REVIEW OF THE 2021 STOCK ASSESSMENT FOR AUSTRALIAN EAST COAST SPANISH MACKEREL (*SCOMBEROMORUS COMMERSON*)

30 June 2023

Simon Hoyle and Alistair Dunn
Hoyle Consulting & Ocean Environmental

Review of the 2021 stock assessment for Australian east coast Spanish mackerel (*Scomberomorus commerson*)

30 June 2023

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 May 2023, the Queensland Seafood Industry Association (QSIA) contracted an independent review of the 2021 stock assessment for Spanish mackerel on the east coast of Australia. It is a difficult stock to assess, with a complex spatial domain, likelihood of hyperstability in the CPUE, and uncertainties about fish 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. We find that the model shows signs of misspecification, with residual trends in the decadal CPUE time series, age structure and length composition data, bias apparent in the estimated growth curve, and instability in model fits and the likelihood profiles. These problems may largely be driven by the lack of recovery in the CPUE index after the large reduction in estimated catch from 2005. These issues should be resolved before the model is used in the development of management advice. We recommend changes to the CPUE standardisation process, including the probability model, the spatial weighting process, and the adjustments for fishing power. We recommend other changes to the model configuration, particularly the approach to steepness, and recommend reducing the reliance on length data given that age data are available.

Citation: Hoyle, S.D.; & Dunn, A. (2023). Review of the 2021 stock assessment for Australian east coast Spanish mackerel (*Scomberomorus commerson*). Hoyle Consulting Ltd., Nelson, New Zealand, 32 p.

Contents

1. Introduction

In May 2023, the Queensland Seafood Industry Association (QSIA) contracted Hoyle Consulting and Ocean Environmental to undertake an independent review of the 2021 stock assessment for Spanish mackerel on the east coast of Australia (Tanimoto et al., 2021b). The project objectives were to review the data and assumptions in the 2021 east coast Spanish mackerel stock assessment. The overall objective and specific objectives of the review are listed in Annex A of this report. The review took place over May-June 2023. As part of the review, the reviewers engaged with industry representatives from QSIA, stock assessment scientists at the Queensland Department of Agriculture and Fisheries (QDAF), 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 east coast Spanish mackerel that has been undertaken by QDAF and all the scientists and fishery experts involved.

The main documents summarising the 2021 stock assessment for Spanish mackerel on the east coast of Australia includes the 2021 assessment (Tanimoto et al., 2021b), the 2018 stock assessment (O'Neill et al., 2018), and the independent review of the 2021 assessment (Klaer, 2021). Other literature and documents reviewed are listed in the references below.

2. Review of the stock assessment

2.1 Stock structure of Spanish Mackerel

Studies of Spanish mackerel stock structure around Australia have identified three genetic stocks: northern/western, Torres Strait, and east coast. At a finer scale, studies within the northern/western stock using parasites and otolith microchemistry identified a metapopulation structure with movements on the scale of 100-300 km (Buckworth et al., 2007). A review of northern Australian populations (Buckworth et al., 2007) stated: "The northern/western Australia stock at least consists of a mosaic of small assemblages (substocks or demes) that, during the adult phase at least, show fine-scale movements and little mixing. The extent to which these are self-replenishing is not known, nor is the mechanism or rate at which such stock units would recover from local depletion. This is in contrast to a previous understanding for this stock of *Scomberomorus commerson*, and shows that an apparently strongly-mixing, large pelagic species may exhibit contrasting fine scale dynamics. It is important to match the scales of management questions and activities and knowledge, and to be aware that cross-scale effects may occur. It may be a critical error to assume that lack of detected genetic differences indicates no spatial dynamic structuring or fine scale implications of management actions".

The report also states: "The project indicated the need to develop management approaches that are resilient to fine scale complexity. This implies that fisheries for Spanish mackerel, as well as fisheries for other species, for which definitive stock structure analysis has not been undertaken, may also need to accommodate strategies that are resilient to a variety of stock structures" (Buckworth et al., 2007, non-technical summary).

Genetic studies indicate that the Queensland east coast stock of Spanish mackerel is separate from populations in the Torres Strait and further west (Buckworth et al., 2007), but investigation of structure within the east coast stock have been limited. Relevant work includes tagging in the 1980's (McPherson, 2007), comparison of parasite loads in Townsville, Rockhampton, Mooloolaba, and Brisbane (Williams and Lester, 2006), comparisons of year class strengths among areas from Townsville south (Welch et al., 2014), and investigation of historic exploitation patterns together with acoustic tagging to track movements (Tobin et al., 2014).

Tagging studies during spawning seasons at reef locations off Ingham, Cairns and Lizard island in the 1980s (McPherson, 2007) found evidence consistent with both large-scale seasonal movements and fine-scale population structure. Of recoveries from the 1976-84 tagging programs, 57% (47 of 82) were within 50 nautical miles (93 km) of the release site. More recent tagging across multiple locations and years reported by the NSW Game Fish Tagging Program (east coast data extracted from annual reports of the [DPI Game Fish Tagging Program (nsw.gov.au)]) showed similar results for movement, with 20 of 28 recoveries recovered within 100 km of their release location, and median displacement of 31 km. Both tagging datasets suggested seasonal movements (e.g., see Figure 2.7 of McPherson, 2007), but displacements after a long period were similar to those after shorter times at liberty. In the NSW Game Fish Tagging Program, the median displacement of the 13 fish recovered after at least one year at liberty was 37 km. These small displacements are inconsistent with a randomly mixing population. Long movements were also observed, with one fish recaptured 1921 km from the site of release after 566 days at liberty.

Parasite studies (Williams and Lester, 2006) showed similar parasite assemblages in adult Spanish mackerel samples from Townsville, Rockhampton, and southeast Queensland (Mooloolaba and Brisbane). These studies have suggested that the south-east Queensland fish appeared to be a random selection each year from the Townsville stock, although we note that this type of random movement would be inconsistent with evidence of mainly small displacements for recovered fish that were tagged south of Townsville.

Spatial variation in year class strength may be informative about mixing before recruitment to the fishery. Relative year class strengths were very similar among 4 areas (Townsville, Mackay, Rockhampton, and South (south of about 24.5° S)) and were negatively associated with sea surface temperature (SST) (Welch et al., 2014), suggesting that fish in each location are mostly from a single spawning area, or are a random mixture of fish from largely the same spawning areas, or that environmental drivers result in similar year class strengths from all spawning areas. In any case we note that this result is consistent with genetic evidence that fish from Townsville south are part of the same breeding population.

Tobin et al. (2014) identified aggregating/concentrating behaviour during the spawning season, and strong reef fidelity, with "a possible strong homing reef ability and behaviour". They suggested that "Such a defined aggregating behaviour suggests that spatial closures are likely to be effective at protecting some Spanish mackerel during spawning".

Given the results of the 1980s tagging (McPherson, 2007) and the acoustic tagging (Tobin et al., 2014), limited mixing of adults among areas from Townsville north appears likely. This would be consistent with the behaviour of Spanish mackerel in the rest of northern Australia – a single genetic stock exhibiting fine-scale dynamics, such as local residency (Buckworth et al., 2007). Parasite and otolith microchemistry work in these areas would be needed to explore these questions. These fine-scale dynamics may include aspects such as differences by age and sex. It is also clear that the feeding range expands to the south and involves long-range movements, particularly after the October-November spawning season. It is unclear what proportion of fish from each northern latitude participate in this expansion, and whether individuals move at random or in a more structured way, but partial migration is very common across fish species (Chapman et al., 2012).

South of Townsville, stock structure and fish behaviour are uncertain. It is likely that a species with local residency behaviour elsewhere would retain some aspects of this behaviour within its seasonally expanding distribution. Null hypotheses about fish behaviour should be based on known behaviours elsewhere by the same or related species. Anecdotal evidence suggesting a degree of local residency in southern areas is reported by McPherson (2007). NSW Game Fish Tagging program data south of Townsville are both consistent with seasonal movements and suggestive of some type of local residency. There is an indication of either local residency or return to the same site after spawning, given observations of long-term recaptures close to the release site. Of the 19 fish released south of 21° S and subsequently recaptured, the median displacement was very low at 28 km. Of these, the southernmost release was at Bermagui (36° S), and this fish was recaptured 31 km from the release site after 1037 days at liberty.

