Review of the 2024 Stock Assessment for King Threadfin (*Polydactylus macrochir*) in the Gulf of Carpentaria, Queensland

20 March 2024

Simon Hoyle and Alistair Dunn
Hoyle Consulting & Ocean Environmental

Review of the 2024 stock assessment for king threadfin (*Polydactylus macrochir*) in the Gulf of Carpentaria, Queensland

20 March 2024

Simon Hoyle[1] & Alistair Dunn[2]

1. Hoyle Consulting, Nelson, New Zealand <simon.hoyle@gmail.com>
2. Ocean Environmental, Wellington, New Zealand <alistair.dunn@oceanenvironmental.co.nz>

Abstract

In February 2024, the Queensland Department of Agriculture and Fisheries (QDAF) contracted an independent review of the 2024 stock assessment for King Threadfin in the Gulf of Carpentaria. The species occupies a complex spatial domain with metapopulation structure, data are limited for much of the time series, and there are uncertainties about the species' behaviour. A considerable amount of work has been undertaken by the stock assessment scientists, who fully engaged with the reviewers in providing detailed information in response to questions and engaged in scientific discussions to allow the review to be undertaken. An ensemble approach is used to investigate a range of model scenarios for steepness, natural mortality, and effort creep.

The estimated biomass trend and stock status appear generally consistent with the catch history and the current age structure, but stock status is likely considerably more uncertain than the current estimate.

Some aspects of the model need further development, and we recommend changes where appropriate. A more conventional data weighting scheme should be applied – this is easily addressed. Further development of the CPUE indices is needed. Regions should be assessed in independent models with composition data only from within each region. Systematic lack of fit to some of the data components should be resolved. More consideration should be given to the metapopulation nature of the stock and its implications for data collection, preparation and assessment modeling.

Citation: Hoyle, S.D.; & Dunn, A. (2024). Review of the 2024 stock assessment for king threadfin (*Polydactylus macrochir*) in the Gulf of Carpentaria, Queensland. Hoyle Consulting Ltd., Nelson, New Zealand, 47 p.

1 Introduction

In February 2024, Queensland Department of Agriculture and Fisheries (QDAF) contracted Hoyle Consulting and Ocean Environmental to undertake an independent review of the draft 2024 stock assessment for King Threadfin in the Gulf of Carpentaria, Queensland. The project objectives were to evaluate the appropriateness of the data, analyses, and inferences in the stock assessment. The overall objective and specific objectives of the review are listed in Annex A of this report. The review took place over February-March 2024. As part of the review, the reviewers engaged with stock assessment scientists at the Queensland Department of Agriculture and Fisheries (QDAF), members of the stock assessment panel, and king threadfin scientists. They also reviewed the published literature and stock assessment documents relevant for the review.

In undertaking the review, we note that a considerable amount of work has been undertaken by the stock assessment scientists, and that they fully engaged with the reviewers in providing detailed information in response to questions and engaged in scientific discussions to allow the review to be undertaken. This review recognises the significant work for the stock assessment of king threadfin that has been undertaken by QDAF and all the scientists and fishery experts involved.

The main documents considered include a draft version of the 2024 assessment (Leigh et al., 2024b), the 2021 stock assessment (Leigh et al., 2021), and the independent review of the 2021 assessment (Klaer, 2021). Other literature and documents reviewed are listed in the references below.

1.1 Review activities

This is a desktop review of the draft 2024 king threadfin (*Polydactylus macrochir*) Gulf of Carpentaria stock assessment report (Leigh et al., 2024b), produced by the Queensland Department of Agriculture and Fisheries (QDAF) (and referred to in this review as 'the report'). The review was funded by QDAF as a competitive tender. We were contracted on 16 February 2024 to begin the review 28 February 2024. Formal terms of reference for the review were included in the contract. We received the king threadfin stock assessment report and associated model input files on 28 February, and details of the assessment were presented to us by the authors during an initial online meeting. Having access to the model files greatly assisted the review as we were able to examine more detailed diagnostics that were not available in the assessment report. During the review period we had a series of further meetings and online discussions with the authors, during which we requested and were provided with further information. These exchanges helped to ensure that any concerns were understood by both the reviewers and the assessment authors. Some of our discussions are reflected in updates to the report. In addition, we communicated with other experts, including Brad Moore regarding threadfin stock structure and the 2010 study, Claudine Ward and Jason Stapley about the commercial fishery, and Ben Bright about the charter fishery.

The draft review was completed on 18 March and submitted for review. We sought comments on the accuracy of statements in the report, and whether the terms of reference had been met. Comments were received on 19 March from George Leigh and Alex Campbell. The final report was submitted on 20 March.

We thank everyone involved in this review for their professional, efficient, and positive contributions.

2 Review of the stock assessment

2.1 Spatial structure of the king threadfin stock

Studies of King Threadfin stock structure around Australia have identified complex spatial structure of the stocks. Patterns of spatial connectivity were evident from otolith elemental data, mtDNA haplotypes (Horne et al., 2010), otolith stable isotope ratios (Newman et al., 2010), life history parameters (Moore et al., 2012b), and parasite assemblages (Moore et al., 2012a). Within the Gulf of Carpentaria, Moore and Simpfendorfer (2014) concluded that king threadfin operated as a series of metapopulations among estuaries, “connected by a degree of non-trivial movement that is neither so low as to negate significant demographic connectivity, nor so high as to eliminate independence of local population dynamics (Kritzer and Sale, 2004)”.

Several views about the threadfin stock structure were expressed in the threadfin stock assessment report. On the one hand it states “Tagging data show that some king threadfin move hundreds of kilometres. In the Gulf, one fish moved 600 km from Weipa in the northern Gulf of Carpentaria to the Flinders River beyond Karumba in the southern Gulf (Infofish 2014). A common hypothesis that king threadfin forms multitudinous stocks of spatial extents less than 100 km (see, e.g., Welch et al., 2010, pp. 148-150) is not supported by available data from the Queensland Gulf of Carpentaria.” On the other hand, the report also notes “Although tagging data show that king threadfin can move long distances (Infofish, 2014) they appear to move much less than many other fish species, and king threadfin populations from different regions may mix only on time scales of many years.”

Tagging data can be informative about movement, and the most relevant (and reliable) information comes from the patterns observed from the majority of returned tags. The behaviour of most of the population is what matters for the stock structure of an assessment, because the behaviour of the majority determines how fishing will affect biomass depletion and population age structure. Rare or small numbers of tagged fish moving larger distances may not be representative of the population. Moreover, all datasets include error, and tag data are no exception. Errors in tag recovery location are not uncommon (e.g., Leroy et al., 2013), and tend to show larger displacements than correctly reported tag locations. Although large displacements are probably reported correctly, they may not be. For these reasons, the best approach is to focus on the majority of the movement distribution.

Moore (2011) analysed the threadfin tagging data available up to 2011, and reported 45 recaptures in the Gulf, of which 38 had enough location information to estimate movement. There was no significant relationship between time at liberty and displacement. One fish was caught after 2599 days with 1km displacement. Of the 38, 70% were recaptured in the same estuary in which they were tagged and 30% (12 fish) were recaptured outside of their tagging estuary. Of these, 20% (8 fish) were caught over 10 km south of the estuary mouth, and up to 568 km away from their release location. All fish with large displacements were tagged in the Weipa region and their recapture was reported from further south. Tagged fish with large displacements were larger in length than average, with none tagged at less than 70 cm, and all were recaptured within less than one year. There are reports of tags recovered since 2011, with some recent recoveries mentioned in the report. However, the data have not been presented so it is difficult for us to make inferences from them. As the sample size of 38 recovered tags is quite small, additional analyses that include tagged fish recaptured since 2011 may be informative for such an important issue as population structure.

The report notes that king threadfin spawn in high salinity inshore coastal waters (Halliday et al., 2008). There is therefore very likely to be some degree of size structuring in riverine versus coastal catches, potentially with larger fish in coastal areas. Since spawning is seasonal with a late winter-spring peak (Garrett, 1997), there may also be seasonal size structure. We discussed the potential for coastal vs riverine size structure with a researcher (Brad Moore), a charter operator (Ben Bright), and two commercial fishers (Claudine Ward and Jason Stapley). They had observed some size structuring on average between riverine and coastal areas, as well as seasonal variation and individual locations with different sized fish. In general, the size differences were not obvious, and some large and small fish could be found in all areas.

2.2 Assumptions of stock structure

The stock assessment assumes that the Queensland Gulf of Carpentaria King Threadfin stock is a single well-mixed stock with no local-scale residency. This assumption is influential for the methods used to derive the observational data and their interpretation in the assessment model. It assumes that the stock behaves homogenously – fishing in one location affects every part of the stock, reducing density and changing the age structure in all locations.

A stock with some independent local population dynamics, however, will not behave like this. Fishing in one location will reduce density and change age structure in that area but will affect other locations only to the extent that there is mixing.

The degree of mixing will affect the length composition, age composition at length, and CPUE data. If locations are sampled in proportion to the catch, locations subject to higher fishing pressure will be overrepresented in the data, areas with low fishing pressure will be under-represented, and areas without fishing pressure (closed or inaccessible) will be unrepresented. If there is a degree of local residency, a preference for more sampling in areas with more fishing will introduce bias into the assessment. The amount of bias will depend on the degree of mixing and the spread of sampling. A possible improvement is to re-stratify to better reflect the population and possibly the fisheries (Maunder et al., 2020).

Similarly, given these dynamics, catch rates will reduce more in areas with higher fishing pressure. Therefore, CPUE models will need to allow for the spatial distribution of fishing effort across the stock.

Given the evidence discussed above that rapid and complete mixing is unlikely, the combination of the current approach that assumes full mixing, together with sampling weighted towards more heavily fished areas is, *a priori*, likely to bias assessment results somewhat towards more pessimistic outcomes. The degree of bias will depend on both the rate of mixing between areas, and the variation in fishing pressure between areas. This is not to say that the model is necessarily too pessimistic - other factors may be present that cause a different bias.

Spatial structure can be characterised at three levels – firstly related to differences among regions, secondly related to the river catchment within region, and thirdly related to foreshore and coast versus river netting. The first level is considered to the extent that the region is included in CPUE models, with interactions estimated between region and year, and region and month. River catchments are not considered because there are insufficient data at that resolution. Differences between foreshore and river netting are also not considered, for the same reason.

At the regional level, the model implements an areas-as-fleets approach, with a separate fleet defined for each of the five regions (Mornington, Karumba, Gilbert, Pormpuraaw, and Aurukun). Three of the fleets (Mornington, Karumba and Gilbert) have their own CPUE index, but selectivity is shared among all fleets, and composition data from all regions are pooled. Data from three periods are allocated to the three fleets with indices, but this is purely for presentation purposes and has no effect on model outcomes due to the sharing of the selectivity parameters.

Moreover, while interannual variation in all CPUE indices affects biomass, estimation of q_{inc} parameters for the Mornington and Gilbert regions means that average CPUE trends in these regions do not affect the overall biomass trend – this is largely driven by CPUE for the Karumba region where q_{inc} is fixed=0.

We note that analyses of data to identify how spatial dynamics may affect stock assessment outcomes will require the following steps:

1. Identification of a time series of size and/or age data with information about location (latitude, longitude, coastal versus river). Information on other factors that affect catch rates and sizes are also needed, such as fisher identity and fishing gear (e.g., mesh).
2. Exploration and analyses of the data to understand the patterns (seasonal, spatial, environmental, gear, reporting biases).
3. Identification of model structures that will work with the biology and data, such as:
    a. Regions / pooled / areas-as-fleets.
    b. Fitting to sizes and/or ages, what kind of age-length key, how to weight size and data.

3 Data inputs

3.1 Data characterisation

Characterisation is one of the key stages of a stock assessment. The process of data characterisation is essential for analysts to understand the data they are modelling, and reviewers and stakeholders need the documentation to be able to judge the assumptions made by the analysts. Thorough characterisation has been recommended as one of the key good practices in CPUE standardization for stock assessment (Hoyle et al., 2024).

We recommend that a data characterisation be a routine part of the assessment and assessment updates. This may help the assessment team develop hypotheses and model assumptions and help aid a more complete understanding of how the assessment model could be structured.

During this review we requested additional plots and figures to better understand the fishery and to help assess the reliability of the CPUE indices. The analysts provided additional information on how catches, catch rates, sizes and ages vary spatially, seasonally, and through time. Additional plots and figures are given in the KTF companion document (Leigh et al., 2024a) and in Annex B of this report.

One issue revealed by the characterisation was that daily catches were usually reported in multiples of 30kg (Figure 12), likely based on the number of 10kg boxes of fillets.

3.2 CPUE

Indices of abundance are usually the most important and influential component of a stock assessment. For king threadfin, the indices of abundance available were based on fishery dependent catch-per-unit-effort (CPUE), using commercial gillnet catch rates from CFISH logbook data. The series covers the period 1988-2022. Logbook data were collated to produce one record per fisher-day, focused on the location where most fish were caught.

The methods were described in the assessment document, but the description was incomplete. Supplementary information was provided to the reviewers in a set of RMarkdown files. The availability of the RMarkdown files was helpful, allowing us to better understand how many of the methods were applied. Much of this has now been included in a more detailed companion report (Leigh et al., 2024a).

The RMarkdown files included some additional diagnostics, but these were limited in scope. Additional diagnostics are needed to evaluate if the error distribution assumptions were appropriate, to show relationships of residuals with covariates, and to show the covariate effects on expected values.

3.2.1 Targeting

The gillnet fishery has historically mainly targeted barramundi as this species had higher catch rates and market price. The dataset also includes some gillnetting targeted at marine species such as grey mackerel and other scombrids.

Prices for threadfin have increased in recent years and are now similar to barramundi, so fishers now target both species depending on availability factors such as location and season. Therefore, the degree of targeting of threadfin versus barramundi has changed through time which will have affected catch rates.

The assessment analysts used a bespoke method, based on associations with other species, to identify targeted effort and thereby filter the data. The method involved calculating correlations between catch rates of king threadfin and other species and identifying species with positive or negative associations. Only a small amount of data was filtered out using this process (2.27%). While this was a small amount of the total data, additional information on the process of determining the final data should be described. In this case, the information was not available in the assessment document and was determined by examining the RMarkdown files provided by the analysts. We recommend that relevant details of potentially influential factors, such as the proportion of effort removed annually, are documented in either the assessment report or in supplementary companion documents.

Fishers expressed concern that the associated species method may not adequately filter the data and result in too much non-targeted data being retained in the model – we agree that this appears likely. In addition, the method is likely to be unable to adjust for what may be a more important issue – an increase in king threadfin targeting compared to barramundi targeting. Failing to account for targeting strategy is particularly problematic when strategy allocations change through time as they can be confounded with the annual abundance indices.

Identifying targeting methods in multi-species fishery is challenging, but we believe that more effective methods are available and should be considered. Multi-species species composition can be informative about both the targeting strategy and the habitat being fished – both of these issues may significantly affect indices. As well as including data from various targeting strategies, the CFISH gillnet fishery catch and effort data includes operations in both riverine and coastal environments. Cluster analysis of species composition (He et al., 1997) is an established approach that uses combined information from multiple species to separate effort into groups, rather than taking one species at a time. For a discussion of this approach and other methods for identifying targeting strategies in catch and effort data, see Hoyle et al. (2024) and references therein.

3.2.2 Covariates

Spatial covariates

As mentioned above, the model was applied jointly to the entire dataset, with region-year and region-month interactions. Although this approach was able to fit the data and estimate all parameters, we consider that a better approach would be to model each region as a separate analysis.

There are many advantages and few disadvantages to modelling regions separately. The model already includes interactions between region and year and between region and month, so the main parameters shared are the covariates. The character of each region is distinct, which may lead to different covariate effects by region, and different error distributions. An individual fisher's knowledge of one region may not transfer to another region, so sharing fisher effects across regions may not be appropriate. Differences among regions may also affect the mesh and net length effects.

Separation by regions would also allow a simple longitude smoother to represent variation along the coast for Mornington and Karumba regions, with a latitude smoother taking the same role in the regions extending north up the Cape York coastline. A numeric smoother has interest because the distance from markets or population centres may be associated with fishing pressure. Other approaches, such as categorical variables for rivers, are also reasonable. Separate regional analyses also make residual patterns and relationships clearer so that it would be easier to diagnose analysis problems.

