_{inc} values by area, but a more compelling justification is required that includes interaction with other influential parameters such as M.

There is some potential to incorporate prior information about acceptable values for q_{inc} from meta-analyses such as Palomares and Pauly (2019) who found an average “creep factor” for vessel fishing power of 2-4% per year across many pelagic and demersal fisheries.

Models that vary in their method of accounting for time-varying catchability can differ substantially in their results, and there has been little formal evaluation of the potential bias of the various approaches, or agreement on standard procedures (e.g., see Wilberg et al. 2010). This indicates that the application of a single method with a tightly constrained assumption about the shape over time of catchability change (i.e., constant increase in ln(q)) considerably reduces the uncertainty about stock status that is caused by the uncertainty about catchability change. Alternative methods for handling the catchability change for king threadfin need to be explored and presented (beyond alternative values for q_{inc}), at least as sensitivity analyses.

I agree that the approach used for time-varying catchability in the assessment for king threadfin is supportable, among other possible procedures. However, the construction of sensitivity tests using alternative time-varying catchability scenarios based on alternative hypotheses is required to describe true uncertainty in stock status for this fishery.

Detailed examination of the assessment model framework and setup shows that it appears to competently apply assumptions common to many other assessments. In my experience, the one unique aspect for this assessment is the constant increase in log catchability which I believe is justified but should be further explored as future research. The assessment model implementation is therefore acceptable.

2.3.1 f. The adequacy of the population dynamic model used in the assessment

I examined detailed diagnostics using R4SS output, particularly for the GoC model which had the most supporting data. For GoC model diagnostics I had questions for the lead assessment author regarding standard warnings produced by the model, minimum length binning, age selectivity in addition to length selectivity, implementation of time-varying retention in forecast years, selectivity vs age and growth for ARGulf_north, implausibly low variance for the initial biomass (B_0) and the estimated series generally, lack of tuning of recruitment deviation bias adjustment, high apparent F of 0.4 to 0.6 since 1980, possible density-dependence of time-varying q, relatively poor fits to index series, some lack of common trend among index series, numbers of fish in the plus group for the equilibrium age distribution, lack of data on ageing imprecision, tuning of effective Ns for length compositions, relatively poor fit of expected to observed retained length frequencies especially for the north, apparent lack of fit of expected to observed mean age by year for the north, and an apparent precision of the $\ln(R_0)$ estimate. These were mostly resolved to my satisfaction although some questions remain about whether some of these diagnostics represent errors in the implementation of R4SS. I was using the development version that contains more plots than the release version, but some questions remain about their reliability that are beyond the capability of this review to resolve.

There are large Pearson residuals in the age at length data fits for AR2-AR5 which should ideally all be below 2. In general, fits to length and age composition data by all models are not particularly good, but this is an acknowledged limitation of the available data.

I believe that there is an issue in the GoC assessment with what seems to be an estimate of R_0 that is mostly pre-determined by the model implementation and/or available data. This leads to a suspiciously precise estimate of B_0 and therefore the whole modelled biomass series. I have seen such precision of absolute biomass levels in other accepted assessments where there has been a recommendation to investigate this further as future research. I would start by attempting a likelihood profile for R_0 (over a narrow range about the estimated value), and a close examination of the interaction with q_{inc} in this case.

The assessment document states that various methods of tuning effective sample sizes were attempted and that none were entirely satisfactory. I recommend that this also be pursued as future research.

This assessment is perhaps more uncertain than many due to uncertainty in historical catches and discards, very recent availability of length and age composition data for most areas, and the apparent unreliability of the fishery-dependent abundance indices. It is therefore important that this uncertainty be sufficiently characterised and conveyed to management as it may influence their decisions based on the reliability of the results.

Uncertainties in model implementation in this report are examined through the presentation of two sensitivity model scenarios that provide high and low productivity alternatives to the base case in each area for management purposes. Set of results are also tabulated for sensitivity analyses of q_{inc} and steepness for all eastern models and high and low harvest, natural mortality, and steepness for the GoC. I believe this is probably sufficient within an assessment report for management for evaluation of the uncertainty for each of the models. It is good that results that seemed to be implausible were also included. Dimensions of uncertainty presented do cover major ones, but a more comprehensive list can potentially be constructed.

Decisions have been made about what constitutes implausible results which was described as those that indicate “very high biomass, on which fishing had hardly any effect”. I believe that results for sensitivities such as q_{inc} 0.02 in AR2 that produce an overall lower -lnL than the base case require further examination of fits to particular data series to refine the rule used for model rejection.

In Table 1 I present a list of uncertainties for the current king threadfin assessment and their associated questions. Most commonly for assessment documentation, such questions are converted to alternative scenarios that are examined more thoroughly via sensitivity analyses. The relative importance of uncertainties is often judged according to their influence on the stock assessment results. Some I have listed are not likely to have great influence (e.g. spawning biology) while some may be substantial (e.g. total historical catch, time-varying selectivity).