2.2 Stock structure in the assessment

The stock assessment assumes that the Queensland east coast Spanish mackerel population 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 like a pool of water – fishing in one location affects every part of the stock, reducing density and changing the age structure in all locations.

A stock with local dynamics, however, will behave quite differently. 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-length relationship, 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 completely 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. Unless mixing is complete, the raw aggregated age structure and size structure data will indicate higher fishing pressure than the true average across the stock. 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 only from fished areas is, *a priori*, likely to bias assessment results towards more pessimistic outcomes. The degree of bias is related to the rate of mixing between areas. We recommend research to better understand stock movement and mixing dynamics.

2.3 CPUE

Indices of abundance are usually the most important and influential component of a stock assessment. For Spanish mackerel, the indices of abundance available were (i) fishery dependent catch-per-unit-effort (CPUE) indices (O'Neill et al., 2018), and (ii) reconstructed historical decadal catch rates (Thurstan et al., 2016b).

A description of the methods used by Tanimoto et al. (2021b) to estimate the CPUE indices can be found in O'Neill et al. (2018). We also note that the description of the CPUE methods within the 2021 assessment document was incomplete and agree with the recommendation of the independent reviewer (Klaer, 2021) that it would be useful to have a separate document that provides a more comprehensive overview.

The overall CPUE approach was related to the delta lognormal method (Lo et al., 1992), in which a binomial model is used to allow for variation in the proportions of zero and nonzero catches, and a CPUE model is fitted to the nonzero catches. However, the approach took a different and less conventional approach to the binomial model, with the objective of adjusting for changes through time in the availability of schools of Spanish mackerel. This is on the basis that catch rates are hyperstable when fishing on schools, so the binomial model index of school availability was needed.

The probability model is not fitted to individual logbook data, because some zero catches are not identifiable in those data. Instead, logbooks are grouped by year, month, and 1° latitude band (latband), with each group used as a row of the dataset. The number of calendar days on which Spanish mackerel were caught by any ACN is standardized as a function of year, season, the number of active Spanish mackerel operations in the group (nACN), and the average wind strength and direction for that group. The idea is that higher fish abundance, or a higher number of fish schools in a group, should correlate with more calendar days on which any ACN catches Spanish mackerel. An ACN is defined as active in a group when it catches at least one Spanish mackerel in that month and latband.

There seems to be potential to improve this model. The response variable is the probability of capturing a fish on each calendar day by any of the ACNs. This is a very aggregated response variable without much potential to adjust for changes in the fleet and other factors. The number of catch days may be less affected by abundance than by other factors, such as whether each day has weather suitable for fishing (currently modelled as a monthly average wind strength from 2 directions), the area available for fishing, and changes in the behaviour of the ACNs.

The probability of nonzero catch will be substantially affected by the behaviour and license of the fisher, who may specialize in Spanish mackerel or allocate some effort to reef fishing. Specialists are more likely to catch Spanish mackerel at a rate that reflects abundance. Non-specialists may catch Spanish mackerel by trolling while traveling to the fishing ground, by chance while fishing for other species, or by intermittently targeting. Changes through time in the proportion of specialists may change the index inappropriately unless this is adjusted for. Target change is a very common and important source of bias affecting CPUE indices. This could be addressed in various ways. The simplest approach would be to include only vessels that exclusively target Spanish mackerel. This approach is commonly applied in mixed fisheries and would at minimum be a useful check. Hyperstability can be associated with targeting aggregations, and this will need to be considered. Data from non-targeting vessels can also be analysed independently, though these data will need to be thoroughly characterised and groomed given the spatial, seasonal, and interannual variation in fishing strategies in this mixed fleet. Including mixed target strategies in a CPUE analysis without accounting for catchability differences is not recommended.

Capture by an ACN of a single fish in a month effectively adds effort for that month (increases nACN), and only adds to the catch variable (number of days with catch) if that catch occurs on a day when others catch nothing. Since Spanish mackerel distributions are relatively predictable, specialists can reliably catch Spanish mackerel when they are available and the weather is suitable. Additional catches by non-specialists with low catch rates may therefore perversely affect the probability index, since increasing abundance may generate more random catches, increasing nACN but not the number of catch days, and thereby reducing the probability index.

It may therefore be best to exclude ACNs that do not consistently catch Spanish mackerel in a latband. Similarly, reducing the fishable area within each latband (due to closure of green zones) may increase the rate of fishing in multiple latbands. Reporting the same ACN in multiple latbands will tend to reduce the index by increasing the nACN on average. In addition, if fleet mobility increases through time due to more powerful engines (Thurstan et al., 2016b), ACNs may progressively become more likely to fish in multiple latbands. This potential bias can be avoided if, when an ACN fishes in 2 latbands, the effort is shared so that each nACN increases by a fraction, not by 1 in both latbands.

The probability model shows a step change down in many locations in about 2005 (see orange line in Figure 1), after the closure of RAP areas and the license buyback. We would expect these management changes to generate more zeroes, given fewer areas to fish and a change in behaviour by the fleet. The change in 2005 looks like a management effect rather than a biomass change but is interpreted as a biomass change. The ACN parameter in the model reduces the decline in 2005 compared to the model without covariates (blue line in Figure 1), but the probability still appears to decline by about 15% in 2005, more than in any other year, which strongly suggests that some of the management change effect remains in the probability index. Apart from the step down in 2005 there is very little change in the index through time, and it seems likely that the probability index is not effectively indexing the availability of Spanish mackerel schools.

Overall, this method seems to have several potential problems. We do not see evidence that it is successfully adjusting for hyperstability, as claimed, and caution against its use unless simulation analysis shows it to be reliable. Simulation is a valuable tool in CPUE standardization, particularly with novel approaches like the one being applied here.

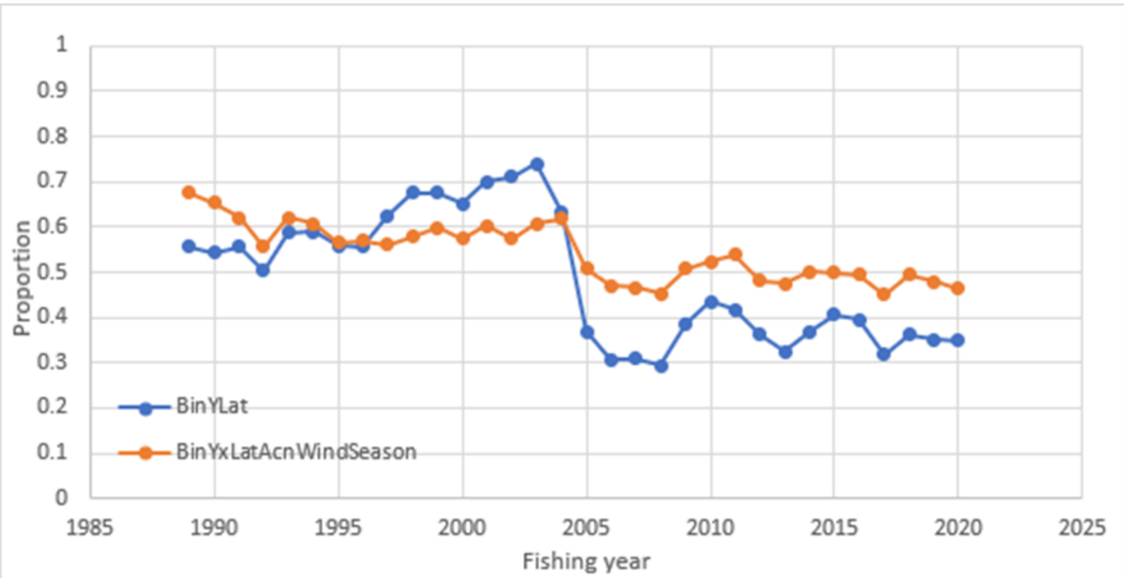

Figure 1: Probability of commercially harvesting Spanish mackerel by fishing year, based on a simple model with year and latitude (BinYLat) and the final model with year, latitude, ACN, wind, and season (BinYLatAcnWindSeason) (QDAF, personal communication).