As a part of this review, and while exploring analysis of the Karumba region, the analysts identified a relationship between longitude and CPUE which was not in the original model. It indicated lower catch rates, which may signify greater depletion, of areas near Karumba where fishing pressure was thought to be higher. This result is consistent with the hypothesis that king threadfin operate as a series of metapopulations (substocks) among estuaries (Moore and Simpfendorfer, 2014).

Given the combined weight of evidence for spatial structure in the population, analyses of individual regions should allow for spatial variation within the region. Allowing for spatial effects in CPUE is essential to avoid bias in the indices (Walters, 2003). The current approach, without a spatial component, will provide more information about trends in substocks with more fishing pressure, which (all else being equal) may result in an overly pessimistic outcome. Adding a spatial component such as a longitude effect would improve estimates. However, its results are still likely to be biased, and a spatiotemporal model with a year-longitude interaction is likely to be a better model. If this approach was used, the longitudes would be weighted by fishable habitat area when predicting annual CPUE to generate the final indices. This approach can be developed using spatial models with software such as VAST, but analysis would be easier using GAMs with spatial smoothers or Gaussian random fields, which make more efficient use of data than using categorical variables. The spatial resolution of the CFISH dataset is relatively high and with the amount of data available, this approach should be feasible.

River flow

The report notes that river flow rates are considered likely to affect catch rates. The mechanism for this effect is unknown, but catchability might increase because flooding provides more food (a high-rainfall event may increase productivity for two or three years) or a higher metabolic rate (metabolism may increase for a year or two after a new high-flow event), which may lead to more movement and therefore vulnerability to netting. Alternatively, or in addition, it is plausible that high river flows will reduce salinity, which may push king threadfin out of riverine areas that are now too fresh for them to tolerate. While moving downriver and returning upriver after floods subside, they may be more vulnerable to netting. Floods may also concentrate king threadfin in downstream areas, increasing their catchability.

Flow rates were developed from Queensland Government flow monitoring data with a series of necessarily subjective decisions. The value for each flow station in a wet season was defined as the maximum value observed in a 14-day moving average of the water discharge over each wet season. Stations were then weighted by catchment area and summed to represent each region.

Flow was not included as a covariate in the CPUE analysis because effects were not thought to occur at the set level but at the season-catchment level, potentially with some delay. Delays were implemented by defining flow as the sum of the current year's wet season, the previous year's wet season, and half of the flow from the wet season two years previously.

Within each region, flow for each year was then categorised as either low or medium-high for that region, regressed against the annual CPUE estimates, and catch rates adjusted to match the medium-high flow setting. This process is the same as taking the average CPUEs of low flow years vs medium-high flow years and then adding the difference to the low flow years' CPUE, so that each group has the same average CPUE.

Overall, this approach is subjective and involves many approximations. We recommend future research to develop a better understanding of the mechanisms through which flow may affect threadfin populations. Freshwater flow has been shown to be correlated with the year class strength of king threadfin recruitment in the Fitzroy river (Halliday et al., 2008), and if similar effects occur in the GoC they may be a confounding factor.

During discussions, a member of the project team indicated that modelled estimates of river flow were available from CSIRO and were likely be an improvement on the estimates used. We agree and recommend that analysts should consider taking advantage of any alternative estimates that are available. Hydrological modelling is a well-developed discipline, and the approach used here was (necessarily) ad hoc.

To increase confidence and understanding, inferences about the influential flow-CPUE relationships should be better supported with analyses. For example, the statement in the report that “inspection of river flows and catch indicated that river flow affected catch rates for several years after a flow event” should be supported by evidence with some analyses provided to demonstrate the relationship.

3.2.3 Model fitting and index development

The CPUE analysis was fitted using a generalized linear model (GLM) in R (R Core Team, 2023). GLMs are flexible, perform well, and are appropriate here. We encourage the use of GAM models implemented in packages such as *mgcv* (Wood, 2011). They can replicate almost any GLM analysis but have the advantage that they can also include a lot of additional flexibility to include testing of alternative assumptions.

The analysis modelled errors with a quasi-negative binomial distribution (QNB). The analysts suggest that the QNB (and its relatives) would be superior to hurdle models because (since they are based on the negative binomial distribution) they make no qualitative distinction between zero catches and small positive catches. However, this is not necessarily an advantage – see below.

The quasi-negative binomial distribution does have good properties that are applicable to king threadfin in that it can provide a good fit for highly skewed data (Li et al., 2011). However, a zero-inflated version may be needed if there are excessive numbers of zeroes (Li et al., 2011). According to Ben Bolker (https://stackoverflow.com/questions/68915173/how-do-i-fit-a-quasi-poisson-model-with-lme4-or-glmmtmb), "quasi-likelihood models really represent a post-fitting adjustment to the standard errors of the parameters and the associated statistics; they don't (or shouldn't ...) change anything about the way the model is fitted."

The QNB distribution may be an appropriate distribution for this dataset, but additional analyses are required to demonstrate that this is the case. Given the lack of diagnostic plots we were only able to make limited judgments about goodness of fit. We note the following potential issues with the use of the quasi-negative binomial distribution in this case.

- With negative binomial distributions, the proportion of zeros is affected by the distribution of non-zeros, which may not be appropriate when the proportion of zeroes is affected by several processes (Lambert, 1992). It is common for fishery catch rates to be affected by multiple processes. In this situation zero-inflated or hurdle models are often used, as is common with fishery data. Note that this addresses the analysts’ argument about a qualitative distinction between zeros and non-zeros – although the simplicity of a unified model is appealing, flexible approaches are needed.
- It is difficult to generate useful diagnostics for the QNB because R packages such as DHARMa (Hartig, 2020) and the simulation tools within R base are not implemented for quasi distributions. Having few diagnostics makes it difficult for analysts to identify and resolve problems in the analysis.

Indices were initially provided to us with the year effects sourced directly from the GLM model object. Although the change did not affect the indices in this case, we encourage the default use of the more flexible prediction-based methods instead - see Hoyle et al. (2024) for a discussion of the advantages, such as the ability to account for density covariates in the indices. The analysts also provided prediction-based estimates.

After discussing these issues, the modelling team provided us with additional CPUE analysis model runs with a number of changes: assuming a delta lognormal distribution, separate analysis of the Karumba region, investigation of spatial covariates within regions, and diagnostic plots (Figure 18) including some developed using the DHARMa package (Figure 19). During this process the analysts identified a relationship between longitude and CPUE which was not in the original model, consistent with significantly greater depletion of areas with higher fishing pressure (Figure 20). This result is consistent with the hypothesis that king threadfin operate as a series of metapopulations among estuaries (Moore and Simpfendorfer, 2014).

3.2.4 Fishing power

Effort creep, or an increase in fishing power over time, is common in fisheries. An increase in effective effort based on target switching (which is to some extent distinct from effort creep) seems likely in this fishery. Given the increased value of threadfin versus barramundi in recent years, net setting techniques in some conditions would have been adapted to catch more threadfin. It is therefore reasonable to include options that set q_{inc} to a non-zero value. Runs that set q_{inc} to zero should also be considered.

There would have been a large increase in catchability with the introduction of monofilament (much less visible to fish) in the 1970s (Darcey, 2015), but this was before the start of the CPUE index. Subsequent changes may have been less influential. We discussed this with industry members Claudine Ward and Jason Stapley:

- GPS (from 1990s) – would have had have some benefit, but not so helpful inshore and in rivers where landmarks have always been available.
- Sounders, plotters (ongoing) – these help to find fish, to the extent that fish distributions are unpredictable.
- Freezers, faster boats (ongoing) – allow fishing further from markets but should not increase threadfin catch per length of net.
- Net reels – make fishing more efficient but perhaps do not increase threadfin catch per length of net.

It is unclear why q_{inc} should be higher for fleets other than Karumba, while keeping Karumba at q_{inc}=0. A justification is needed, based on an underlying process. Was there more of a switch to threadfin targeting in Mornington and Gilbert than in Karumba?

Estimating q_{inc} inside the model does not seem appropriate – there is no plausible source of information about q_{inc} in the data, particularly given the quality of the early composition sampling data. An estimate for a single region model would be driven by conflict between CPUE trends and the composition data, conditioned on model structure and biological parameters.

Estimating q_{inc} in regions outside Karumba while holding Karumba q_{inc} constant would simply adjust their mean CPUE trends to be the same as Karumba. The result would be to effectively ignore any trend information in the other CPUE indices. From a results perspective we are comfortable with that outcome, because Karumba appears to be the region with the highest proportion of the stock and the most information. It is best to focus on what's happening there. However, a more appropriate approach would be to run the Karumba model by itself; in other words, to run each region independently and to set q_{inc} to alternative plausible values. The fact that the q_{inc} estimate for each other region was positive suggests that trends in these regions' CPUE indices were to some extent more optimistic than the CPUE trend in Karumba.

We note that increasing q_{inc} for Karumba and/or adjusting for more recent targeting of threadfin would tend to make Karumba CPUE decline more. This would reduce early depletion, which would tend to better fit both the early (1985-1993) age-at-length data and the early CPUE.

3.3 Catch estimates

Early catch estimates are very uncertain. The commercial catch reconstruction method is plausible, and we have little basis for questioning it. The only (minor) issue is the assumed decrease in CPUE of 3% per year from 1970 to 1977, based on lower abundances. This assumes no change in catchability, which was not the case. Monofilament nets were introduced during the 1970s and widespread by 1976 (see page 2 of report). They are known to have substantially increased catchability because they are invisible to fish, whereas fish could see the nylon that was used previously (Darcey, 2015).

A possible cause for concern is the inability of the model to predict the observed ages at length in the 1986-1993 composition data. These fish are observed to be older than the model predicts, which is consistent with total mortality being lower than the model estimates. This inconsistency could be caused by overestimation of catch during the previous 10 years. However, there are other possible explanations for the large ages-at-length, and other changes could be made to the data and/or model structure that would improve the fit.

Recreational catch estimates are a small proportion of the total catch, so we have not spent time considering them.

The total catch in the Karumba area has been declining since 1977 after peaking at over 400 tonnes (Figure 1). Over 20 years from 1974 until 1995 catch was above 200 tonnes, but since then it has been well below that level apart from brief peaks in 2001-2002 and 2010. Since 2013, catch in the Karumba region has been less than 100 tonnes per year. This is a substantial reduction in catch which, unless abundance is very low, would be expected to benefit the stock.

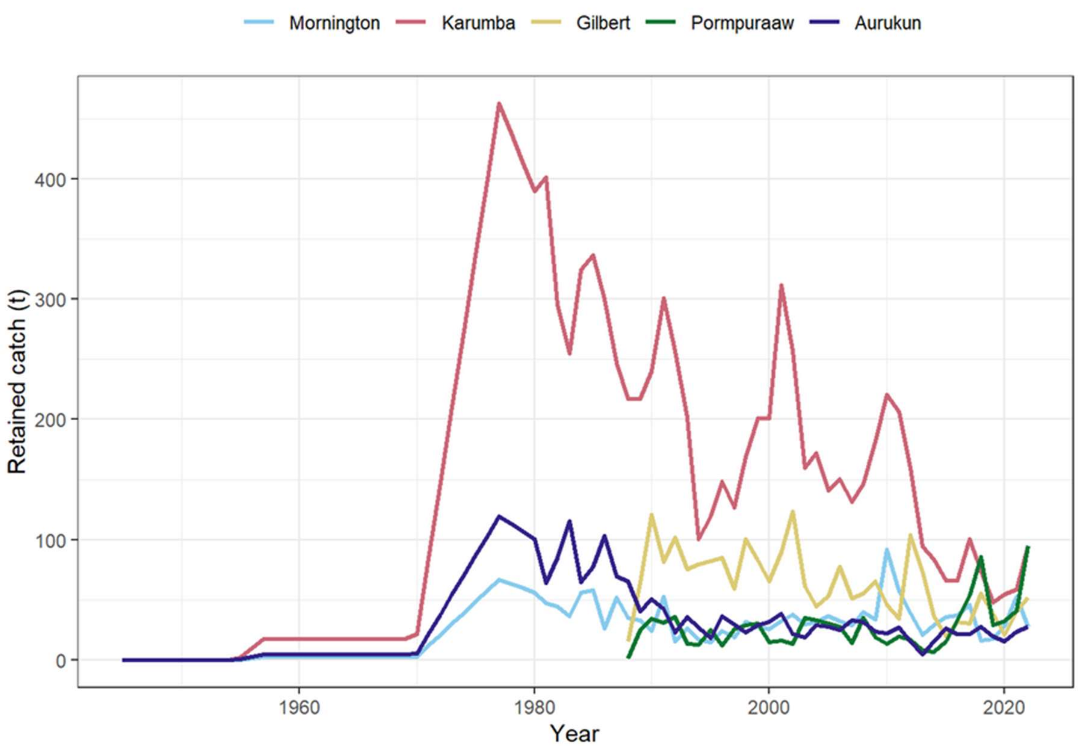

Figure 1: Line plot of total estimated catch by year and region.

Catch in other areas has been less variable. Mornington and Aurukun catches have been relatively stable since 1995. Gilbert catch has been declining since 1990, while Pormpuraaw catch has been increasing since 2015 after being stable since 1990.

3.4 Age and length data

Composition data are included in the model as length compositions, combined with observations of conditional age at length data from otolith readings. Length samples from the three southern Catch Rate Regions (Mornington, Karumba and Gilbert) were merged on grounds of small sample sizes in most years in Mornington and Gilbert, and similarity of the length distributions from the three regions. Similarly, the age-at-length data were merged across all 5 catch rate regions.

We advise against merging data from the different areas. Mixing between areas is believed to be low to very low, and fishing pressure is likely to differ between regions. Composition data should only be combined among areas if there is full mixing. Merging the data loses information.

There is also no need to remove strata with low sample sizes. Data weighting should be sufficient to give appropriate influence to those strata with small sample sizes. The model assumes selectivity to be invariant through time, so lack of data in some years is not a problem for the analysis.

Both the length data and the age-at-length data appear to be quite noisy. Patchiness in the size data is probably due to patchy size distributions in the catch and the population. There is not a lot of information in these data about changes in the population – in other words, the effective sample size of these data appears to be low, when considered as sources of information for the assessment. For example, size data for Karumba have a very different pattern in 2020 compared to 2021.

Interpretation of the causes of the observed patterns requires information about the monitoring process. There is likely to be variation - how many samples, sets, fishermen, months, etc? Compare sampling with the timing of the fishery. It looks like all this info is available? Should be analysed to identify the sources of the variation, e.g., how much is random and how much between locations, months, net types, etc. sets, etc.

An important source of spatial variation is likely to be spatially varying fishing pressure. The monitoring programme is designed to sample composition data that are representative of the size and age structure of the catch. This is a common approach in fishery management. However, it has a major drawback when the stock has a metapopulation structure with limited mixing: fishing pressure varies among substocks, and the catch is weighted towards substocks experiencing higher fishing pressure. As a result, samples from the catch will be smaller and younger than if sampling was weighted to be representative of the population distribution. The composition data are used in the model as if they were representative of the population (after allowing for fishery selectivity). This mismatch results in a bias that tends to overestimate fishing pressure. The size of the bias will depend on the rate of mixing and the amount of variation in fishing pressure among substocks. Results of CPUE analyses and discussions with fishing industry suggest significant variation in depletion and allocation of fishing effort to different parts of the Karumba region.

A potential approach to resolving this problem would be to reweight the composition data to be representative of the population (Maunder et al., 2020), based on spatial CPUE. This reweighting approach requires knowledge of the location (at least to the catchment level) where each composition sample was taken. It is currently unclear how much of this information is available in the monitoring data, much of which is obtained at the trip level. It is not possible to infer fishing location from the identity of the fisher, because for confidentiality reasons that information is not recorded. Some data are available however (Figure 17), and the related age-at-length data may be sufficient to test the hypothesis that depletion varies along the coast.

Another potential source of variation in fish size data is coastal versus riverine fishing. However there is little available information on coastal versus riverine sampling (Figure 14, Figure 15, Figure 16), and it is likely confounded with sampling in different catchments.