Table 1 Dimensions of uncertainty and level potentially addressed via alternative model scenarios

Uncertainty	Degree addressed	Comments/questions
Spatial structuring	Yes	Separate assessments mostly by management area/geographic region.
Total catch	Partial	High, middle and low reconstructed harvest series examined for GoC. Potential to explore models (especially for GoC) with initial F estimated at the start of more reliable catch data. Indigenous catch not included.
Fishery CPUE	Partial	Not all available series used.
Time-varying catchability	Partial	Fitted (GoC) or fixed values for other areas. Narrow range of possible mechanisms considered. Values of q_{inc} used as a major dimension of uncertainty for all areas except for GoC.
Spatial variation in growth	Yes	Growth is assumed to differ among areas. Variation among areas in parameters estimated where supportable by available data.
Spatial selectivity	Yes	All fleets and areas have separate selectivity.
Spawning biology	No	Is ignoring hermaphroditism an acceptable approximation? Stocks that transition male to female are probably less problematic than the reverse when treated within a one sex model.
Discards	No	Fixed retention curves used to account for changes in minimum size regulations through time. Historical discard rate and mortality is uncertain.
Major productivity parameters (*M*/*h*)	Yes	*M* and *h* examined for the GoC. Other areas examined *h*.

I agree with the authors that the two sensitivity cases for each area span a range of alternative model structures that are useful in conveying model uncertainty to fishery managers. My suggestions here apply to a wider range of sensitivity and other tests that that could be examined and presented to allow improved scientific judgement of the behaviour and uncertainty of the assessment models.

Likelihood profiles provide useful insight to model behaviour, and those for M for the GoC model and q_{inc} for AR2-AR5 were presented. It was a little concerning that the q_{inc} profile for AR4 showed opposing trends to those in other eastern areas. Additional profiles for R_0 should be considered as those examine the influence of data sources on absolute biomass estimates by the models.

Retrospective pattern analysis involves the sequential dropping of data from the assessment model, backwards, one year at a time, to examine whether the assessment appears to systematically over- or under-estimate biomass or recruitment levels. Diagnosis of retrospective bias in stock assessments has received considerable attention in the literature, research is on-going, and means for diagnosis and correction for them are not agreed. A rule of thumb that can be used to diagnose retrospective patterns is by Hurtado-Ferro et al. (2014) that says that “values of Mohn’s ρ higher than 0.20 or lower than -0.15 for longer-lived species, or higher than 0.30 or lower than -0.22 for shorter-lived species should be cause for concern and taken as indicators of retrospective patterns.” Retrospective bias provides evidence for model misspecification, but the lack of a retrospective bias does not prove that the model is correctly specified. Examination of retrospective patterns should be considered for future assessments as a further diagnostic of model reliability.

The population dynamics models used for this assessment are adequate.

2.3.1 g. Appropriate recommended biological catch

Policy for the estimation of catch levels to achieve a target spawning biomass is outlined by QDAF (2020). The assessment document states that “model results were projected forward 20 years, to 2039. A harvest control rule was employed, consistent with the Sustainable Fisheries Strategy (Department of Agriculture and Fisheries 2017). This control rule assumed that the fishery was closed when the spawning biomass fell below 20% of the unfished level (B_{20}). Above B_{20}, the instantaneous fishing mortality rate F was assumed to equal F_{60}; i.e. the rate corresponding to a steady-state spawning biomass level of 60% of unfished (B_{60}). We did not use any precautionary buffer on fishing mortality.” A buffer is a discount factor applied to the control rule to account for risk under uncertainty. For this assessment, no buffer value was used, but this decision was not explained in the document. I agree that the form of the harvest control rule and therefore projections follow from the policy and are appropriate.

2.3.2 Provide comment on the accuracy of key statements in the report summary and conclusion. How well are they supported by available data, analysis and literature?

For management purposes, the report adequately describes important aspects of the species biology, fishery extent, current assessment input data, model construction, and model results. Uncertainty in the spawning biomass trajectory within the current base case for each area is presented, as well as the spawning biomass series for various sensitivity analyses. Future harvest levels according to the harvest control rule for base models as well as high and low sensitivity models were also provided. These are adequate to describe central values for future harvest recommendations and also the uncertainty of those to some extent.

I support the conclusions that application of the Sustainable Fisheries Strategy to stock assessment results determines that: (a) future harvest levels for the East Coast appear to be sustainable at near equilibrium harvest levels in line with the target reference point and (b) the GoC requires a three-year period of no harvest to allow rebuilding to above 20% of unfished spawning biomass levels.