The model for probability of nonzero catch could be further discretized in stages to improve the use of the information in the data, via the following suggestions. However, we do not know if these changes will result in a usable index.

- Weather is an important driver of fishing effort, and conditions change through time with factors like La Nina and climate change. Better use could be made of weather info and its effect on catch by running the model daily, which should be straightforward while still using monthly groups to define the nACN variable. There may be more relevant covariates for fishable weather than wind strength from 2 directions, which should be available as daily records.
- Individual ACNs have different probability of catching Spanish mackerel. As suggested by QDAF, the nACN value could be adjusted based on the catchability parameter estimated for each ACN in the catch rate model, though this requires choosing how to convert between the two. Alternatively, the dataset could be discretized to the ACN level and a catchability parameter estimated for each ACN. This changes the model significantly and makes better use of information by removing the aggregation of catch across all ACNs, and should increase its responsiveness to abundance change. There would be many zero catches, which are straightforward for a binomial model. There would be incomplete factor combinations at the ACN-month-latband level, which is very common in this type of model. It is unlikely to cause any significant problems and simulations can be used to check that the estimates are robust.

2.3.1 Catch rate

The model of catch rate is implemented as a linear mixed model in GenStat. The data set is large and the model estimates many parameters, with a separate fixed effect for each latitude and year combination, and independent terms for seasonal effects at each latitude, and 2 terms at each latitude for sinusoidal lunar effects. Weather effects are implemented using quadratic polynomials on wind strength components NS and EW, assuming the same relationship for all locations. There is a huge amount of data, and the parameters are estimable. However, not all the data is necessarily informative – particularly given the lack of information in logbooks about fishing methods and targeting. The catch rate analysis fits a single model across the whole spatial domain with the same error distribution assumptions. With so much data available, there are opportunities to examine relationships at finer scale, and focus in on the more informative parts of the dataset.

For ACNs that target other species, zero catches for Spanish mackerel should be identifiable in daily logbooks on days when they catch other species but not Spanish mackerel. These records contain information about the abundance of Spanish mackerel; including them in the index may be better than including only positive catches.

It is hard to recommend specific methods for the CPUE analysis without full details of the catch and effort dataset, which would be provided by the process of characterisation. A documented characterisation of the dataset would give analysts and stakeholders a better basis for identifying the appropriate fleet components. Information is also needed on changes in spatial and temporal distributions, and issues such as catch composition and gear use.

In a mixed-fleet fishery with shared licensing, changes in targeting behaviour through time can substantially affect the CPUE. This may be particularly important before and after the management changes that occurred in 2005. Approaches such as cluster analysis on species composition and exploration of fishing behaviour (He et al., 1997; Hoyle et al., 2022; Parsa et al., 2020) are useful for identifying patterns of targeting. Based on multiple characteristics including the catch composition, season, and latitude, it may be possible to classify ACNs into groups such as those that consistently target Spanish mackerel via trolling, those that fish for reef fish while anchored and occasionally catch Spanish mackerel, and those with a mix of behaviours.

In addition, we recommend running and evaluating models that standardize the CPUE of subsets of vessels with consistent behaviour over the time series (including before and after the management changes in 2005), including models that only use vessels that have exclusively targeted Spanish mackerel.

We suggest considering fitting these models as GAMMs in the R package mgcv (Wood and Wood, 2020), which allow the implementation of cyclic splines for month and lunar cycles, and also smooth across space (latitudes) and season. Polynomial approaches can behave in problematic ways away from the majority of the data, driven by the functional form. Splines are more flexible and tend to be better behaved. Spatial smoothing would avoid having to estimate multiple independent categorical variables for spatial effects, and would take advantage of similarities between adjacent latitudes. GAMs and GAMMs are straightforward to implement for experienced analysts, with a wide array of diagnostic tools that can be used to identify modelling problems.

As indicated by Walters (2003), and also in follow-up papers (Carruthers et al., 2010; McKechnie et al., 2013), when filling spatial gaps (i.e., locations without fishing effort) it is important to consider why there's a gap and to apply appropriate methods on a case-by-case basis. We note that approximately 35% of the reef was closed to fishing in 2005, which would represent a significant increase in gaps over a short period, depending on the proportion of Spanish mackerel habitat within these areas. Uncertainty will need to be resolved about the proportion of habitat this represents.

The probability model and the catch rate model have no data from closed areas after 2005, so the current approach assumes that abundance trends in closed areas are the same as in open areas. Areas that are unfished due to spatial closures could act as reservoirs of higher abundance both within a fishing season (Tobin et al., 2014) and potentially longer term, depending on the pattern and time scale of population mixing. If this is the case, their abundance would be expected to trend upwards compared to fished areas. Similar effects will occur in areas with lower fishing pressure. Since full and rapid mixing is unlikely, these types of effect are likely, though the degree of mixing would be important in choosing an appropriate method for addressing this.

2.3.2 Hyperstability

Hyperstability is a concern since it can change the relationship between catch rates and abundance, but we note that hyperstability will affect CPUE during both population decline and recovery. If catch rates were perfectly hyperstable there would be no relationship between CPUE and abundance, and the current approach would always provide a declining index whether the population was declining, stable, or recovering, because of the probability and fishing power adjustments.

If the spawning fishery in the Lucinda latband is the most concentrated period and therefore could be most subject to hyperstability, a model that excluded this area during the spawning period should be considered as it may provide a better index of abundance.

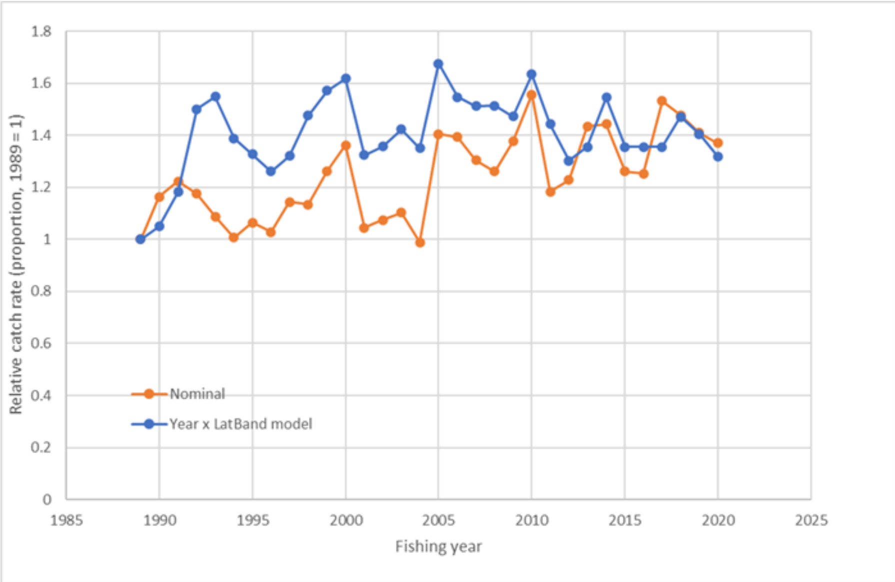

Figure 2: Nominal (orange) and standardized (blue) indices. QDAF, personal communication.

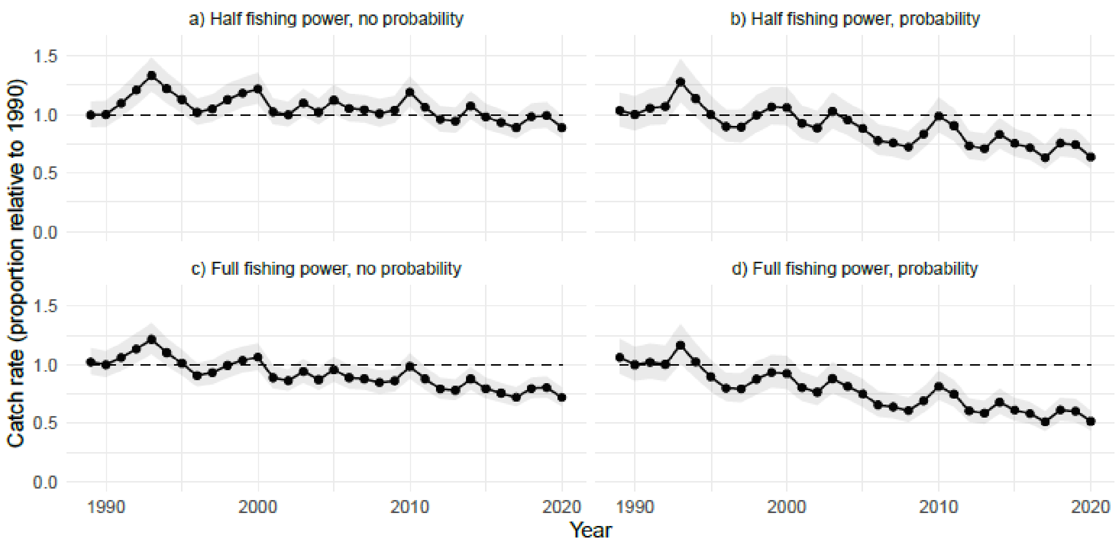

Figure 3: (Figure 3.5 from Tanimoto et al 2021): Annual standardised catch rates (95% confidence intervals) for Queensland commercial line-caught Spanish mackerel between the years of 1989 and 2020, for four scenarios.