Sampling occurs across seasons (Figure 15), which may have three effects on the model. Fish may move to spawn seasonally, so selectivity may change seasonally. The model includes only one fishery, and spatial substock effects are likely far more important, so this cannot be dealt with. Secondly, fish grow within each season, but the data file represents the size data as though all sampled in July. Thirdly, fish growth within seasons also affects the distribution of age at length. These two issues may have only a small impact but can be addressed in part by including sampling month in the SS input data, rather than the current approach of including them as sampled in July. This also implies taking care about the month when ring increments are laid down – see the Stock Synthesis manual for details.

The length composition data were sampled across all seasons but were included as a single fishery without timing information. See Punt et al. (2020) "An Achilles heel of contemporary assessment methods that fit to size-composition data is their temporal resolution. While animals retain the same integer age throughout a year or season, they grow at various rates during the year, so the time at which animals are sampled needs to be aligned with the appropriate time resolution along the continuous growth axis."

4 Biological parameters

4.1 Length-weight relationship

This comes from McPherson (1997). L-W relationships can change seasonally and spatially, and vary with covariates, but updating it is not a high priority unless a change would affect the assessment.

4.2 Recruitment

Values for σ_R of 0.3 and 0.6 are used in the ensemble. The 0.6 value is consistent with values used in many other fish stock assessments. The value of 0.3, like the 0.35 value estimated in the previous assessment (Leigh et al., 2021), is lower than used in most stock assessment for fish species and gives a lot of influence to the assumed value of steepness. The penalty associated with a recruitment deviate is proportional to the square of σ_R, so a value of 0.3 penalises deviations 4 times as much as 0.6. Since deviations are on a log scale, this can make a huge difference to recruitment estimation.

It is not good practice to estimate the variability of recruitment estimates within a maximum likelihood assessment. Estimates tend to be biased lower than the true variability among recruitments (Methot and Taylor, 2011). This is particularly the case in a model fitted with maximum likelihood – Bayesian methods are a minimum requirement. Age data can provide good information about relative recruitment strength, but the signal is always affected by model misspecification and ageing uncertainty. This dataset includes only a small amount of ageing data and there are only a few cases where it appears to track cohorts.

The assumed σ_R is often negatively correlated with B0 and R0, so choosing a value that is too low will tend to increase depletion estimates. It is usually better to assume a high σ_R when fitting a model because this gives the model more freedom to fit the data. We recommend that the commonly assumed value of 0.6 should be used when fitting the model to estimate recruitments, instead of 0.3.

4.3 Natural mortality

The lowest instantaneous rate of natural mortality (M) in the models (0.325) was the value estimated for the Queensland East coast in the 2021 assessment (Leigh et al., 2021).

Information about natural mortality can be inferred from other stocks but should be considered carefully. Natural mortality has both intrinsic and extrinsic components. Intrinsic factors are those associated with the organism itself and are likely to be similar across stocks and through time. Extrinsic factors vary among stocks, locations (e.g., Strøm et al., 2019), and through time, and include abiotic challenges, predation pressure, competition, and exposure to pathogens. The GoC environment is very different from the Queensland east coast, and extrinsic natural mortality factors may also be different. Threadfin demography can vary substantially between areas (Moore et al., 2011). There are also genetic differences between the GoC and east coast stocks (Horne et al., 2010), but these are likely small enough not to affect intrinsic natural mortality.

Estimates of natural mortality based on *Amax* (i.e., following Hamel and Cope, 2022) are also affected by the history of fishing mortality experienced by the stock, because the relationship M=5.4/Amax is based on the lightly fished and unfished stocks in the Then et al. (2015) database. Estimates of *Amax* from a heavily fished stock (such as Gulf of Carpentaria king threadfin) are therefore likely to be biased low, and *M* biased high. In such cases, *M* based on the A_{max} of a conspecific in another location that is less heavily exploited and better sampled can provide guidance for determining *M*. Given some degree of demographic independence among king threadfin populations in estuaries across northern Australia, there may be opportunities to find lightly fished threadfin stocks and sample their age structure.

We also caution that M is hard to estimate from fishery data (Maunder et al., 2023) and misspecified models (which all models are to some extent) tend to produce biased estimates of M. We were unable to review the 2021 assessment (Leigh et al., 2021) in detail, but suggest caution about accepting as reliable its estimate of M for the Queensland east coast. For the current assessment, given the spatial complexity of this stock, the simplified nature of the assessment across a large and complex spatial domain, the uninformative CPUE series with little contrast, uncertain catch estimates, uncertainty about the pattern of selectivity, and evidence of poor fit to age structure data, we recommend against estimating M inside the model. We endorse the approach of fixing M at a range of values across the prior. Representing the uncertainty in M and how this influences estimates of management quantities is an important component of conducting stock assessments (Maunder et al., 2023).

When data from lightly fished populations are available, we recommend considering a prior based on M=5.4/Amax (Hamel and Cope, 2022), using values of Amax obtained from populations in the Gulf of Carpentaria and elsewhere. This approach has been recommended in recent reviews of methods for estimating natural mortality for fish stocks (Maunder et al., 2023).

We also recommend the biologically well-justified approach of setting M inversely proportional to body length (Lorenzen, 2022). With this approach the prior mean is usually applied to the mature age classes.

4.4 Steepness

Steepness (h) represents a measure of density-dependence in the stock recruitment relationship (SRR) (Zhou et al., 2020). It is always a major source of uncertainty in stock assessments. We recommend using the standard approach of representing that uncertainty in management advice via a wide range of alternative values. The standard approach for steepness is to represent uncertainty by considering a range of alternative values, and basing management advice on them all (i.e., an ensemble approach). An alternative recommended approach (Brooks, 2024) is to use the null SRR as a default – i.e. steepness of 1, combined with a large estimate of σ_R (Miller and Brooks, 2021).

When the model is allowed to estimate steepness, the estimate tends to be low: 0.56 in an early version of the base case. However, this is equivalent to assuming that the assessment contains reliable information about steepness, which is not our view. Lee et al. (2012) conducted a simulation study to explore the ability of stock assessment models to estimate steepness and found that a high proportion of estimates occurred at the bounds – i.e., the best fit was obtained at impossible values either below 0.2 or above 1. They concluded that reliable estimation of steepness is attainable for relatively unproductive stocks with good contrast of stock spawning biomass, where the model is correctly specified.

The threadfin stock assessment does not meet these criteria. There is not good contrast of spawning biomass, because the age and CPUE data are sampled only when the stock is depleted, and the CPUE series shows little contrast. The age and size data appear to have much lower effective sample size than they are given in the model. The model fits the early age-at-length data very poorly and consistently predicts younger ages, which indicates that something is wrong, and so estimates of productivity parameters such as steepness should not be trusted. The early periods were opportunistically sampled and the later period, though well sampled to represent the catch, is not representative of the population as the model assumes. It is not feasible to correctly specify the model because spatial dynamics are complex and not well understood: the stock is modelled as fully mixed using an areas-as-fleets approach, whereas there appear to be significant spatial dynamics, both between and within regions. There are strong environmental drivers with a spatial-temporal component, which are not well understood.

The lower estimates of steepness appear to be driven by the slow recovery in the estimates of Karumba CPUE and age-at-length, despite the large reduction in catch. For similar reasons, the model preferred lower estimates of natural mortality – lower values of both represent lower resilience and productivity.

We endorse the approach used by the analysts of applying a wide range of values for steepness. They chose a range from 0.55 to 0.75. Values of 0.45 and 0.85 were rejected due to strong trends in the recruitment deviates. However, these trends are likely due to problems with other model components, such as the very high weights given to composition data. We consider very low steepness unlikely: Moore et al. (2017) provided clear evidence of king threadfin's strong density-dependent response to depletion. We suggest a range from 0.55 to 0.95. This would represent the substantial uncertainty about steepness, and result in a median value close to the 0.75 estimate used in the 2021 assessment.

5 Model structure

The model is implemented in Stock Synthesis, which is a well-developed modeling platform that provides a wide range of options for setting up the model, and excellent diagnostic tools. It is an appropriate tool to use for this assessment.

5.1 Growth

The early age-at-length data fit badly, with consistently more old fish than expected. The otoliths are thought to have been aged correctly. They may have been sampled from a catchment with low fishing pressure, but this is unlikely for multiple years of data.

There are a number of possible reasons for these observations, which are not mutually exclusive. The early population may be less depleted than the model predicts. Another possible explanation is that the somatic growth rate of the population has increased in recent years as a density-dependent response to depletion. The growth curve is estimated inside the model, mostly informed by recent data since 2015. If growth was slower for fish sampled in the 1986-1993 period, they would have smaller lengths at age than currently, and older ages at length.

This hypothesis can be tested by comparing otoliths from the current period with the earlier samples, using similar methods to Moore et al. (2017). It would also be useful to identify otoliths from the current sampling program that come from different catchments, to explore whether growth varies among catchments, and what are the covariates if so.

5.2 Recruitment

Apply lognormal bias correction (Methot and Taylor, 2011). This will likely have a very small effect, but it is good practice to apply it.

No relationship is assumed between recruitment and river flow. Relationships between recruitment and environmental variables are difficult to estimate and use in stock assessments (Haltuch et al., 2019; Maunder and Thorson, 2019). The model has a lot of flexibility to estimate recruitment, and age data can help the model provide good estimates. However, the temporal coverage of age and size data is very limited and there is considerable medium-term autocorrelated variation in flows through time, which may drive large shifts in recruitment. A strong relationship between flow and recruitment might considerably impact biomass through time.

Mention of healthy recruitment during an extremely severe drought in 2016-17 as "described by Duke et al. (2022)" was puzzling. Duke et al. (2022) show normal rainfall for the Karumba region in 2016-17 (Figure 2 – copied from Figure 4 in Duke et al 2022) after a period of low rainfall. Similarly, Figure 3.4 in the assessment document shows moderately low rainfall in 2016-17 but higher than the previous 3 years. This may reinforce the need for the analysts to consider alternative approaches for estimating river flow and its relationships with both recruitment and catchability.

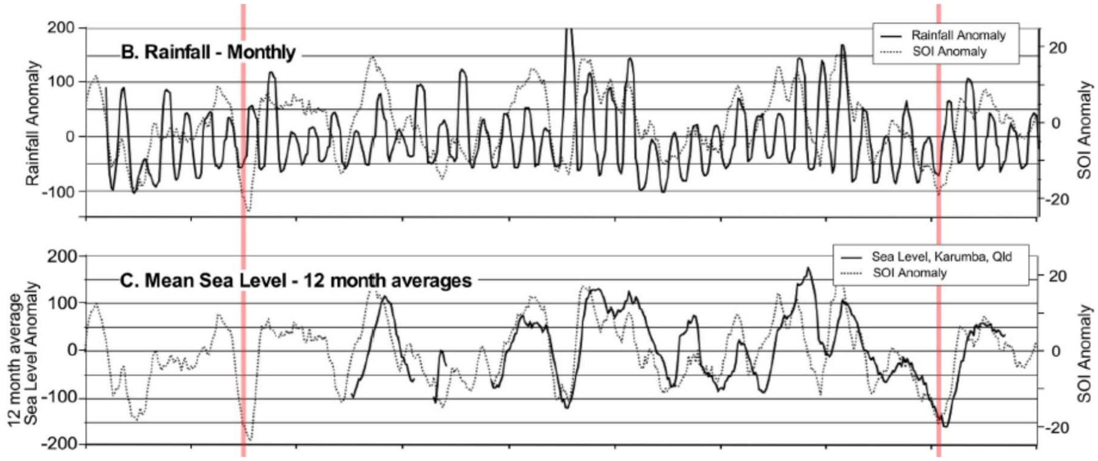

Figure 2: Timelines of influencing variables shown in climate (B), and sea level change anomaly (C) compared to the SOI anomaly for the severely impacted site near Karumba in the Gulf of Carpentaria (E). Vertical red lines indicate major desiccation events in 1982, and in 2015 when there was major loss of mangrove canopy density.

5.3 Fleet definitions

The model is structured using an unconventional version of an areas-as-fleets approach, with one fleet for each region. Areas-as-fleets models (Punt, 2019) are often used for species where a stock shows ontogenetic movement between areas based on age or size. Estimating movement rates requires informative data – without it, multi-region models can be unstable. The areas-as-fleets approach assumes that all areas are part of a single pooled population and addresses size/age differences with selectivity. This approach can be effective (Goethel et al., 2024; Lee et al., 2017), but can also cause bias (Hurtado-Ferro et al., 2013) – it depends on the situation. The king threadfin model is unusual because it assumes the same selectivity everywhere, so there is no obvious purpose for defining separate fleets. They are used to include multiple abundance indices, but in a single region model it is not generally useful to include more than one CPUE series. Each series provides different information about the trend of the same population. When there are alternative CPUE trends, the preferred approach is for different indices to represent alternative states of nature and to use an ensemble approach (Hoyle et al., 2024). However, as discussed elsewhere in this review, a better approach may be to model each region individually.

5.4 Selectivity

The model includes 5 fleets but specifies selectivity as constant and asymptotic across all areas, time periods, and fishing sectors. As stated earlier, given the limited mixing and apparent spatial variation in fishing pressure, we recommend running a separate model for each region.

Asymptotic selectivity can be an influential assumption in determining the outcome of the assessment. True selectivity is rarely asymptotic (Sampson, 2014; Waterhouse et al., 2014), but it can help model stability to assign asymptotic selectivity to the fishery catching the largest individuals.

In this case, population-level selectivity is probably not asymptotic. If selectivity is a mixture of size-selective meshing and non-selective brailing, the end result will still be size-selective. Fishers consider that although brailing occurs, it is not how most of the catch is taken. Available figures on catch size structure by mesh size (Russell, 1988) indicate strong size selection by mesh size, though sample sizes are quite small, there is no data for 162.5 mm, and no information on the size distribution of the population being sampled.

Fishers also noted that market demand and prices for large fish above 100 cm is low, with restaurants preferring a standard size of fillet. If there is a preference to fish in locations or with fishing gear that avoid the largest fish, this would tend to make selectivity non-asymptotic.

Just as important as gear selectivity is spatial and seasonal size variation, which can affect population-level selectivity. Selectivity may vary spatially, e.g., with different size fish in rivers versus coastal.

We recommend exploring non-asymptotic selectivity for the commercial catches.

Misfits to composition data increase uncertainty and may bias the assessment. The assessment shows some misfit to both the age and length data. Some of this may be due to selectivity assumptions, but the variability between years is more likely due to spatial structuring of the population, and spatial variation in fishing pressure.

5.5 Data weighting strategy

The method described in Appendix D is novel and differs from the methods used by other fisheries scientists. It appears mathematically sophisticated, but we did not have time to examine it in detail. However, we believe that a method should be validated by simulation testing and peer review before applying it to a production assessment.

One potential concern is that the method estimates a separate weighting for each year. The description also suggests that it may be valid to apply this weighting algorithm to individual lengths of an age-at-length sample. However, data are noisy, and in normal circumstances some years (or lengths) will by chance fit better than others. This algorithm will upweight the years that fit better by chance and downweight the years that fit worse, which is wrong in principle. In contrast, the Francis method reweights an entire data series with the same adjustment. In that case the principle is that if the whole data series is based on the same sampling design, sampling will be equally reliable (or unreliable) for all years (after allowing for differences in sample sizes and other relevant factors among years), and the weighting adjustment can be estimated from the fit of the whole series. Relative sample sizes among years can be estimated outside the model – see a description of options in the data weighting section of the SS manual.

If there are periods when the sampling design changes, then it is reasonable to use those periods as separate blocks for reweighting purposes.

The effective sample sizes (ESS) produced by the algorithm (Table 1) are very high, particularly at the end of the time series. The size data, despite being very noisy, have effective sample sizes in the hundreds. For data with this level of variability and quality of fit, we would expect ESS in the ones or tens, rather than the hundreds. We also note that sample sizes for 2009 age-at-length data are still in whole numbers and were not adjusted in the data file. The Francis scalars (based on input sample sizes that have already been adjusted using the method above) for the 3 periods of length data are 0.30, 0.26, and 0.077. The Francis scalars for the 3 periods of age-at-length data are 0.44, 0.49, and 0.81.