2.3.3 Provide comment on recommendations for management and monitoring and inclusion of additional data in future assessments

The assessment report included a fairly comprehensive section on recommendations separated as they apply to data, monitoring, management and the stock assessment. I agree with those recommendations. I can see some advantage in working towards an integrated multi-region model that could estimate common parameters as informed by all areas but acknowledge that no current off-the-shelf modelling framework is currently available for that. I also agree that manual tuning of model results currently required (balancing effective sample sizes for input data, recruitment deviation bias adjustment) does not support automation and hope that this might be rectified by future model package developers.

I have included recommendations for additional exploration of model uncertainty in this report. I also recommend that additional procedures for accounting for catchability increase be further investigated and at least accounted for in sensitivity tests for future assessments.

Several potential anomalies were highlighted via examination of R4SS diagnostic plots for this assessment. Those should be investigated with the help of the R4SS development team where required.

It is a standard research recommendation to develop fishery-independent abundance indices for fisheries that do not have them. Whether this is possible is normally determined by the value and importance of the fishery. How this might be cost-effectively achieved for king threadfin should be considered – e.g., close-kin genetic analysis. This is a significant problem for this fishery as it is reasonable to expect that catchability change will be an on-going problem for future stock assessments.

2.3.4 Any other outputs or graphical figures that the report could have provided to aid fishery management processes

Outputs and graphical figures provided in the report were sufficient for fishery management purposes. However, they were not sufficient to allow scientific review of the stock assessment. As I was provided with model input files I was able to run my own diagnostics to support this review. I believe that it has become necessary to provide such files to scientific reviewers to allow a thorough examination of the assessment implementation.

3 References

Francis RICC. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68: 124–1138.

Hurtado-Ferro F, Szuwalski CS, Valero JL, Anderson SC, Cunningham CJ, Johnson KF, Licandeo R, McGilliard CR, Monnahan CC, Muradian ML, Ono K, Vert-Pre KA, Whitten AR and Punt AE. 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES J. Mar. Sci. doi: 10.1093/icesjms/fsu198.

Lee H-H, Piner KR, Methot RD and Maunder MN. 2014.Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: an example using blue marlin in the Pacific Ocean. Fisheries Research 158:148-156. DOI: 10.1016/j.fishres.2013.12.017.

Leigh GM, Tanimoto M and Whybird OJ. 2020. Stock assessment of king threadfin (*Polydactylus macrochir*) in Queensland, Australia. Queensland Department of Agriculture and Fisheries Report.

Maunder MN and Piner KR. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science 72(1):7–18. doi:10.1093/icesjms/fsu015.

Methot, RD., Wetzell, CR. 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142: 86–99.

Moore BR, Welch DJ and Simpfendorfer CA. 2011. Spatial patterns in the demography of a large estuarine teleost: king threadfin, *Polydactylus macrochir*. Marine and Freshwater Research 62:937–951.

Moore BR, Stapley JM, Williams AJ and Welch DJ. 2017. Overexploitation causes profound demographic changes to the protandrous hermaphrodite king threadfin (*Polydactylus macrochir*) in Queensland's Gulf of Carpentaria, Australia. Fisheries Research 187: 199–208

Palomares MLD and Pauly D. 2019. On the creeping increase of vessels' fishing power. Ecology and Society 24(3):31. Doi:10.5751/ES-11136-240331.

QDAF. 2020. Reef line fishery harvests strategy 2020-2025. Queensland Department of Agriculture and Fisheries.

Welch D, Gribble N and Garrett R. 2002. Assessment of the Threadfin Salmon Fishery in Queensland – 2002. Queensland Government Department of Primary Industries Report.

Wilberg MJ, Thorson JT, Linton BC and Berkson J. 2010. Incorporating time-varying catchability into population dynamic stock assessment models. Rev. Fish. Sci. 18, 7–24. doi:10.1080/10641260903294647.

Annex 1: Bibliography of materials provided for review

Leigh GM, Tanimoto M and Whybird OJ. 2020. Stock assessment of king threadfin (*Polydactylus macrochir*) in Queensland, Australia. Queensland Department of Agriculture and Fisheries Report.

Leigh GM and O'Neill MF. 2017. Stock assessment of the Australian east coast tailor (*Pomatomus saltatrix*) fishery. Queensland Department of Agriculture and Fisheries Report.

Moore BR, Welch DJ and Simpfendorfer CA. 2011. Spatial patterns in the demography of a large estuarine teleost: king threadfin, *Polydactylus macrochir*. Marine and Freshwater Research 62:937–951.

Moore BR, Stapley JM, Williams AJ and Welch DJ. 2017. Overexploitation causes profound demographic changes to the protandrous hermaphrodite king threadfin (*Polydactylus macrochir*) in Queensland's Gulf of Carpentaria, Australia. Fisheries Research 187: 199–208

Welch D, Gribble N and Garrett R. 2002. Assessment of the Threadfin Salmon Fishery in Queensland – 2002. Queensland Government Department of Primary Industries Report.

Supplementary data files

SS starter, .ctl, .dat, and forecast files for the base case models for each area and the high and low sensitivity tests for the Gulf.