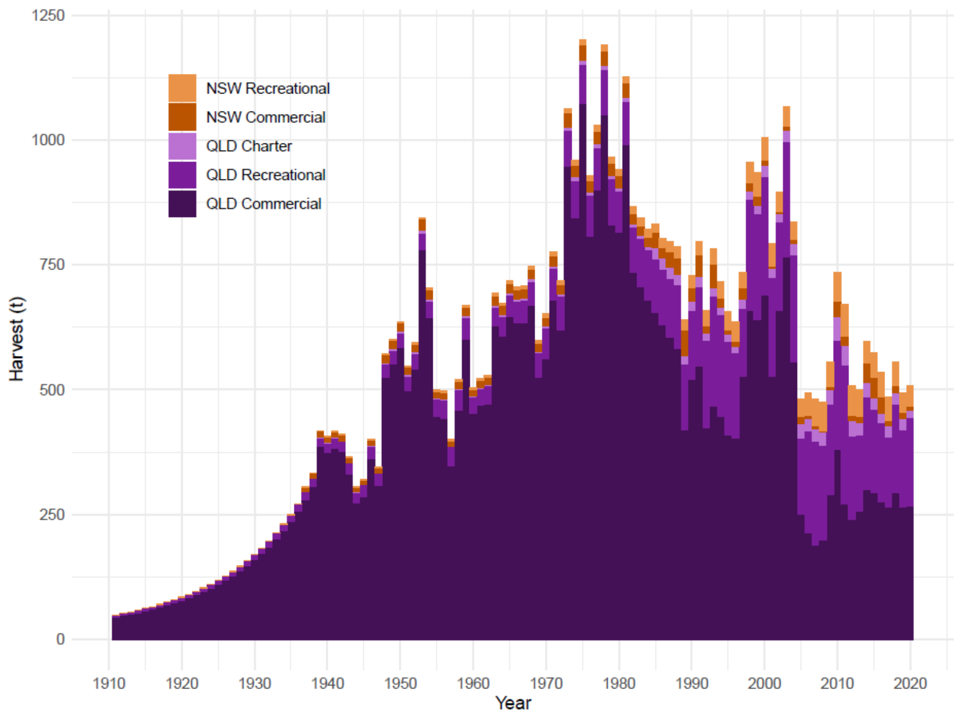

Figure 4: Annual estimated harvest from commercial, recreational and charter sectors between 1911 and 2020 for Spanish mackerel. (Figure 1 from Tanimoto et al 2021).

2.3.3 Fishing power

The fishing power estimates are based on fisher interviews (Buckley, 2016; Buckley et al., 2017; Thurstan et al., 2016b), also summarised by O'Neill et al. (2018). The PhD thesis on which the analyses in the paper and report are based (Buckley, 2016) is unfortunately not publicly available. Buckley et al. (2017) report that estimates of the proportional increase in fishing efficiency as a result of adopting each new fleet characteristic were obtained from interviews with 10 fishers. O'Neill et al. (2018) report that estimates of technology effects were obtained from raw data provided by Buckley for interviews with 41 commercial and 23 recreational fishers (at least 25 commercial fishers, and between 7 and 21 recreational fishers for 3 technologies, and 11 commercial and 4 recreational for the other (QDAF, personal communication)). For the current analysis, data on technology effects were averaged across all fishers, both commercial and recreational (O'Neill et al., 2018). Estimates of technology effects on catch rates by both O'Neill et al. (2018) and Buckley et al (2017) did not differentiate between trolling and anchored line fishing. Buckley et al. (2017) report selection of 10 fishers who had been operating prior to adopting each new technology. It is unclear whether the fishers used by O'Neill et al. (2018) had similar levels of experience, and if there were other differences between the samples.

Appendix S2 of Buckley et al (2017) gives estimated fishing power effects for commercial fishers due to GPS, colour sounders and paravanes of 30.63%, 16.56%, and 21.1%, which are smaller in aggregate than the equivalent estimates in the 2018 report (which were used in the stock assessment) of 42%, 27%, and 21%. O'Neill et al (2018) also included live bait (26%), which was not reported by Buckley et al (2017). In the data used in the 2018 report, gear effects estimated by commercial fishers were higher than by recreational fishers for GPS, paravanes, and live bait, and similar for colour sounders.

There is concern from industry about the relevance and representativeness of these fishing power estimates based on both commercial and recreational fishers for the commercial fishing-based index. They have considerable influence on the outcomes of the stock assessment. There is a view that the Spanish mackerel fishing power benefits of GPS, colour plotters, and paravanes are greatly overestimated for experienced commercial fishermen targeting Spanish mackerel via trolling with lures. These fishermen are said to usually locate Spanish mackerel in well-known and predictable locations during the peak spawning season without GPS and colour sounders, and have used alternative paravanes / line sinking devices such as streamlined lead weights since the 1930s. Live trolling is uncommon for commercial fishers except in the south (Industry, personal communication). GPS and colour sounders are said to be more relevant to anchored fishing, which should not be mixed with troll fishing. They are also considered likely to help recreational fishermen more than commercial fishermen, given differences in experience and local knowledge. We recommend a review of these estimates and consideration of their relevance to the fishing methods included in the CPUE index before they are used in the base case assessment model. For example, it may be necessary to consider fishing methods and their spatial and seasonal variation.

If GPS and colour plotter technologies do provide benefits in some circumstances, it would be to help fishers locate schools when they are present. They may therefore be more appropriately applied in the probability model as an adjustment to the nACN, rather than in the catch rate model as increases in individual catch per day, as they currently are.

The makeup of the fishing fleet and its effect on fishing power may also be worth considering, since Spanish mackerel fishing is physically demanding and older fishermen may change their behaviour, reducing the effectiveness of a day's effort. A survey of economic and social indicators in the Queensland East Coast Spanish mackerel fishery 4-6 years ago (2017/18 and 2018/19) found that, at that time, 53% (23/43) were aged 56 or more (Magnusson et al., 2020).

We also note that industry have expressed concern about lack of transparency regarding these fishing power estimates. In response to our questions, QDAF provided us with some additional background information. It would help understanding and support a better-informed discussion if these and other relevant details were available more widely.

2.3.4 Creating the combined index

In the description of modelling methods (O'Neill et al., 2018), the description of how the binomial and catch rate indices were combined was missing some details. We recommend that a detailed description of the method be included in the documentation of the CPUE. We note that it would be appropriate to (i) predict values from each model at the latband by year stratum level, (ii) multiply the probability of nonzero catch by the predicted catch rate for each stratum (without first normalizing the catch rates), (iii) multiply (weight) each stratum by habitat area, because catch rate is an index of density, (iv) sum strata across latbands to obtain annual indices, (v) normalize by the mean across years to obtain an index. For a thorough description of these processes see Campbell (2015).

Note that when predicting the binomial component, it is helpful to keep the means of the annual predictions for each latitude close to the means of the equivalent observations. Over- or under-scaling probability estimates can bias the trend by changing the degree of saturation against the upper boundary (Hoyle et al., 2022).