Giving very high ESS to composition data that nevertheless fit poorly will tend to lock the model into place, giving these data priority over other data types.

Table 1: Total effective sample size by year of age at length data, after reweighting.

Year	N samples	
	Length	Age-at-length
1986	85.3	28.7
1987	84.3	
1988	65.2	14.5
1989	25.3	5.7
1990	92.6	2.5
1991	211.4	27.7
1992	3.7	4.3
1993	66.4	5.2
1994	53.6	
2008	31.2	17.2
2009	75.1	221.0
2010	70.5	17.6
2011	91.9	14.1
2015	13.1	18.8
2016	241.5	19.0
2017	348.3	31.8
2018	162.5	54.9
2019	327.4	174.4
2020	124.1	46.7
2021	644.5	306.4
2022	170.2	408.3

5.6 Ensemble scenarios

Estimates were compiled from a grid of plausible scenarios (Assessment Table 2.5) with the following settings. The term ‘plausible’ should be defined. In this case we were informed that models with steepness of 0.45 and 0.85 were excluded due to implausible sequences of recruitment deviates.

Steepness (0.55, 0.65, 0.75)

Natural mortality (0.325, 0.35, 0.4)

q_{inc}: 3 options: (spatially varying with Karumba=0, 0.01, spatially invariant with q_{inc} = 0)

6 Model outputs

6.1 Model diagnostics

The fit to the main CPUE index for the Karumba region is moderate (Figure 3). There is a trend at the start of the time series where the index does not match the estimated decline. After 2005 the residuals are more positive than negative, with the index slightly more optimistic than the estimated biomass trend. Indices for the other regions are of less interest. No trend is expected or observed, because of the estimated q_{inc} parameters.

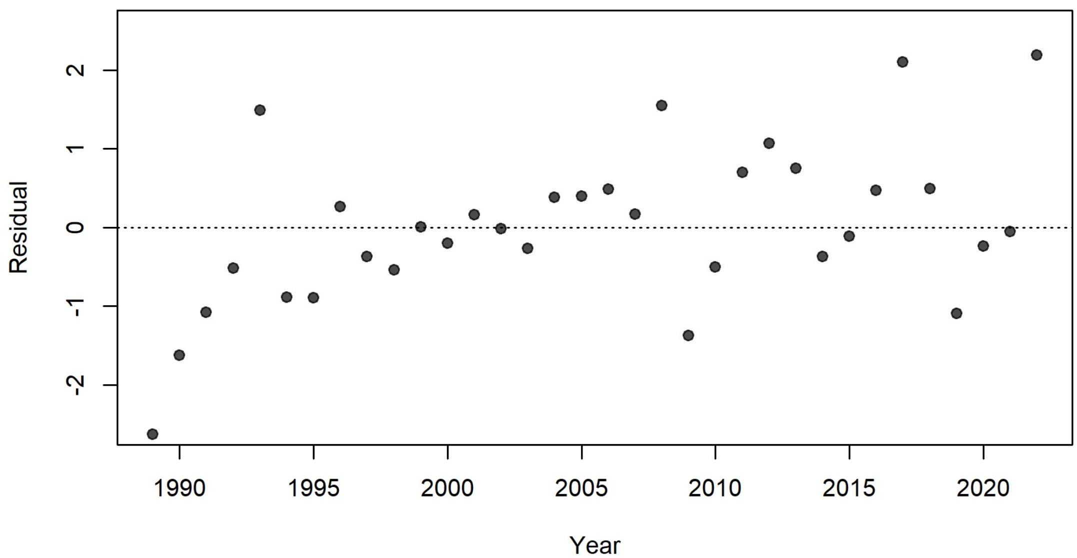

Figure 3: Residuals of fit to index for the Karumba region from the reference case model, scenario 5.

Age at length data are more interesting (Figure 4, 5, and 6). The early period shows substantial and consistent lack of fit (Figure 4). At almost all sizes and in all years there are more old fish than the model expects. In the middle period, the data fit quite well in 2009 and 2011 but there is some bias towards older fish than expected in 2008 and 2010. In contrast, most years of the later period 2015 to 2022 fit very well, though with a slight bias (except in 2016) towards fish being younger than expected. The later period samples also have considerably more statistical weight in the model than the earliest samples, with more lengths represented and higher effective sample size at each length (Table 1).

The consistently poor fit to the early size data is an unresolved conflict in the model which may be contributing to the convergence problems apparent in many scenarios. We recommend developing and testing hypotheses to address it. It is risky to use models with unresolved internal conflicts to develop management advice. If there is no clear explanation, the recommended approach is to develop alternative scenarios that resolve the conflicts, such as dropping the early composition data, or increasing qinc / reducing early catch / increasing recent catch (?) / exploring other options, to remove the conflict between predicted and observed early age-at-length data.

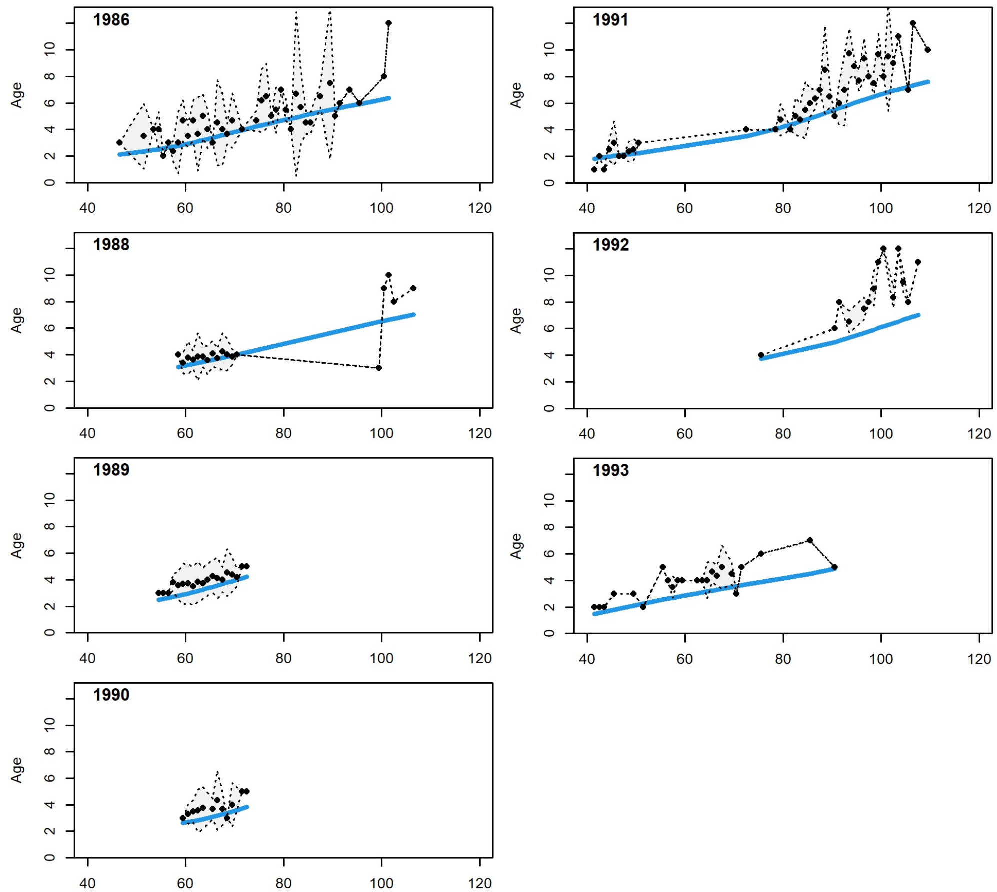

Figure 4: Conditional age at length plot for data from 1986-1993. These plots show mean age at length by size-class (obs. and exp.) with 90% CIs based on adding 1.64 SE of mean to the data.

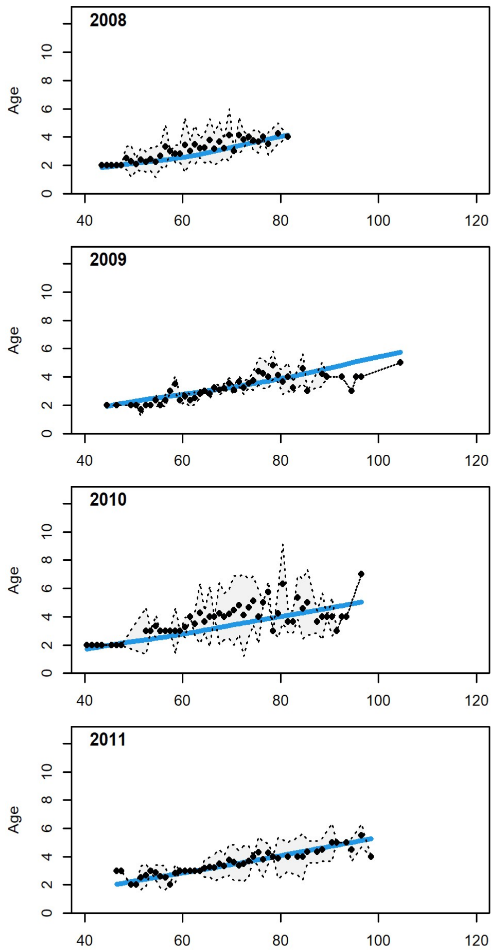

Figure 5: Conditional age at length plot for data from 2008-2011. These plots show mean age at length by size-class (obs. and exp.) with 90% CIs based on adding 1.64 SE of mean to the data.

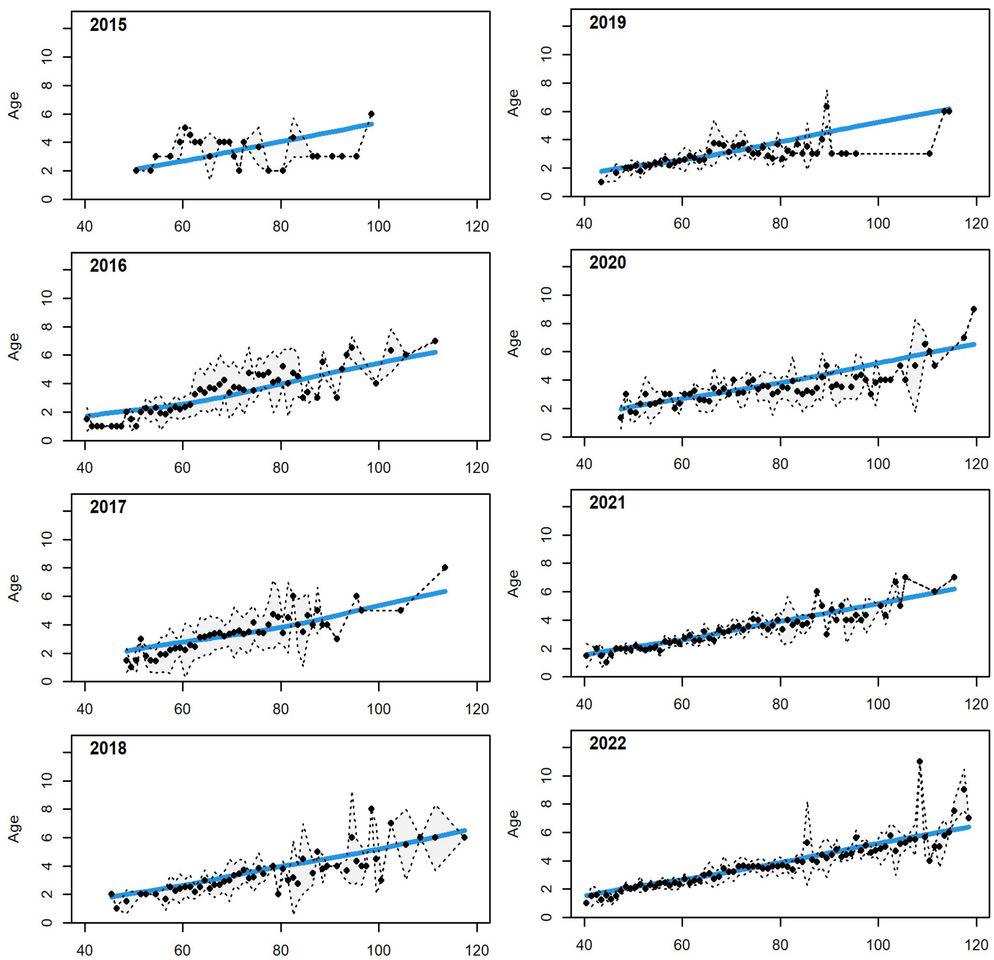

Figure 6: Conditional age at length plot for data from 2015-2022. These plots show mean age at length by size-class (obs. and exp.) with 90% CIs based on adding 1.64 SE of mean to the data.

Size data show some residual trends. There is a trend of increasing observed size from 2016-2022 which the model does not fit well.

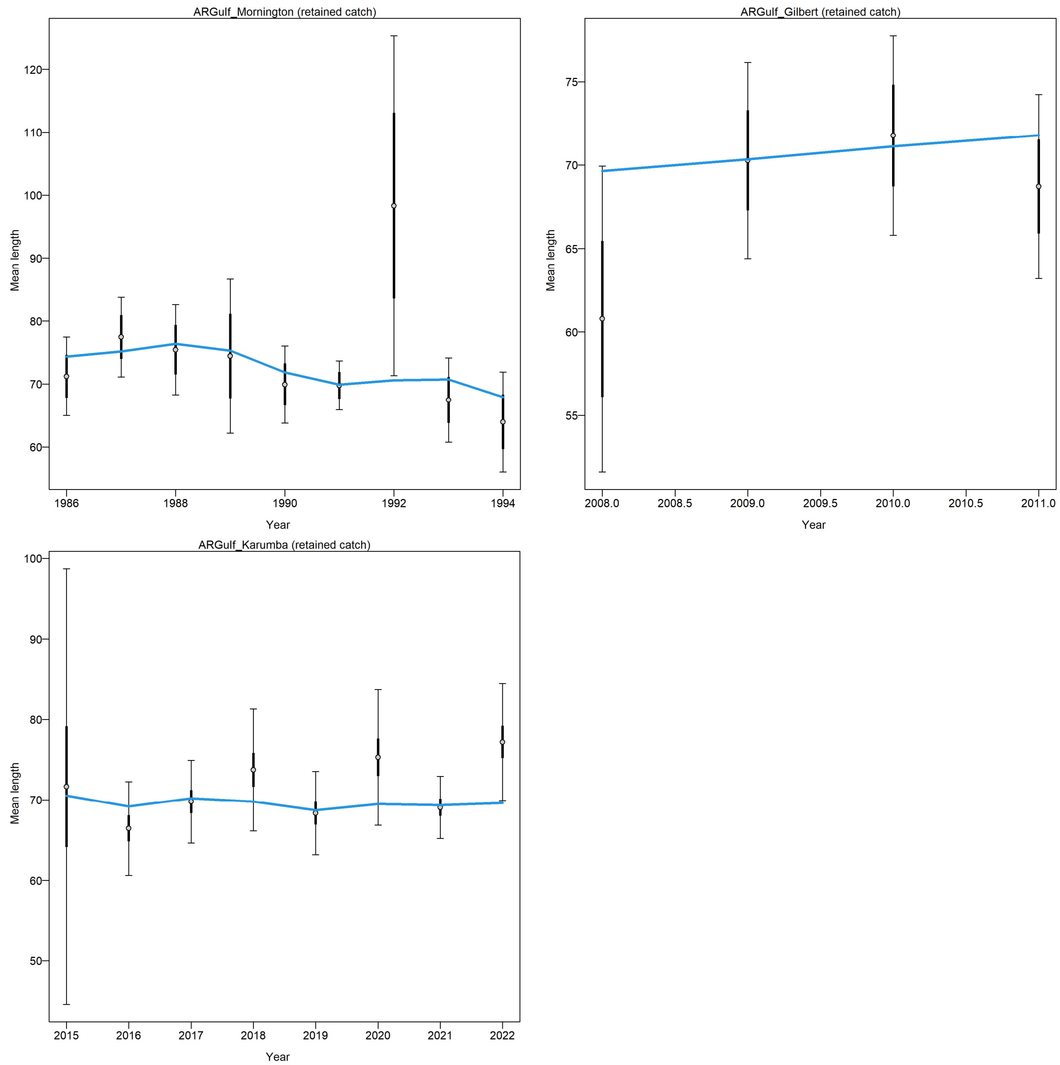

Figure 7: Mean length for the three periods of length data with 95% confidence intervals (thick bars) based on current sample sizes. The thinner intervals (with capped ends) show the results of further adjusting sample sizes based on the Francis data weighting method TA1.8, with a suggested multiplier (with 95% interval) for each period (0.3005, 0.2607, and 0.0769 respectively).

Likelihood profiles on the population scaling parameter (R0) are a well-established tool for exploring the influence of different data components in stock assessments (Lee et al., 2014). They are extremely useful for understanding which data component is driving the model.

A worrying characteristic of the likelihood profile on R0 (Figure C.8) is that scale is largely determined by recruitment penalties, at both upper and lower ends of the R0 confidence interval. This is unusual in our experience, especially with sigmaR of 0.6. It would be useful to examine the sensitivity of the model to higher values of sigmaR such as 2.0, so that the recruitment likelihood has no influence on population scale. The index provides little information to inform population scale. The age and size data also make relatively small contributions, partly because different periods of age and length data are pushing in different directions. Early age data wants to increase population scale (Figure C.12), as does the 2008-2012 age data, probably because both have more old fish than the model expects, and higher biomass would reduce fishing mortality. For the 2015-2022 period both the age-at-length data and the length composition (Figure C.11) data fit better with lower population scale.

It is likely that the dependence of the model on recruitment penalties for scaling is because recruitment is the main thing changing at different levels of the scaling parameter. This may be caused by the high penalties on the composition data.

A major advantage of using Stock Synthesis 3 for stock assessment is the availability of a wide range of diagnostic tools. Some, such as residual plots, are available automatically in r4ss (Taylor et al., 2013) but others are not. For future assessments we recommend applying diagnostics from the R package ss3diags (available at github.com/JABBAmodel/ss3diags) (Carvalho et al., 2021) and improvements as they become available, since this is an area of active development.

6.2 MCMC and convergence diagnostics

MCMC (Markov chain Monte Carlo) chains in earlier scenarios presented to the reviewers showed evidence of substantial non-convergence for many of the scenarios, with even the better scenarios showing evidence of autocorrelation. This issue needed to be resolved before MCMC could be used to provide confidence bounds. We understand that new models with better MCMC diagnostics have been developed which are less seriously affected by this issue.

Chains should also be investigated for evidence of non-convergence using multiple-chain comparisons, $\hat{r}$ statistics, and the effective sample size estimates (Vehtari et al., 2017).

Convergence diagnostics indicated that the R0 parameter for the diagnostic case (scenario 5) had a gradient of 1.4e-3, which is not fully converged to the 1e-4 criterion, and the Hessian was not positive definite. Of 27 scenarios (close to final) provided to the reviewers, only 7 scenarios (3, 9, 15, 17, 18, 22, and 23) converged to the criterion, and none generated positive definite Hessians. The models have difficulty converging to stable estimates.

Jittering would help to identify if models have reached optimal solutions. The convergence issue may be a consequence of overweighting the composition data. However, the causes remain unclear and further investigation will be needed.

6.3 Abundance and recruitment trends

Here looking at results from the reference case model (scenario 5) which has M=0.35, h=0.65, qinc set to 0 in the Karumba region and estimated elsewhere.

Recruitment estimates are characterised by runs of high and low deviates. A series of 5 high recruitments is estimated 1978-1982, followed by 3 low recruitments from 1985-1987 and 2 more in 1990-1991 (Figure 8). Estimates are moderate from 1988 through to 2008, but then there is a series of low recruitments from 2009-2015, apart from one moderate recruitment in 2014. Finally, there are 3 relatively high recruitments in 2019-2021.

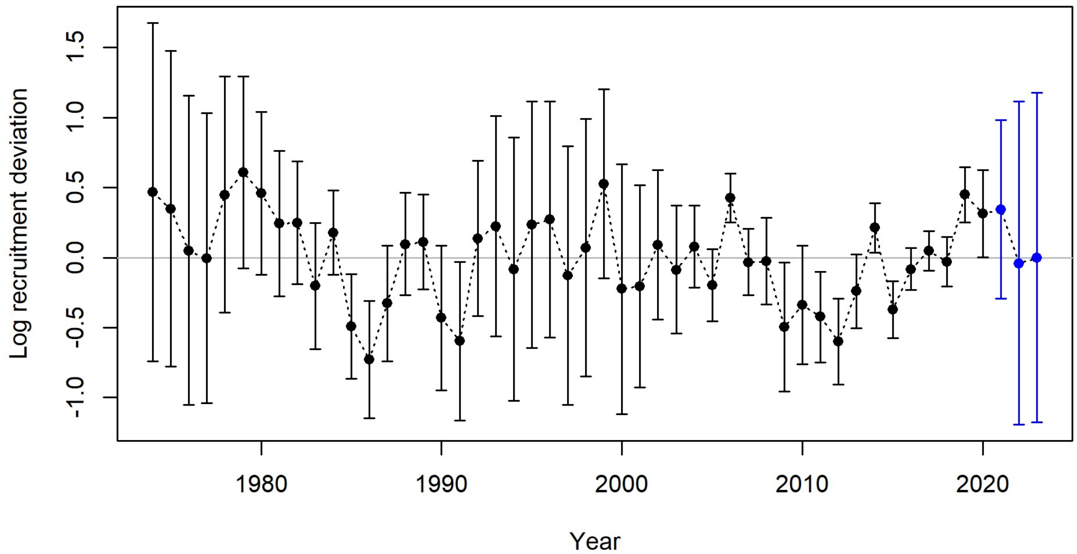

Figure 8: Recruitment deviates estimated for scenario 5, the reference case.

The impact of these recruitment trends is seen in the no-fishing plot (Figure 9). The biomass decline that starts in the mid-1970's with the surge in fishing pressure is mitigated by an increase in recruitment. Recruitment then declines and contributes to a continued decline in biomass, which reaches a high level of depletion in the mid-1990s. Following a period of median recruitment, the reduced recruitment from 2009 helps reduce biomass to its greatest level of depletion, before a subsequent increase with higher recruitment in 2019-2020.

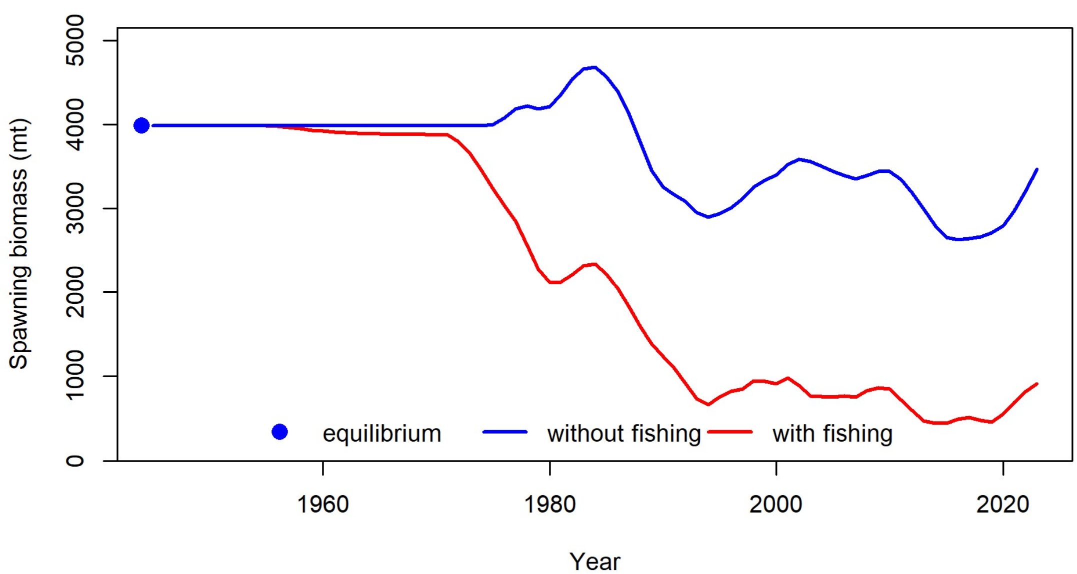

Figure 9: Dynamic B0 plot estimated for scenario 5, the reference case. The lower line shows the time series of estimated Spawning biomass (mt) in the presence of fishing mortality. The upper line shows the time series that could occur under the same dynamics (including deviations in recruitment), but without fishing. The point at the left represents the unfished equilibrium.

These runs of recruitment can be explained by the model's fits to the composition data. These data have very high statistical weight, so that the biomass trend is strongly constrained to be heavily depleted at the end of the time series (given the lack of old fish). The 1986-1993 data have less statistical weight, but the model still tries to generate as many old fish as possible to reduce the misfit. It seems likely that when R0 increases or decreases in the likelihood profile, the model estimates recruitment deviates that maintain the same biomass trend, as much as possible. It has enough flexibility to do this without much change in the composition data fits, which is why those datasets contribute much less to the likelihood profile on population scale.

From about 2010 there is a reduction in catch, but the CPUE declines at a similar time, so the model reduces recruitment to reduce biomass and predict the observed catch rate.

The apparent lack of recovery in CPUE after such a large reduction in catch (given the estimated depletion level) is consistent with a very unproductive stock, which may also explain why the model estimates low steepness – with higher steepness the recruitment deviates would need to trend more strongly to predict the lack of recovery in CPUE. However, the Karumba region CPUE trend may be affected by the spatial distribution of the effort included in the index.

7 Discussion

We have considered many aspects of the stock assessment for GoC king threadfin. It is a very interesting but difficult stock to assess, with a complex spatial domain and limited data for most of the time series. We note that a considerable amount of work has been undertaken by the stock assessment scientists, QDAF, and the scientists and fishery experts involved in the assessment.

One of the most important aspects is the spatial subdivision of the stock. As the analysts note, mixing is likely to be very slow. Each region includes many rivers and estuaries, and mixing is likely quite limited even between catchments in the same region. Fishing pressure varies spatially, given the distribution of the population of commercial and recreational fishers, the market access from particular sites such as the town of Karumba, and the need to quickly transport fish to market or to a freezer in a hot environment.

Applying spatially varying fishing pressure to a metapopulation with limited mixing will result in spatially varying levels of depletion.

The data used in the assessment do not allow for this spatial variation. The CPUE analyses allow for regional differences, but not for differences within regions. The length frequency and age-at-length data are aggregated across all regions. This should, almost inevitably, result in more composition data and more CPUE data from areas that are more depleted.

There are various possible solutions to this problem. The first, which is impractical, is to conduct a separate assessment for each important catchment. The second approach is to assess at the regional level, but to reweight the data so that it allows for different levels of fishing pressure and depletion by catchment. This is mainly a modelling exercise for the CPUE, though it will need an estimate of proxies for habitat area. For composition data there will be a need for more accurate information in future on fishing locations, and an effort to recover fishing locations of data already sampled.

For data weighting, we recommend using the Francis method. The method used here is unproven and has provided unusually high effective sample sizes for composition data. The size data are given more statistical weight than the age-at-length data, but we think the age-at-length data are likely to be more informative than the size data – as indicated by the estimated Francis weights.

The estimated biomass trend and stock status appear to be generally consistent with the catch history, the CPUE series, and the current truncated age structure. However, there remains uncertainty associated with some unresolved issues. Overweighting of data from areas with more fishing would tend to overestimate the current level of depletion for the overall stock, since the areas that are most heavily fished will be the most depleted. From the information available to us the locations of these areas are uncertain, but they may be close to Karumba. On the other hand, Karumba region CPUE indices may be missing the effects of increased fishing power and increased targeting of threadfin in recent years, which would tend to have the opposite effect on stock status estimates. Consequences for stock status of addressing the poor fit to early age-at-length data and the overweighting of composition data are uncertain but may be significant, and addressing these issues may help the model to converge.

8 Recommendations

1. Expand the data characterisation analyses as much as possible, including but not limited to the following.
    a. Investigate monitoring data for spatial and temporal patterns in age and size structure.
    b. Show representativeness of sampling through time with respect both to the catch and to locations.
2. Revise CPUE series.
    a. Use an error distribution model that will allow full diagnostics.
    b. Provide enough diagnostics to determine whether the model is appropriate for the data.
    c. Explore alternative targeting analyses based on cluster analysis.
    d. Conduct separate analyses by region.
    e. Apply models that allow for spatial and spatiotemporal variation in catch rates within regions.
    f. Consider using a revised river flow dataset.
    g. Consider running models with the R package mgcv.
3. Data preparation.
    a. Allow for spatial structure in size and age comps when developing composition datasets, since the model assumes that catch is sampled uniformly from the whole population.
4. Biological inputs
    a. Use a range of steepness values from 0.55 to 0.95, and sigmaR of 0.6.
    b. Consider Lorenzen M.
5. Modeling
    a. Apply standard Francis data weighting methods.
    b. Run a separate assessment model for each region.
    c. Identify hypotheses to explain why the model cannot fit the ages-at-length in the early period, and develop alternative models based on these.
    d. Ensure that issues with incomplete convergence, no Hessian, and poor MCMC diagnostics are resolved.
    e. Allocate composition data to the region where it was sampled.
    f. Consider non-asymptotic selectivity.
    g. Include month of sampling in the composition data.

9 References

Brooks, E.N. Pragmatic approaches to modeling recruitment in fisheries stock assessment: A perspective. Fisheries Research. 270:106896; 2024.

Carvalho, F.; Winker, H.; Coutrtney, D.; Kapur, M.; Kell, L.; Cardinale, M.; Schirripa, M.; Kitakado, T.; Yemane, D.; Piner, K.; Maunder, M.; Taylor, I.; Wetzel, C.; Doering, K.; Johnson, K.; Methot, R. A Cookbook for Using Model Diagnostics in Integrated Stock Assessments. Fisheries Research. in press; 2021.

Darcey, J. Oral history of the Australian Fishing Industry. Queensland. Final Report to FRDC for Project Number 1987-02190 Fisheries Research and Development Corporation (FRDC); 2015

Duke, N.C.; Mackenzie, J.R.; Canning, A.D.; Hutley, L.B.; Bourke, A.J.; Kovacs, J.M.; Cormier, R.; Staben, G.; Lymburner, L.; Ai, E. ENSO-driven extreme oscillations in mean sea level destabilise critical shoreline mangroves—An emerging threat. PLoS Climate. 1:e0000037; 2022.

Garrett, R.N. Biology and harvest of tropical fishes in the Queensland Gulf of Carpentaria gillnet fishery: Queensland Department of Primary Industries; 1997

Goethel, D.R.; Berger, A.M.; Hoyle, S.D.; Lynch, P.D.; Barceló, C.; Deroba, J.; Ducharme-Barth, N.D.; Dunn, A.; Fu, D.; Izquierdo, F. 'Drivin'with your eyes closed': Results from an international, blinded simulation experiment to evaluate spatial stock assessments. Fish and Fisheries; 2024.

Halliday, I.A.; Robins, J.B.; Mayer, D.G.; Staunton-Smith, J.; Sellin, M.J. Effects of freshwater flow on the year-class strength of a non-diadromous estuarine finfish, king threadfin (Polydactylus macrochir), in a dry-tropical estuary. Marine and Freshwater Research. 59:157-164; 2008.

Haltuch, M.; Brooks, E.; Brodziak, J.; Devine, J.; Johnson, K.; Klibansky, N.; Nash, R.; Payne, M.; Shertzer, K.; Subbey, S. Unraveling the recruitment problem: A review of environmentally-informed forecasting and management strategy evaluation. Fisheries Research. 217:198-216; 2019.

Hamel, O.S.; Cope, J.M. Development and considerations for application of a longevity-based prior for the natural mortality rate. Fisheries Research. 256:106477; 2022.

Hartig, F. DHARMa: residual diagnostics for hierarchical (Multi-Level/Mixed) regression models. 2020

He, X.; Bigelow, K.A.; Boggs, C.H. Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. Fisheries Research. 31:147-158; 1997.