The methods description (O'Neill et al., 2018) reports that each latitude's prediction was first normalised as a proportion of the rate for the same latitude in 1990, and then weighted by (multiplied by the latitude's proportion of) total harvest over years 1989-2016. (Weighting by catch was also applied to the decadal mean catch rates). QDAF (personal communication) reported that the 2021 assessment did not weight by total harvest but weighted by the number of data, which is similar to weighting by fishing effort.

Weighting normalised estimates either by total catch or by the number of records is not usually recommended. CPUE is an index of abundance, but these weighting methods ignore the spatial abundance information in the CPUE and assume that an area's abundance is proportional to either the average annual catch or fishing effort. They assume there is less abundance in latbands that are more remote or have larger closed areas and are therefore fished less. They give the most weight (39% in the 2018 assessment) to the latitude 19 band with the most effort and catch, which assumes that this area has 39% of the vulnerable population on average. Unless the population is very well mixed, abundance will tend to be more depressed in areas that are more heavily fished, so this approach will tend to bias the index towards more pessimistic results.

Instead, the objective should be to index the abundance of the whole stock rather than giving more weight to the fished areas. CPUE should be proportional to local density, which is the reason for using it in the first place, so it is technically more appropriate to weight latband estimates by relative area, or (better) by the relative area of Spanish mackerel habitat (Campbell, 2015). An appropriate method will need to be identified for determining relative habitat areas.

The nominal index shows an increasing trend through time, while the standardized index before adjustments is relatively flat since 1992 (Figure 2). The declining trends of the indices used in the assessment (Figure 3) are driven by the combined adjustments due to the probability model and the fishing power estimates. Running model scenarios both with and without the probability model goes part way to addressing the concern about this approach, but we recommend dropping the runs that include the probability model until the issues with it have been addressed.

2.3.5 Decadal CPUE

The decadal CPUE used in the assessment is based on estimates reported by Thurstan et al. (2016a) and Thurstan et al. (2016b). Based on data from multiple species, Thurstan et al. (2016a) estimated a trend in recall bias of 0.65% per year. "Compared to mean catch values, fishers' recall became more exaggerated as time passed, increasing at a rate of 0.65% per year elapsed between the event and timing of recall". Buckley et al. (2017) applied similar methods to their much smaller Spanish mackerel catch dataset, and not surprisingly found no significant trend in recall, given the much lower statistical power with fewer individuals. Since Buckley et al. (2017) did not find a trend, O'Neill et al (2018) assumed no trend in recall bias, but in this situation, it is more appropriate to use the estimate from the larger dataset than to assume no trend, particularly since the existence of temporal trends in recall bias is well established (Lawson, 2015). The 0.65% per annum trend in the rate of recall should be used to adjust the decadal CPUE index.

We note that the assessment model is unable to fit the unadjusted trend in the decadal CPUE index, which indicates less decline in abundance than predicted by the assessment model (Figure 5). In most of the model run scenarios there is a trend from negative to positive residuals (Figure 6), which suggests that the index underestimates the rate of decline, and/or that the productivity or scale of the stock is being underestimated by the model.

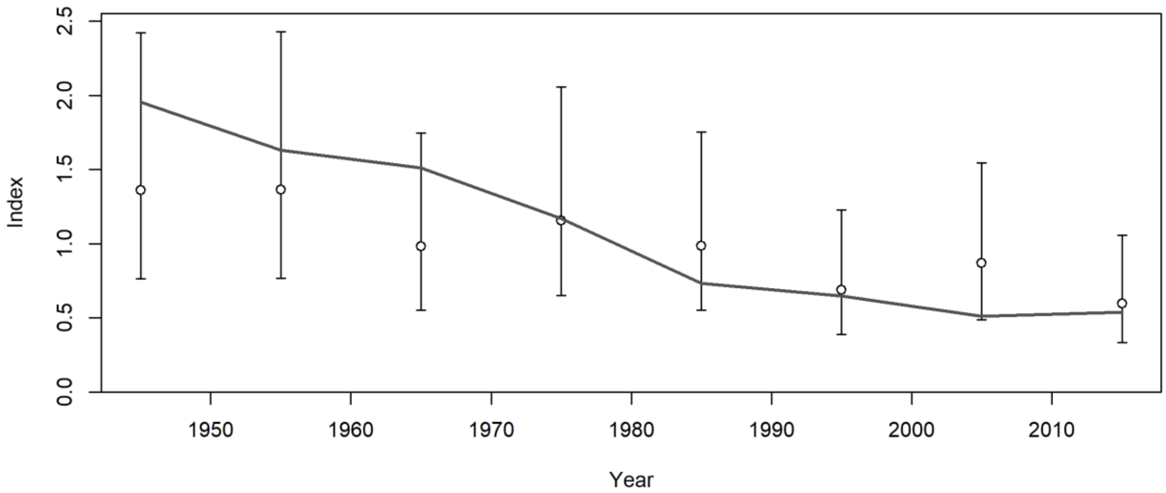

Figure 5: Model predictions (grey line) to historical decadal catch rates for Spanish mackerel for the base case scenario (Figure B.2 from Tanimoto et al., 2021b).

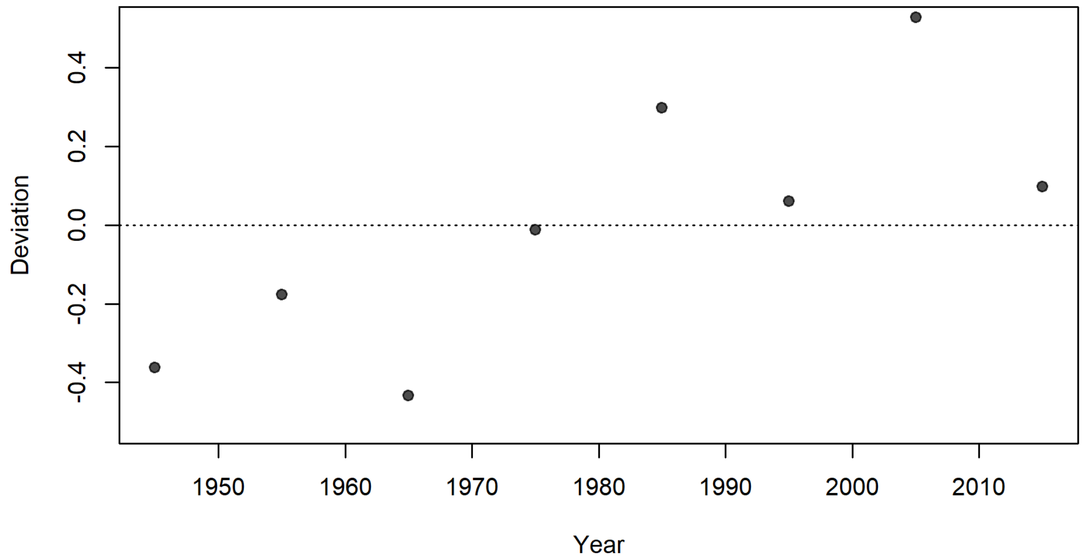

Figure 6: Time series of residuals from base case fits to the historical decadal time series.

2.4 Catch estimates

Recreational catch estimates from RFish have been reduced to 0.34 for harvested and 0.26 for released Spanish mackerel (Tanimoto et al., 2021b), following Leigh et al. (2017). The assumption in this scaling was that the RFish estimates were overstated by the same fraction in all survey years in which the RFish methodology was employed. The adjustments are based on both the existence of recall bias (Lawson, 2015) and a desire to align RFish estimates with the results of NRIFS surveys. However, these are very significant adjustments for an iconic species with a bag limit of 10 (before 2003) or 3 (since 2003), which should be recalled more accurately than most species. It also represents an adjustment factor more substantial than estimated by Lawson (2015) for any species – most required adjustments factors over 0.5, where 1.0 represents no adjustment. We note that Lawson suggests "the unpredictable and multifaceted nature of the reporting bias observed in the RFish data may preclude any reliable adjustments". We recommend a review of the approaches used to develop the recreational catch time series. We also recommend that analysts explore the sensitivity of assessment outcomes to alternative assumptions about recreational catches.