Horne, J.; Momigliano, P.; Welch, D.; Newman, S.J.; van Herwerden, L. Stock structure of *Polydactylus macrochir* across northern Australia based on mtDNA. in: Welch D.J., Ballagh A., Newman S.J., Lester R., Moore B., van Herwerden L., Horne J., Allsop Q., Saunders T., Stapley J., Gribble N.A., eds. Defining the stock structure of northern Australia's threadfin salmon species: Final Report to the Fisheries Research & Development Corporation, Project 2007/032. Fishing & Fisheries Research Centre, James Cook University, Townsville, Qld; 2010

Hoyle, S.D.; Campbell, R.A.; Ducharme-Barth, N.D.; Grüss, A.; Moore, B.R.; Thorson, J.T.; Tremblay-Boyer, L.; Winker, H.; Zhou, S.; Maunder, M.N. Catch per unit effort modelling for stock assessment: A summary of good practices. Fisheries Research. 269:106860; 2024.

Hurtado-Ferro, F.; Punt, A.E.; Hill, K.T. Use of multiple selectivity patterns as a proxy for spatial structure. Fisheries Research; 2013. 10.1016/j.fishres.2013.10.001

Infofish. Weipa tagging 1985-2014. Infofish Australia; 2014

Klaer, N. Independent review of the stock assessment of King threadfin (*Polydactylus macrochir*) in Queensland, Australia.: Prepared for Queensland Department of Agriculture and Fisheries; 2021

Kritzer, J.P.; Sale, P.F. Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. Fish and Fisheries. 5:131-140; 2004.

Lambert, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 34:1-14; 1992. Doi 10.2307/1269547

Lee, H.-H.; Piner, K.R.; Maunder, M.N.; Taylor, I.G.; Methot Jr, R.D. Evaluation of alternative modelling approaches to account for spatial effects due to age-based movement. Canadian Journal of Fisheries and Aquatic Sciences. 74:1832-1844; 2017.

Lee, H.-H.; Piner, K.R.; Methot Jr, R.D.; Maunder, M.N. 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. 158:138-146; 2014.

Lee, H.H.; Maunder, M.N.; Piner, K.R.; Methot, R.D. Can steepness of the stock-recruitment relationship be estimated in fishery stock assessment models? Fisheries Research; 2012.

Leigh, G.M.; Campbell, A.B.; Tanimoto, M.; Whybird, O.J. Companion document to 'Stock assessment of king threadfin (*Polydactylus macrochir*) in the Gulf of Carpentaria, Queensland, Australia, with data to December 2022'. Fisheries Queensland, Department of Agriculture and Fisheries; 2024a

Leigh, G.M.; Campbell, A.B.; Tanimoto, M.; Whybird, O.J. Stock assessment of king threadfin

(*Polydactylus macrochir*) in the Gulf of Carpentaria, Queensland, Australia, with data to December 2022. Fisheries Queensland, Department of Agriculture and Fisheries; 2024b

Leigh, G.M.; Tanimoto, M.; Whybird, O.J. Stock assessment of king threadfin (Polydactylus macrochir) in Queensland, Australia. 2021.

Leroy, B.; Nicol, S.; Lewis, A.; Hampton, J.; Kolody, D.; Caillot, S.; Hoyle, S. Lessons learned from implementing three, large-scale tuna tagging programmes in the western and central Pacific Ocean. Fisheries Research; 2013.

Li, S.; Yang, F.; Famoye, F.; Lee, C.; Black, D. Quasi-negative binomial distribution: Properties and applications. Computational Statistics & Data Analysis. 55:2363-2371; 2011.

Lorenzen, K. Size-and age-dependent natural mortality in fish populations: Biology, models, implications, and a generalized length-inverse mortality paradigm. Fisheries Research. 255:106454; 2022.

Maunder, M.N.; Hamel, O.S.; Lee, H.-H.; Piner, K.R.; Cope, J.M.; Punt, A.E.; Ianelli, J.N.; Castillo-Jordán, C.; Kapur, M.S.; Methot, R.D. A review of estimation methods for natural mortality and their performance in the context of fishery stock assessment. Fisheries Research. 257:106489; 2023.

Maunder, M.N.; Thorson, J.T. Modeling temporal variation in recruitment in fisheries stock assessment: a review of theory and practice. Fisheries Research. 217:71-86; 2019.

Maunder, M.N.; Thorson, J.T.; Xu, H.; Oliveros-Ramos, R.; Hoyle, S.D.; Tremblay-Boyer, L.; Lee, H.H.; Kai, M.; Chang, S.-K.; Kitakado, T. The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. Fisheries Research. 229:105594; 2020. 10.1016/j.fishres.2020.105594

McPherson, G.R. Reproductive biology of five target fish species in the Gulf of Carpentaria inshore gillnet fishery. in: Garrett R.N., ed. Biology and harvest of tropical fishes in the Queensland Gulf of Carpentaria gillnet fishery Final Report of FRDC project no 92/145. Brisbane: Queensland Department of Primary Industries; 1997

Methot, R.D., Jr; Taylor, I.G. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Canadian Journal of Fisheries and Aquatic Sciences. 68:1744-1760; 2011.

Miller, T.J.; Brooks, E.N. Steepness is a slippery slope. Fish and Fisheries. 22:634-645; 2021.

Moore, B.; Welch, D.; Newman, S.; Lester, R. Parasites as indicators of movement and population connectivity of a non-diadromous, tropical estuarine teleost: king threadfin Polydactylus macrochir. Journal of Fish Biology. 81:230-252; 2012a.

Moore, B.; Welch, D.J.; Simpfendorfer, C.A. Spatial patterns in the demography of a large estuarine teleost: king threadfin, Polydactylus macrochir. Marine and Freshwater Research. 62:937-951; 2011.

Moore, B.R. Movement, connectivity and population structure of a large, non-diadromous, tropical estuarine teleost. James Cook University; 2011

Moore, B.R.; Simpfendorfer, C.A. Assessing connectivity of a tropical estuarine teleost through otolith elemental profiles. Marine Ecology Progress Series. 501:225-238; 2014.

Moore, B.R.; Simpfendorfer, C.A.; Newman, S.J.; Stapley, J.M.; Allsop, Q.; Sellin, M.J.; Welch, D.J. Spatial variation in life history reveals insight into connectivity and geographic population structure of a tropical estuarine teleost: king threadfin, Polydactylus macrochir. Fisheries Research. 125:214-224; 2012b.

Moore, B.R.; Stapley, J.M.; Williams, A.J.; Welch, D.J. 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; 2017.

Newman, S.J.; Allsop, Q.; Ballagh, A.C.; Garrett, R.N.; Gribble, N.; Meeuwig, J.J.; Mitsopoulos, G.E.; Moore, B.R.; Pember, M.B.; Rome, B.M. Variation in stable isotope (δ18O and δ13C) signatures in the sagittal otolith carbonate of king threadfin, Polydactylus macrochir across northern Australia reveals multifaceted stock structure. Journal of Experimental Marine Biology and Ecology. 396:53-60; 2010.

Punt, A.E. Spatial stock assessment methods: A viewpoint on current issues and assumptions. Fisheries Research. 213:132-143; 2019.

Punt, A.E.; Dunn, A.; Elvarsson, B.Þ.; Hampton, J.; Hoyle, S.D.; Maunder, M.N.; Methot, R.D.; Nielsen, A. Essential features of the next-generation integrated fisheries stock assessment package: A perspective. Fisheries Research. 229:105617; 2020.

R Core Team. R; A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2023

Russell, D.J. An assessment of the east Queensland inshore gillnet fishery: Department of Primary Industries, State of Queensland; 1988

Sampson, D.B. Fishery selection and its relevance to stock assessment and fishery management. Fisheries Research. 158:5-14; 2014.

Strøm, J.F.; Rikardsen, A.H.; Campana, S.E.; Righton, D.; Carr, J.; Aarestrup, K.; Stokesbury, M.J.; Gargan, P.; Javierre, P.C.; Thorstad, E.B. Ocean predation and mortality of adult Atlantic salmon. Scientific reports. 9:1-11; 2019.

Taylor, I.; Stewart, I.; Hicks, A.; Garrison, T.; Punt, A.; Wallace, J.; Wetzel, C.; Thorson, J.; Takeuchi, Y.; Monnahan, C. r4ss: R code for Stock Synthesis. R package version. 1; 2013.

Then, A.Y.; Hoenig, J.M.; Hall, N.G.; Hewitt, D.A. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science. 72:82-92; 2015.

Vehtari, A.; Gelman, A.; Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing. 27:1413-1432; 2017.

Walters, C. Folly and fantasy in the analysis of spatial catch rate data. Canadian Journal of Fisheries and Aquatic Sciences. 60:1433-1436; 2003.

Waterhouse, L.; Sampson, D.B.; Maunder, M.; Semmens, B.X. Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. Fisheries Research; 2014.

Welch, D.J.; Ballagh, A.; Newman, S.J.; Lester, R.; Moore, B.; van Herwerden, L.; Horne, J.; Allsop, Q.; Saunders, T.; Stapley, J.; Gribble, N.A. Defining the stock structure of northern Australia's threadfin salmon species. FRDC; 2010

Wood, S.N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 73:3-36; 2011.

Zhou, S.; Punt, A.E.; Lei, Y.; Deng, R.A.; Hoyle, S.D. Identifying spawner biomass per-recruit reference points from life-history parameters. Fish and Fisheries. 21:760-773; 2020.

Annex A: Contract details

Schedule 1 – Requirements

The Supplier must provide the Goods/Services specified below to the Customer, in accordance with the Requirements described in this Schedule.

Requirements for Services

The Queensland Department of Agriculture and Fisheries is seeking an independent review of the stock assessment of king threadfin (*Polydactylus macrochir*, KTF) in the Gulf of Carpentaria (GoC), Queensland, Australia.

The assessment aimed to collate all key data for stock assessment and to provide estimates of stock status (spawning biomass ratio and fishing pressure) as of the final year of data in 2022. The stock assessment was conducted using Stock Synthesis (SS) software.

The objective of the independent review is to evaluate the appropriateness of the data, analyses, and inferences in the stock assessment.

The document and independent reviewer identification will be made publicly available.

1. Contract Scope

Specifics to review include:

1. Comment on all aspects of the catch per unit effort modelling, including but not limited to: the fishery was appropriately defined (species, fleets, spatial and temporal scope), the data has been thoroughly explored and prepared (well characterised, thorough assessment of misreporting and sources of bias, appropriate level of aggregation, reasoned identifications of catchability and density covariates, reasoned inclusion of environmental variables), the modelling framework has been well chosen and appropriately used (statistical framework choice, handling of spatial structure, handling of correlation structure and error distributions, handling of zeros), the model fitting has been appropriately conducted (estimation of uncertainty, model diagnostics, model selection), the final indices input to the population model have been appropriately assembled.
2. Comment on all aspects of the historical catch reconstruction which formed the basis for the dead catch time series input to the population model, including the assumed start year and sufficient consideration of plausible catch history scenarios.
3. Comment on all aspects of the preparation of age and length data which forms the basis for the length-composition and conditional age-at-length composition data sets input to the population model, including the appropriateness of this data preparation given the sampling design protocols that governed how the age and length data were collected.
4. Comment on all aspects of the population modelling, including but not limited to: appropriate choice of modelling framework (in this case Stock Synthesis), appropriate handling of spatial and stock structure, appropriate handling of process error, appropriate handling of the key processes of growth, selectivity, natural mortality and recruitment, appropriate handling of other relevant biological settings (sex ratio, sex change, maturity, fecundity, steepness), appropriate handling of (for relevant scenarios)

model-based catchability increase parameters, appropriate choices with regard to the inclusion or not of environmental covariates in the model, appropriate handling of priors, appropriate and sufficient model diagnostics, in particular diagnostics that relate to the relative weighting of catch per unit effort and age-length data sets, appropriateness of the approach used to the characterisation of uncertainty (including within-scenario and across multiple scenarios where relevant).

5. Assess the Stock Synthesis model configuration settings, results, and diagnostics performed. Were they adequate to achieve the assessment objectives for stock status?
6. Comment on the accuracy and reliability of key statements in the report summary and conclusion. How well were they supported by the data, analysis, and literature?
7. Make recommendations for additional analyses to support the stock assessment.
8. Provide comment on any other important aspects you see or that the report should have completed, provided they relate to the estimation of stock status.

2. Supplier Method/Service Approach

The Supplier must detail their proposed methodology and approach to delivering the contract requirements (including project planning, supply / delivery scheduling, resourcing plans etc

Resources allocated to the work will include 15 days of time by Simon Hoyle (SH) and 3 days by Alistair Dunn (AD). The work will be conducted as a desktop study, with SH and AD based in New Zealand.

Data Verification and Validation: We will verify and validate all data used in the stock assessment, including but not limited to catch and effort data, reconstructed catch data, biological data, composition data, and other relevant information. This verification process will involve examining data sources, assessing data preparation, analysis methods, and key assumptions, and identifying any inconsistencies and potential biases.

Model Evaluation: We will evaluate the models and methodologies used in the stock assessment, assessing their appropriateness for the target species, ecosystem, and fishery given the available data, and their suitability for estimating stock status. This evaluation will include an examination of model assumptions, parameterization, sensitivity analyses, and model performance evaluation.

Comparison with Best Practices: We will compare the stock assessment methodology and results with established best practices, guidelines, and protocols. This comparison will help identify any deviations or deficiencies in the current assessment approach and recommend improvements accordingly.

Documentation and Reporting: We will document all findings and recommendations in a comprehensive report. This report will include a detailed description of the review methodology, key findings, identified strengths and weaknesses of the stock assessment, recommendations for improvement, and any other relevant information.

Client Consultation: Throughout the review process, we will engage in regular communication and consultation with the Client to ensure alignment with project goals, address any concerns or questions, and facilitate the timely completion of the review.

Stakeholder consultation: During the review we will consult with stakeholders in order to understand fishing methodology and changes through time, and to canvas stakeholder views about possible sources of bias in the assessment. Client documents will not be shared with stakeholders except with the express permission of the Client.

We agree to perform the review with professionalism, diligence, and expertise, adhering to the highest standards of scientific integrity and ethical conduct.

3. Timeframes

A document of the findings of the review is to be provided to Fisheries Queensland by 3pm 20th March 2024.

4. Requirements and Deliverables

The Supplier must detail their plan to provide draft review responses to be checked and accepted by DAF prior to final submission.

A draft version of the review document will be provided to Fisheries Queensland by 3pm 15th March 2024.

Annex B: Additional figures provided by stock assessment analysts

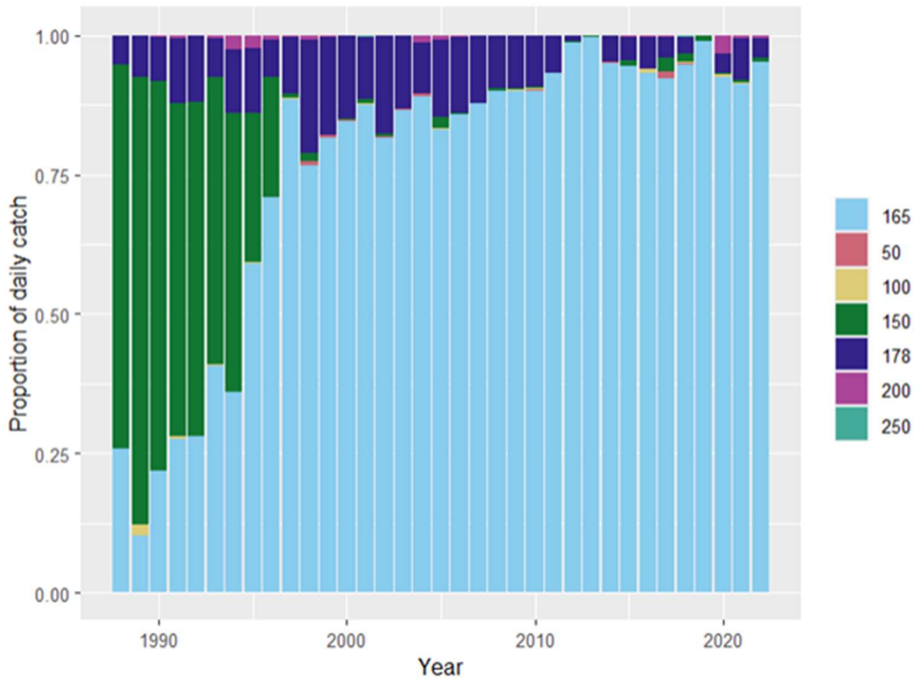

Figure 10: Characterization of mesh size distribution over time.