Shark depredation can substantially affect catch estimates. Rates that trend upwards from 0% in 2009 to 20% in 2020 were explored as a scenario in the assessment. The peak estimate of 20% was based on the highest depredation rate reported by Mitchell et al. (2018a), which represented fishermen's estimates in the Australian longline tuna and billfish fishery (Gilman et al., 2008). However, depredation rates are likely to vary substantially between fisheries, locations, fishing methods, and fishing sectors, and local estimates are needed, particularly for recreational fisheries which have been relatively little studied (Mitchell et al., 2018a). Depredation may be more prevalent in the recreational sector given the financial incentives to avoid depredation in the commercial sector. Depredation can be more prevalent in areas with higher fishing effort, which suggests the development of a behavioural association (Mitchell et al., 2018b), i.e., sharks may learn to target fishermen. Research is currently under way by QDAF (Dr Jonathan Mitchell) to estimate depredation rates in the commercial and recreational sectors of the Spanish mackerel fishery. When preliminary estimates of depredation rates become available, we recommend testing their effects in the assessment.

2.5 Biological parameters

2.5.1 Recruitment

The assumed σ_R of 0.35 is lower than used in most stock assessment for fish species. This value is based on the variability of recruitments estimated in the previous assessment (O'Neill et al., 2018). However, the variability of recruitment estimates from an assessment is biased to be lower than the true variability among recruitments (Methot and Taylor, 2011), so this assumed value is likely to be too low. Age data usually provides good information about recruitment strength, but the signal is always affected by model misspecification. The assumed σ_R is often negatively correlated with B0 and R0, and 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 0.35.

2.5.2 Natural mortality

The instantaneous rate of natural mortality (M) in the final models was estimated with and without the prior from Then et al. (2015). We recommend updating the prior mean to M=5.4/Amax, based on Hamel and Cope (2022), who identified a modelling error in the earlier approach.

M is hard to estimate from fishery data (Maunder et al., 2023) and misspecified models (which all models are to some extent) will tend to produce biased estimates of M. 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, and evidence of poor fit to age structure data, we recommend fixing M in the model 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).

We also recommend replacing constant M at age with the biologically well-justified Lorenzen 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.

2.5.3 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. Basing stock assessment steepness primarily on an estimate from FishLife (Thorson, 2020) is not recommended.

The FishLife database predicts median steepness of 0.45 for *S. commerson*, with a (left-shifted due to the log scale) distribution truncated at 0.2 and extending to 0.8. We identified 5 estimates for *S. commerson* in the FishLife database: 0.400, 0.461, 0.689, 0.736, 0.733, which must be based on 5 SB & R series in the RAM legacy database, fitted using a model described by Thorson (2020) (further details are difficult to find). The estimate of 0.45 will be based on these 5 values, along with information from higher level taxa and correlations with other life history parameters. Median for the genus *Scomberomorus* is higher, which suggests that the value of 0.45 is driven by life history correlates.

FishLife has not yet been widely used so its utility is not well understood. The model was developed to estimate multiple life history parameters across a very large number of species, and some parameters and species are better informed than others. The 5 input steepness estimates are derived from assessment results in the RAM legacy database, and the quality of the data sources is unknown. The utility of this method is uncertain for steepness but appears to be low: Szuwalski et al. (2015) found that recruitment and spawning biomass were not positively related over the observed range of stock sizes for 61% of 224 stocks in the RAM Legacy Stock Assessment Database.

FishLife estimates of hard-to-estimate parameters like steepness and natural mortality are unlikely to be reliable, simply because the data sources do not provide much information. Based on the same information sources, FishLife predicts natural mortality for *S. commerson* of 0.25 to 0.7 with median M of 0.45. The M values used in the current assessment are at the very low end of that distribution, around 0.25 to 0.3. These lower values are reasonable given the observed age structure, i.e., there is biological evidence for them, whereas there is little evidence for the steepness estimates.

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).

As noted by Zhou et al 2020, life history correlates for steepness are not well established. Munyandorero (2020), summarized three inconsistent perceptions in the literature about the relationship between steepness and life-history traits. (1) h is higher in short-lived than in long-lived species. (2) h is higher in long-lived species than in short-lived ones (Goodwin et al., 2006; Myers et al., 2002). And (3) there are no relationships among h and M or other LHPs (Shertzer & Conn, 2012; Thorson, 2019). Results in Zhou et al (2020) tended to support Perception 2, but this is contrary to some popular beliefs and practices.

We emailed Jim Thorson to ask about the FishLife steepness, suggesting that we're inclined to recommend giving the FishLife estimate only a small amount of deference, and consider a wide range of options. He agreed this was fair and said his model needs better diagnostics to understand how much results are being driven by related taxa vs life history correlations. Jim said he has previously been contacted about this same assessment and has always recommended exploring a wide range of options. The independent reviewer (Klaer, 2021) made similar points and suggested that using the FishLife median of 0.45 as the base value was inconsistent with accepted practice, which was reinforced by the recommendations by reviewers of other assessments that he noted. SEDAR review panels for other Scomberomorus assessments in the US (2014 Gulf of Mexico king mackerel and 2014 South Atlantic king mackerel) recommended the use of high steepness values, despite the estimation of lower values in those assessments.

The authors (Tanimoto et al., 2021a) disagreed with the advice of the independent reviewer on this issue. They argued that, because models with higher steepness were unstable and had poor fit to the data, those steepness estimates should not be used. We think this is equivalent to assuming that the assessment contains reliable information about steepness, which is not our view. Model misspecification is a much more likely explanation for the modeling problems with high steepness. We observed model instability when making reasonable changes to model parameters, and the model also appeared unstable in the analysts' likelihood profiles on model scale.

We recommend that the analysts should apply a wide range of values for steepness from 0.45 to 0.95. This represents the substantial uncertainty about steepness, and results in a median value close to the 0.69 estimate for the Scombridae Family.

It has been argued that steepness is likely to be low because Australian waters are relatively unproductive. Productivity will affect the carrying capacity (Bzero), but we are not aware of evidence for carrying capacity affecting steepness. Changing the assumed steepness can change the estimated carrying capacity in some modelling situations, but that does not imply a relationship between system productivity and steepness.

As Mangel et al. (2013) points out, fixing both steepness and natural mortality largely determines the productivity of a stock. Since it is unrealistic to reliably estimate either parameter inside this assessment, we recommend addressing the uncertainty in these parameters by using an ensemble approach.

2.6 Age and length

Composition data are included in the model as length compositions, combined with observations of conditional age at length data from otolith readings.

The age at length composition data shows a trend of bias through time in the fit (Figure 7). In early years the observations are younger than predicted, while in later years the observations are older than predicted. The same pattern is apparent for both males and females (see Figures B.4 and B.5 of Tanimoto et al., 2021b). Note that each 95% confidence interval represents the uncertainty distribution for a single year and is designed to estimate effective sample size, not the overall fit of the model to the data. Although all predictions are inside the 95% CIs for individual years, the residuals show a trend across multiple years that indicates substantial lack of fit of the model to the age data. See also figure D.1 in the assessment (Tanimoto et al., 2021b), which shows increasing mean age through time.

This pattern suggests that (if selectivity is constant) there has been a reduction in total mortality which the model is underestimating. This is consistent with what you would expect if the population was recovering more than predicted by the model. An alternative explanation is a change in sampling selectivity through time.