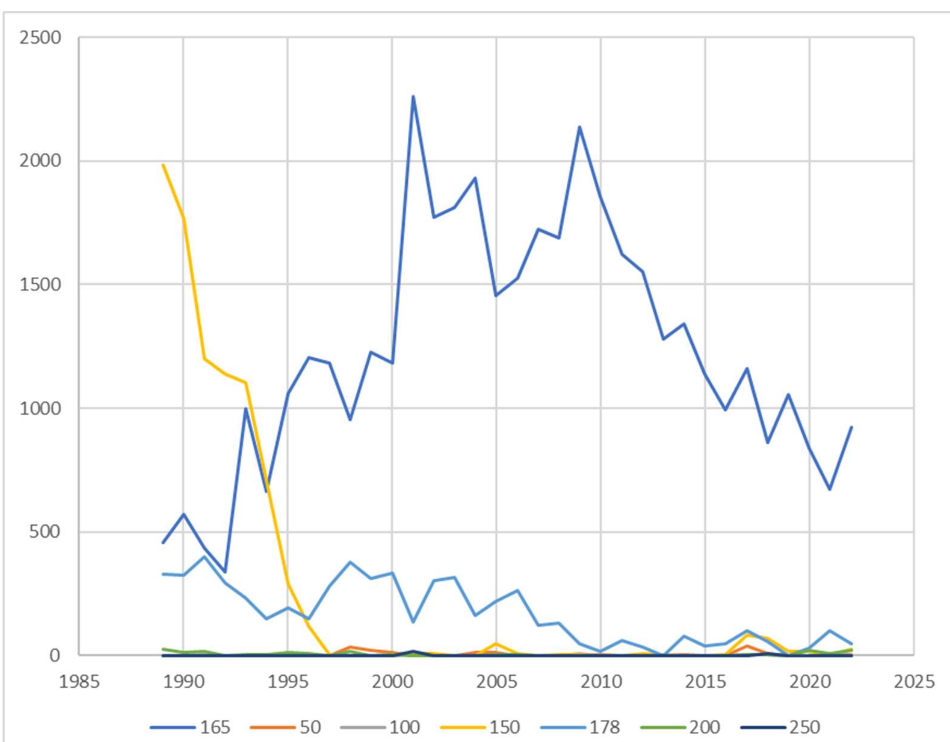

Figure 11: Catch records by mesh sizes per year.

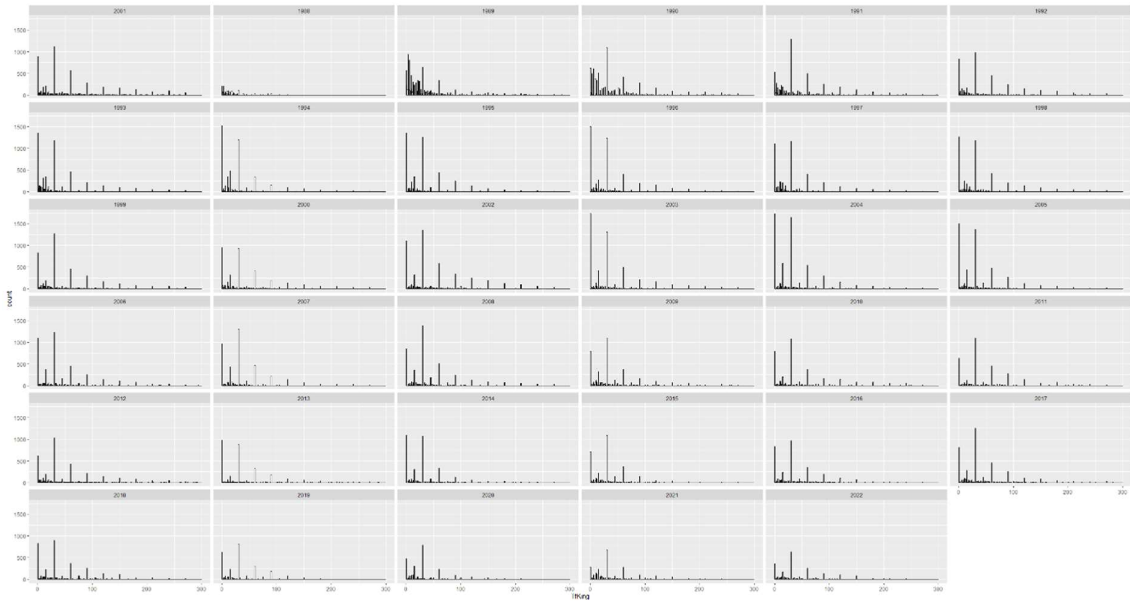

Figure 12: Distribution of catch per fisher per day by year, at 1 kg resolution, showing the high proportion of daily catches reported at 0 kg, 30 kg, and multiples of 30 kg.

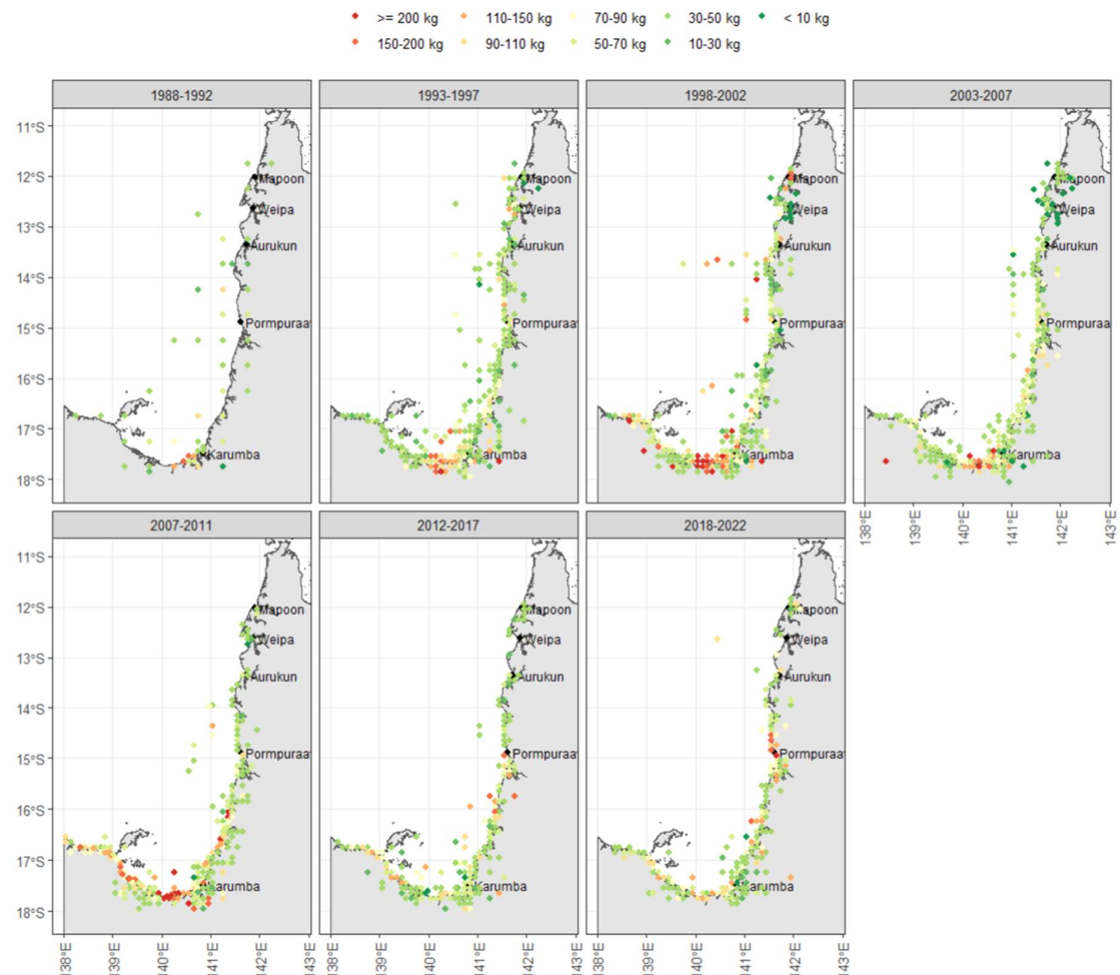

Figure 13: Maps of mean CPUE by CFISH cell across a 5-year period.

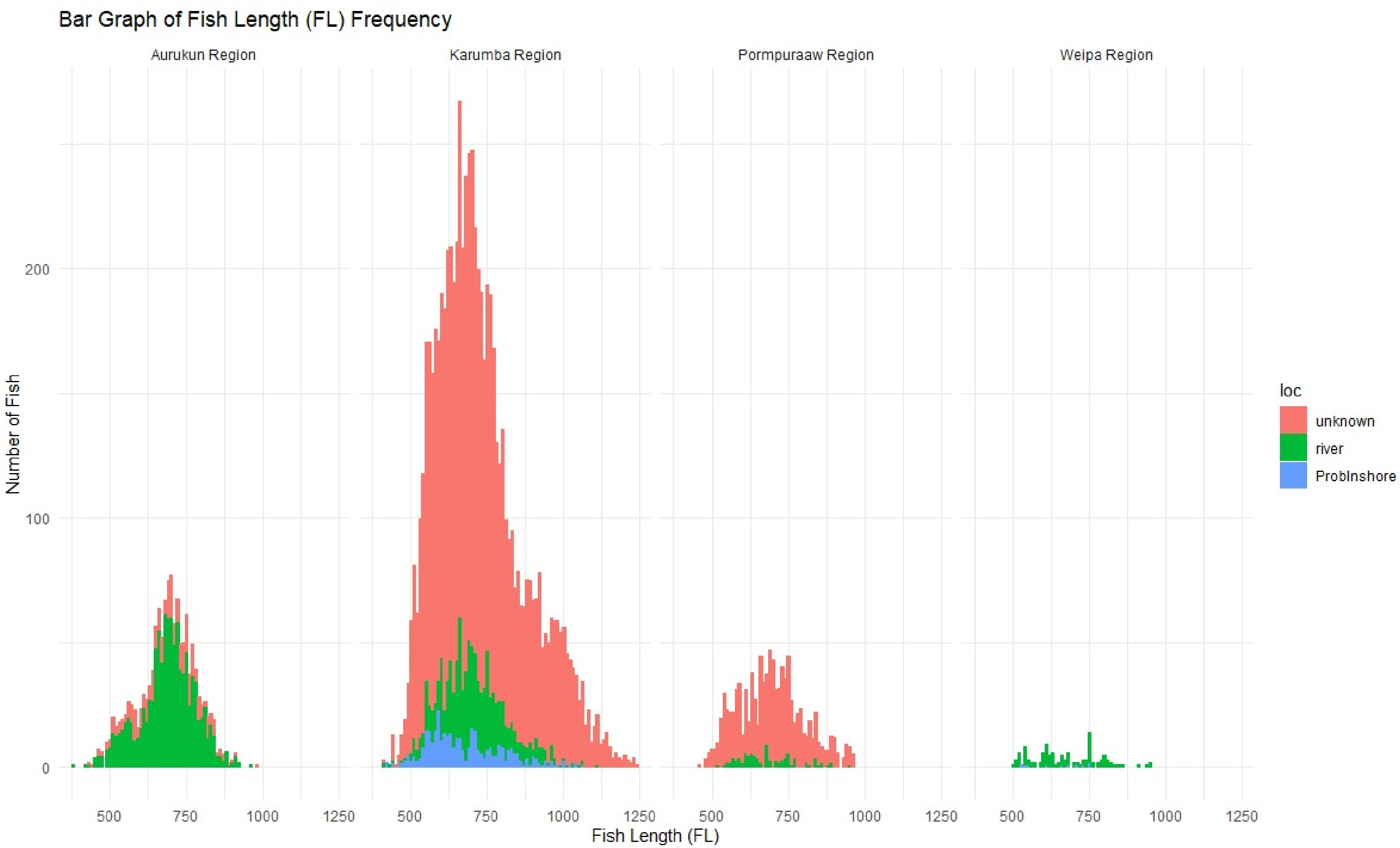

Figure 14: Aggregated data on lengths of fish sampled by region, categorised by catch location into river (green), probably inshore (blue), and unknown (red).

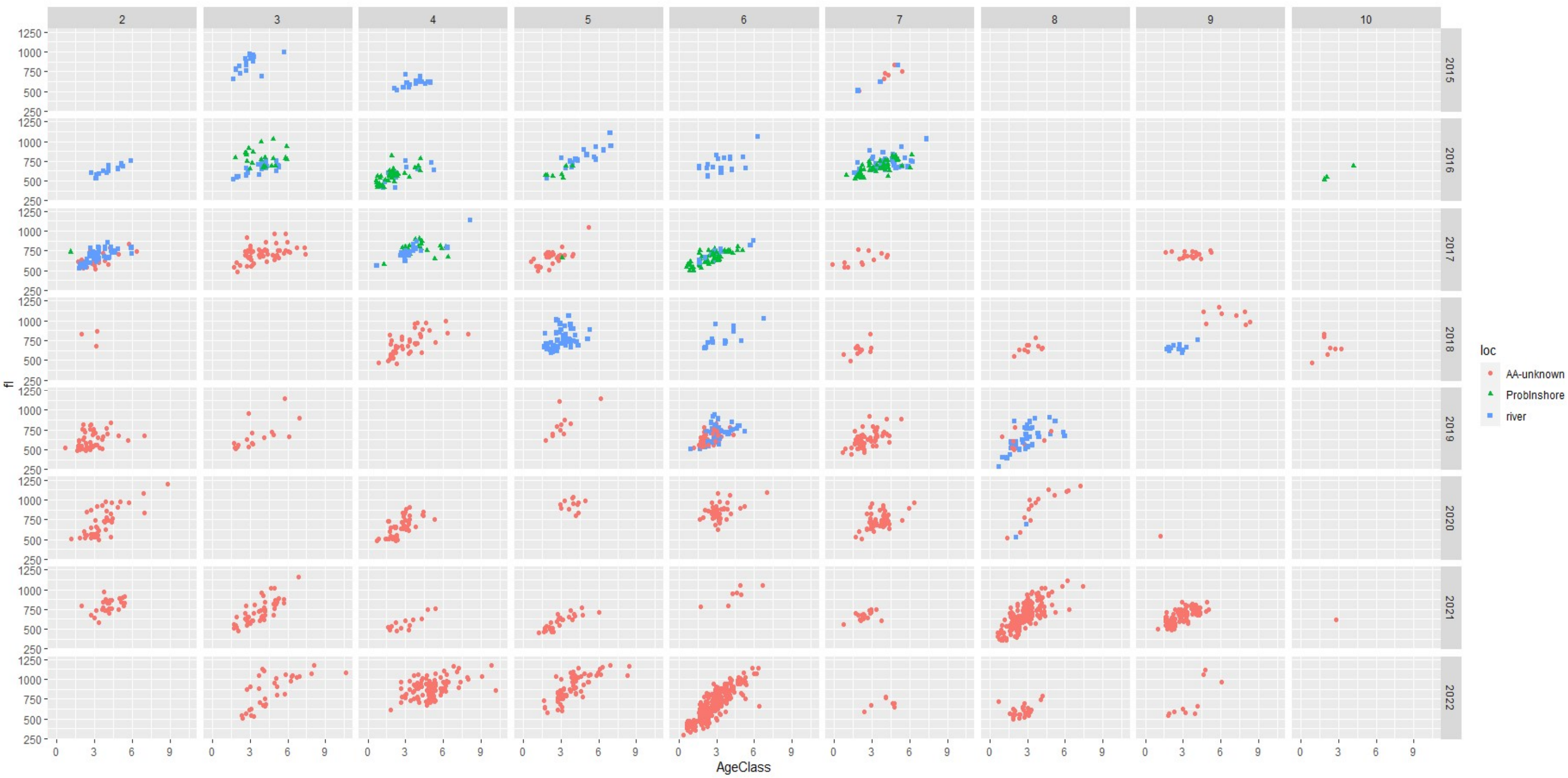

Figure 15: Lengths at age by year and month of fish sampled in the Karumba region, categorised by catch location into river (blue), probably inshore (green), and unknown (red).