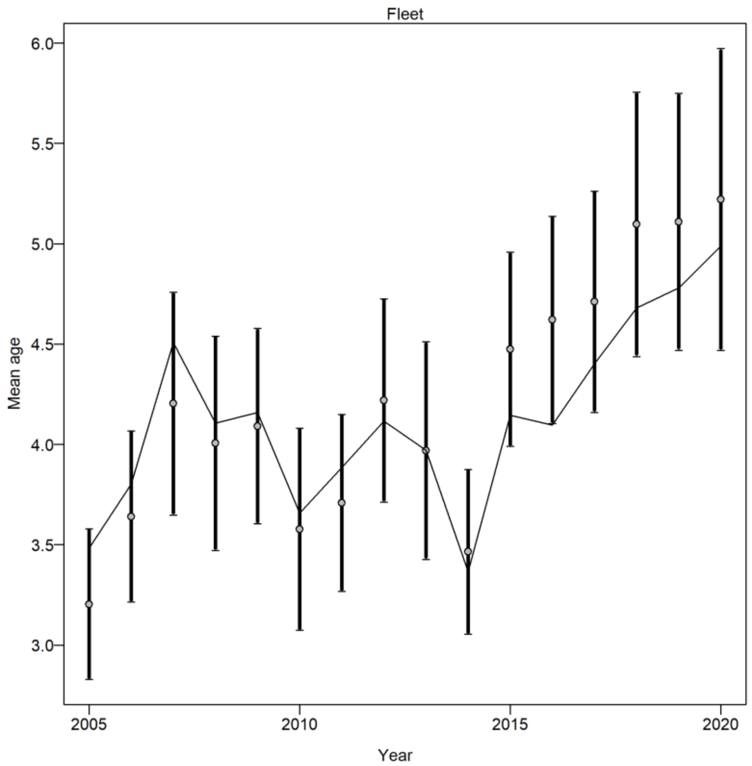

Figure 7: Diagnostic plot generated by r4SS from the Stock Synthesis 3 model 1, showing mean age by year from conditional data (aggregated across length bins) with 95% confidence intervals.

However, fits to the length composition data show the opposite pattern, with more large fish observed than expected early in the time series, and vice versa late in the time series (Figure 8). This inconsistency between length and age data may be due to changes in sampling selectivity through time, bias in the growth curve (see next section), or time-varying length selectivity associated with commercial effort targeting particular lengths. This also suggests conflict in the model between the length composition and age-at-length data sets.

The length composition data are given effective sample sizes based on Francis (2011) method, which is appropriate. However, we question whether including length composition data in the assessment is helpful when much more informative age composition data are available from the same sampling program. The use of length composition data in an age-structured model is usually recommended only when adequate age composition data are not available. The model's use of both length compositions and conditional length at age data assumes that the distribution of length-at-age does not change over time due to fishing induced effects on growth or length selectivity. As noted by Punt et al. (2020), "A key assumption of this approach to including size-composition data in a stock assessment is that the distribution of size-at-age does not change over time due to fishing induced effects on growth or size-selectivity. While such effects are likely small when fishing mortality is low, they could be substantial (and lead to biased assessment outcomes) when fishing mortality is high (e.g., Taylor and Methot, 2013)."

The use of conditional age length data can result in biased growth estimates in the presence of unaccounted for age-based movement when length-based selectivity is assumed (Lee et al., 2017). Alternative approaches that should be considered include fitting to the age compositions calculated from the length compositions and age-length keys directly, or using cohort slicing (e.g., McGarvey et al., 2007, Modelling fish numbers dynamically by age and length).

The length composition data were sampled in many latitudes and across all seasons but were included as a single fishery. 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."

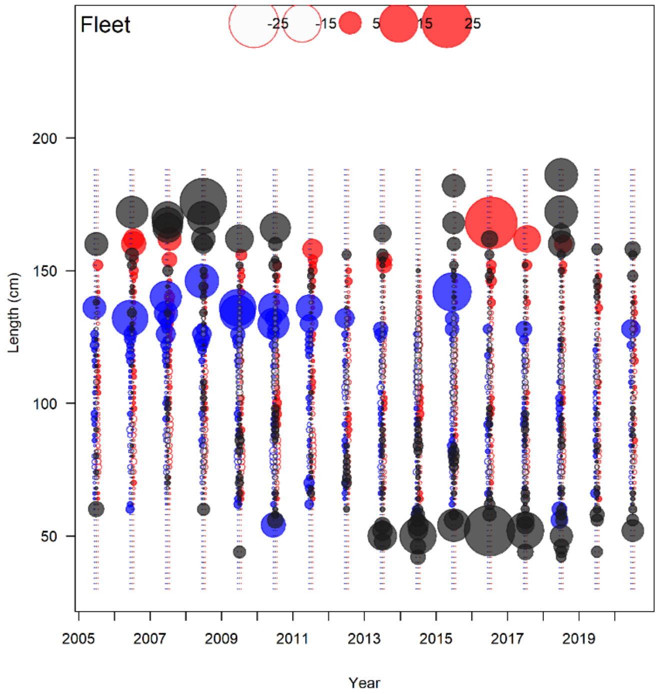

Figure 8: Pearson residuals showing fit to length composition data across time. Closed bubbles are positive residuals (observed > expected), and open bubbles are negative residuals (observed < expected).

2.7 Growth

The assessment's estimated growth curve indicated mean lengths at age lower than published values based on both observed and back-calculated data (Ballagh et al., 2006; McPherson, 1992). (Aside from the von Bertalanffy models of Ballagh et al. (2006) applied to back-calculated data, which do not appear to fit well). This difference raised questions about why the growth curves differed.

The growth model in the assessment fits the age at length data reasonably well (Figure 9). However, fitting a length-at-age growth curve to the assessment's age & length data in a spreadsheet gave higher female L_∞ of 140 cm (Figure 10). A Schnute-Richards growth curve (Schnute and Richards, 1990) fitted much better for older ages, which appear to show continuing growth not well represented by the von Bertalanffy model (see also McPherson, 2007). These estimates were a good match for the empirical estimates of mean length at each age, whereas the base case model growth curve lengths at age were biased low.

We ask why does the assessment, which fits to the age-length data, estimate lower L_∞ than fitting directly to the same data as length at age? Length at age fits outside the model will be biased high by selectivity for younger ages (mainly ages one and two) so the model estimate is preferred for these ages, but that issue should not affect older age classes or L_∞. The Stock Synthesis 3 manual (Methot et al., 2021) notes that fitting the data as age at length will cause bias in the presence of (i) length-based selectivity and age-based movement, (ii) when other age-based processes (e.g., mortality) are not accounted for, or (iii) based on the age-sampling protocol.

In this way, the difference between estimates of age-at-length versus length-at-age is a useful diagnostic that may indicate a problem in the assessment. If the assessment's estimate of the growth curve represents mean length at age in an unfished state, the inconsistency indicates that, to accurately estimate the observed mean age at length, the model may be overestimating mean age at length in an unfished state (i.e., underestimating mean length at age). This is what we would expect if the total mortality estimate was too high (see Lee et al., 2019). Overestimating total mortality is also consistent with the poor fits seen in the age data. Given this inconsistency, we recommend further work to determine its cause.

Regarding the form of the growth curve, we reran the assessment model using a Richards curve (equivalent to one of the forms allowed by Schnute) and obtained a slightly better likelihood, similar depletion, and a growth curve with similar shape to the generalised Schnute curve below (Figure 10) but still with lower female L_{∞} (148 cm) than the curve fitted outside the model (L_{max} = 165 cm) (Schnute-Richards model parameterised with length at age 26). We recommend applying the Schnute-Richards curve in the assessment, although we noted a few fitting hurdles, in that the additional shape parameter for each sex needed to be estimated in a later phase to avoid instability.

Ballagh et al. (2006) estimated that growth varies spatially, which is plausible given local residency, and if correct has significant implications for stock assessment methods. Unfortunately, these analyses appear to have been affected by the poor fit of the von Bertalanffy growth model to older ages, which reduces confidence in their estimates. The long-term monitoring program has collected a large amount of length and age data from across the state. This is a valuable resource which should be leveraged to improve understanding of stock structure. These data could very easily be analysed to investigate factors affecting growth.