Figure 16: Lengths by year and month of fish sampled in the Karumba region, categorised by catch location into river (red) and probably inshore (blue), with unknown locations omitted for clarity.

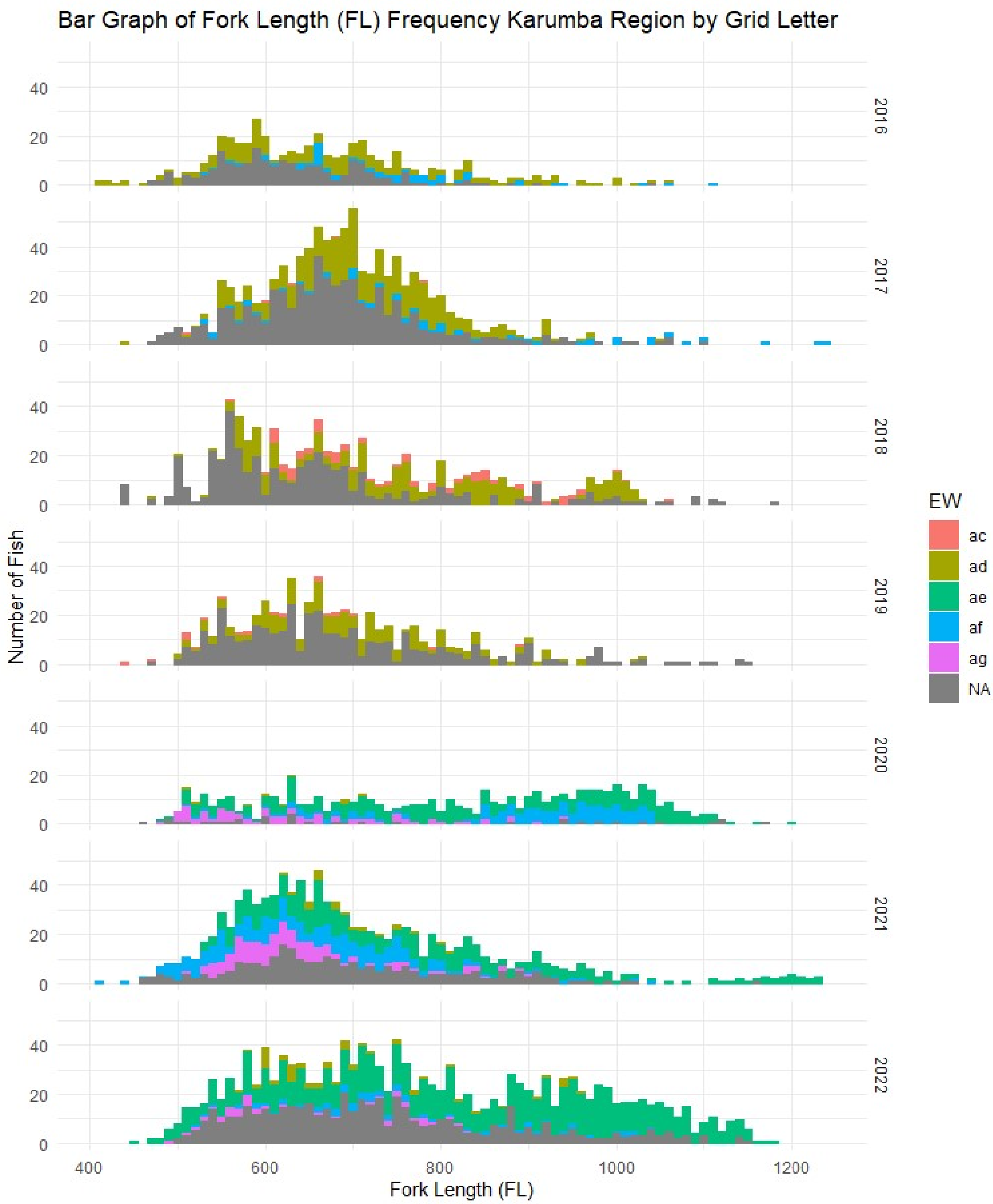

Figure 17: Length frequency by year (rows) and reporting grid (colours) for fish sampled in the Karumba region. The reporting grids represent different locations along the coast.

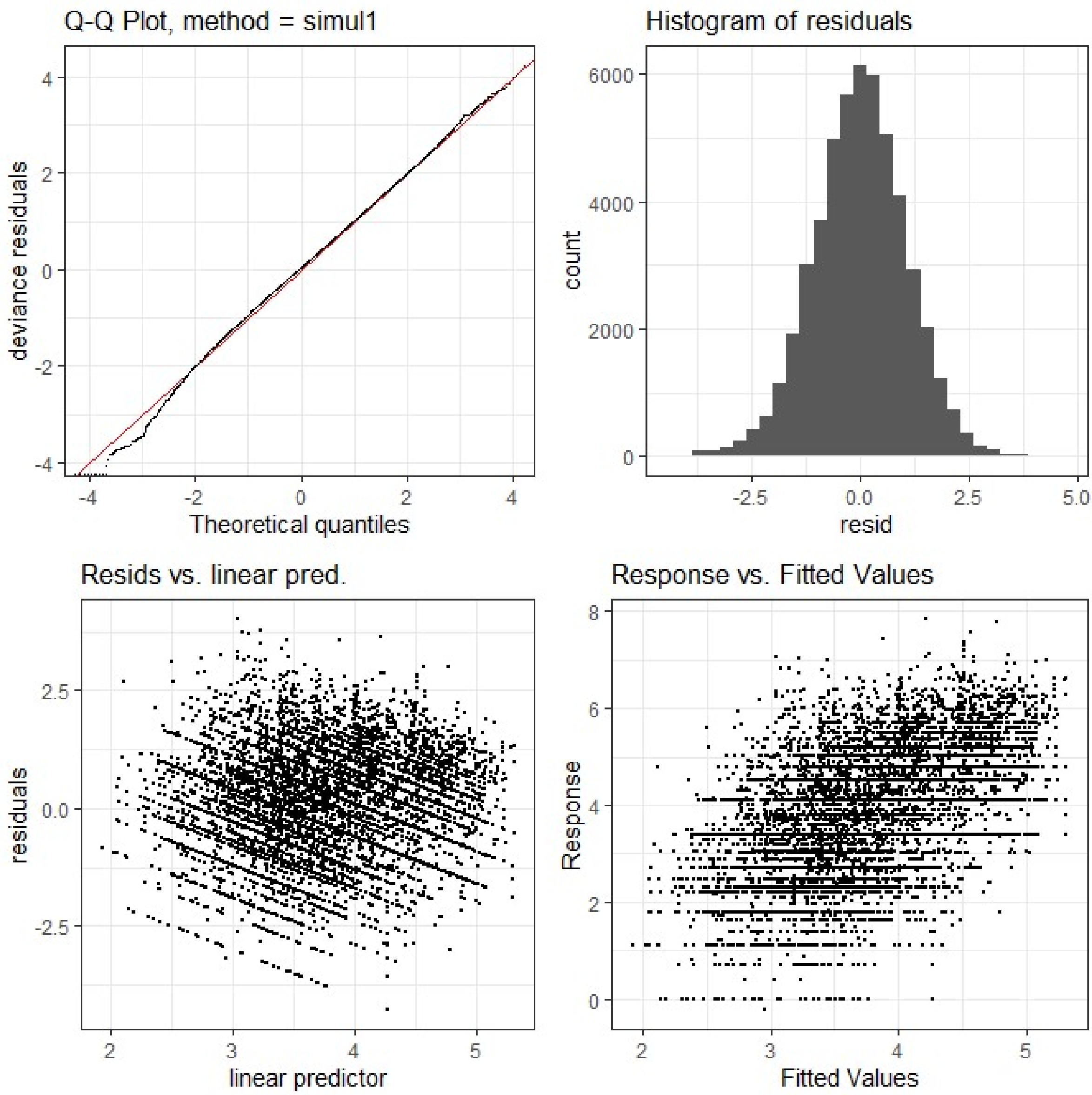

Figure 18: Preliminary standard residual plots generated from a draft lognormal model fitted to non-zero catch rate data from the Karumba region.

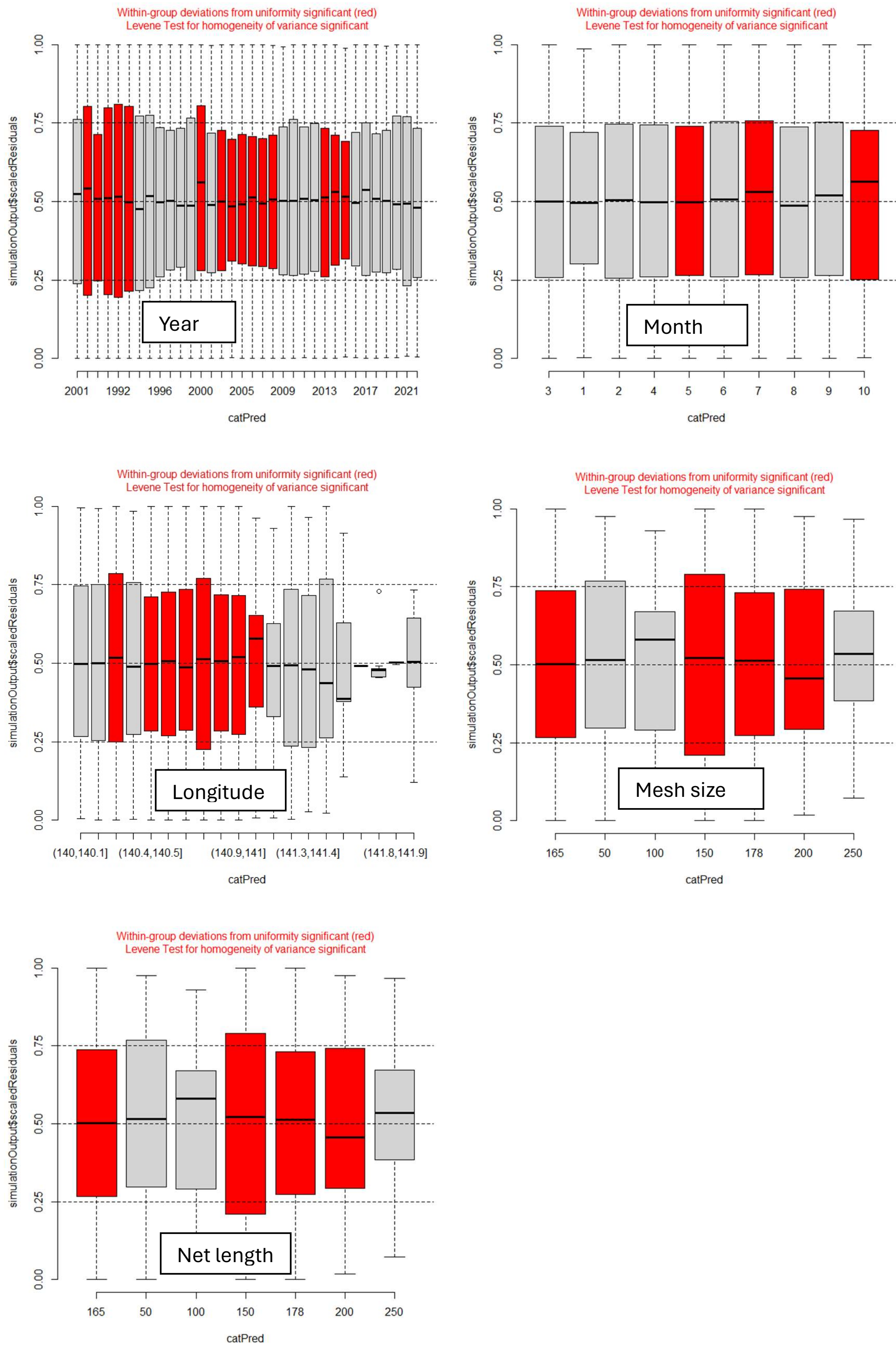

Figure 19: Preliminary DHARMa diagnostic plots for a draft lognormal model fitted to positive catches from the Karumba region. These figures are included to show the types of diagnostics suitable to help identify appropriate models.

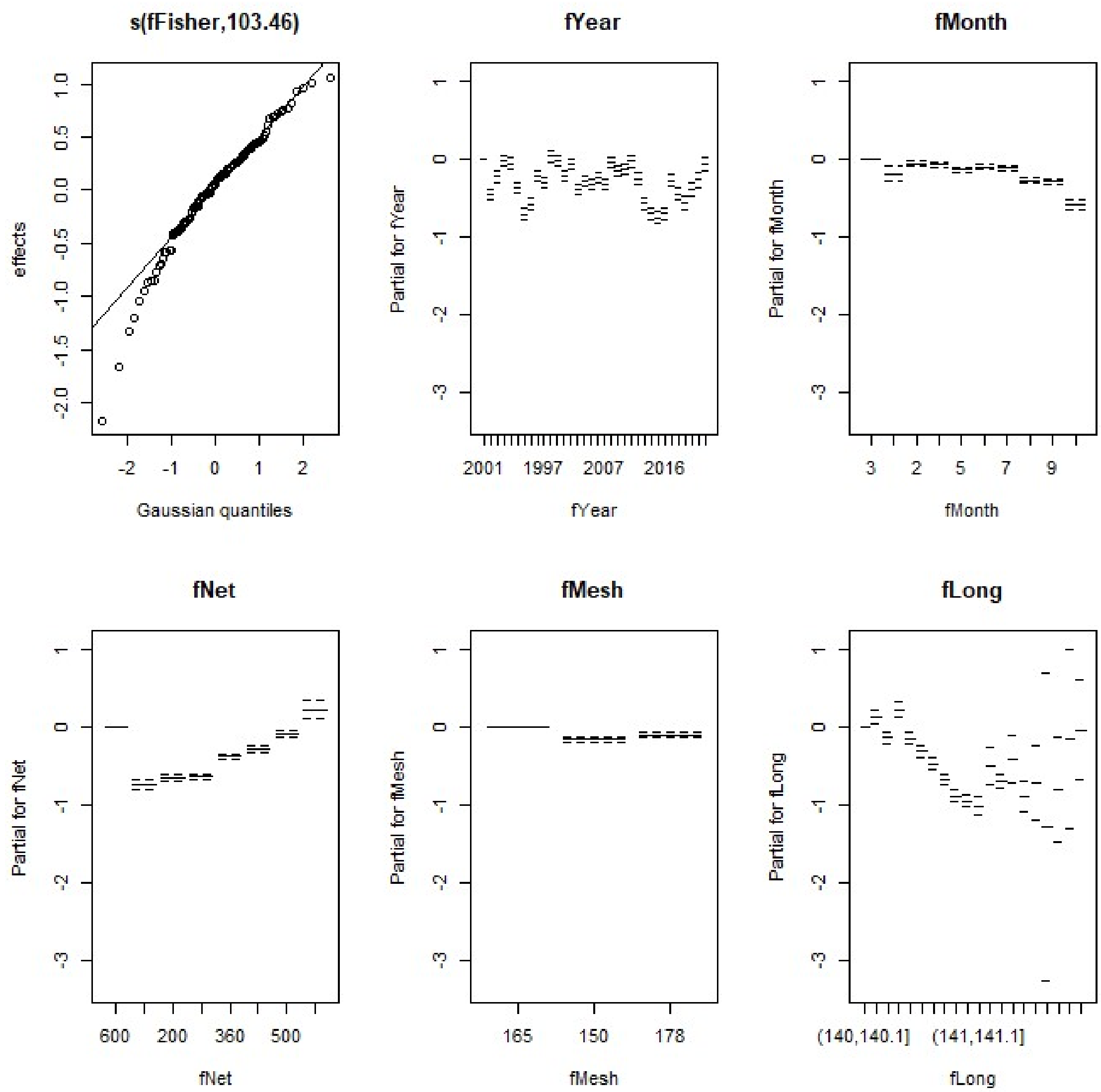

Figure 20: Preliminary effect plot from a draft lognormal model for positive catches from the Karumba region. It shows strong effects associated with year, fisher, net length, longitude and month, and weaker effects associated with mesh size.