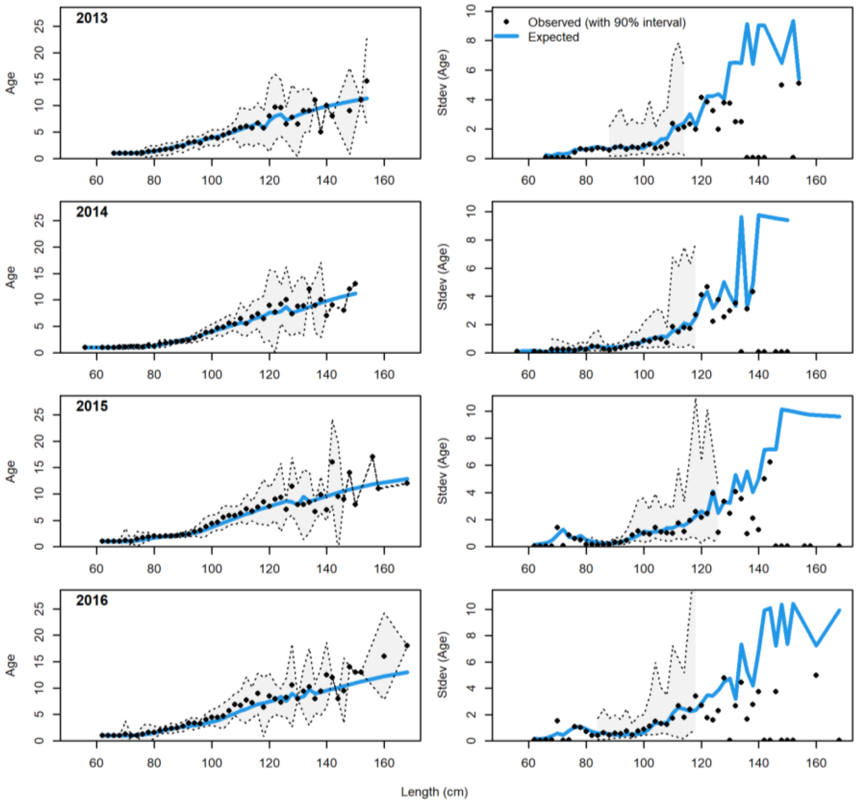

Figure 9: Conditional age at length plot from r4SS model 1, for the years 2013-2016, showing mean age and std. dev. In conditional age at length. Left plots are mean A@L by size-class (obs. And exp.) with 90% Cis based on adding 1.64 SE of mean to the data. Right plots in each pair are SE of mean A@L (obs. And exp.) with 90% Cis based on the chi-square distribution.

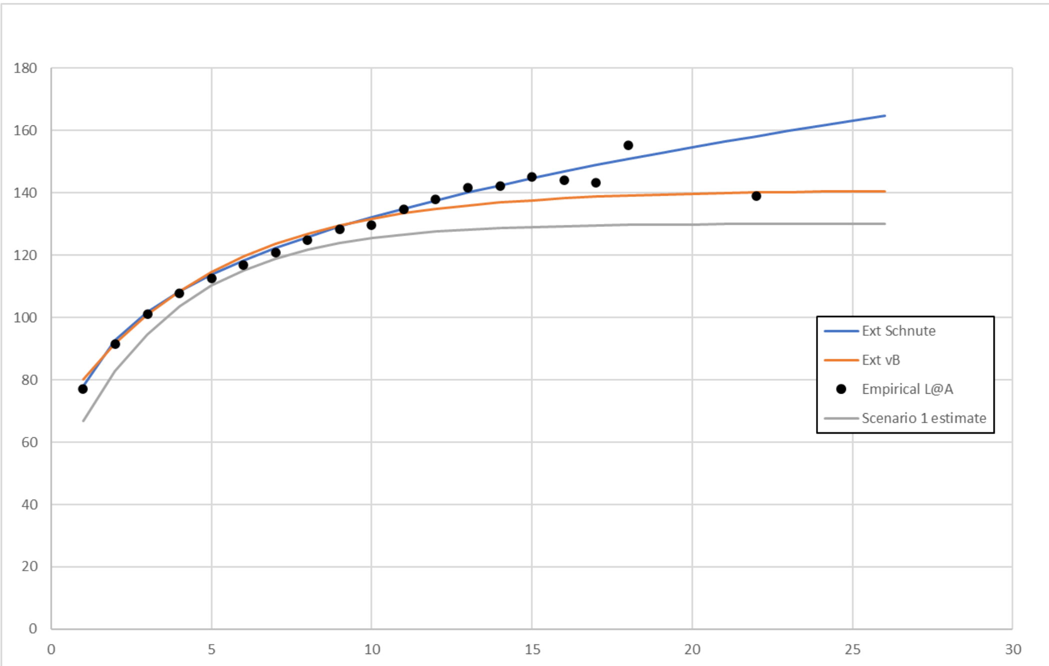

Figure 10: Growth curves fitted to length at age data extracted from the Stock Synthesis 3 model input files, together with the growth curve estimated for Scenario 1, and empirical length at age estimates from the data.

2.8 Selectivity

The model includes only one fishery and specifies selectivity as constant and asymptotic across all areas, time periods, and fishing sectors. This appears to be inconsistent with the results of Tobin and Mapleston (2004), who found a number of sources of variation. They identified differences between sectors, showing that commercial fishing gears select for larger, faster growing young mackerel, whereas recreational fishers tend to harvest smaller and younger mackerel, apart from a subset of the recreational fishery who target very large fish. Length of fish were affected by the gear used by commercial fishers, with cryptic rod and reel methods taking larger fish than wire. The commercial catch saw variation in age structure between months. Coastal region was also associated with variation in length and age structure, though mainly for the recreational catch. Tobin and Mapleston (2004) also noted that commercial fishermen target size ranges favoured by the market, which may explain the smaller length variation of commercial catches. Some studies report older Spanish mackerel becoming solitary, with large fish more likely to be solitary and potentially to occur in different locations (Mackie et al., 2005), which would lead to selectivity varying spatially and between fishing methods.

Analyses of Queensland data (Tobin and Mapleston, 2004) were based on just two years' sampling. We note that length composition data presented in the 2018 assessment (O'Neill et al., 2018, page 92) also appears to show a narrower and more consistent length range for commercial length frequencies, with recreational length frequency data including larger fish in most years and appearing more variable through time. We recommend exploring the long-term monitoring length and age dataset to identify spatial, seasonal, and sectoral effects on age and length structure in the catch.

Misfits to composition data increase uncertainty and may bias the assessment. The assessment shows some misfit to both the age and length data. Adding fisheries as appropriate would likely improve the fit to the composition data, make better use of the information in the data, and potentially reduce the conflict between data sets. It is not necessary to model all selectivity patterns, but it is important to avoid misfits to the composition data. A model with size-based selectivity that fits to size data may be particularly sensitive to these issues.

Initially, we suggest applying separate fisheries for commercial and recreational fishers, because selectivity is likely to differ. Further, selectivities for the commercial fishery pre- and post- the management change in 2005 may change which should be investigated.

Asymptotic selectivity has been applied, which may 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. Commercial selectivity is likely best modelled as double normal rather than asymptotic, since commercial fishing is less likely to target or capture the largest length classes (see figures 8.6-8.7 of Tobin and Mapleston, 2004). Recreational selectivity could be modelled as asymptotic since they take a wider size range than commercials, including very large fish, but with smaller length on average. However, if larger Spanish mackerel become solitary then they are less likely to be caught by fishers targeting schools, so the best approach may be to allow the recreational fishery selectivity of large fish to decline but constrained to remain above zero. Selectivity estimates in the model can be checked using the empirical selectivity diagnostic (Minte-Vera et al., 2021).

2.9 Abundance and recruitment trends

The total catch has been declining since 1980, and since 2005 following a license buyback, catch has been almost half of earlier periods (Figure 4). In 2005 approximately 35% of the Great Barrier Reef Marine Park was closed to fishing, providing numerous potential spatial refuges from fishing. These are substantial management changes which would be expected to benefit the stock.

Consistent with lower fishing mortality, the average age of fish in the catch has steadily increased since 2015 (Figure 7). Nominal catch rates were also higher after 2005 than before (Figure 2). However, the CPUE time series used in the assessment have all been declining since 2005, with (if the CPUE is reliable) no apparent response of the abundance to the reduction in estimated catch (Figure 3).

To match this trend and predict the observed decline in CPUE across the 1989-2005 period, the assessment needed to use recruitment, via high recruitment deviates before 2005 and low deviates after 2005. Before 2005, 12 deviates are above the median and 4 below, while from 2005 on, 10 of the 14 deviates are below the median and only 3 above. The model uses this recruitment trend to generate a declining abundance (and therefore CPUE) trend, which is otherwise inconsistent with stock productivity and the greatly reduced catch. Unusual recruitment trends are sometimes used as a diagnostic of modeling problems (Merino et al., 2022). The lack of recovery in CPUE after such a large reduction in catch 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.

The model fits poorly to the decadal time series, with residual trends that suggest less decline than estimated by the model. It also fits poorly to the age data, with residual trends that suggest overestimation of total mortality.

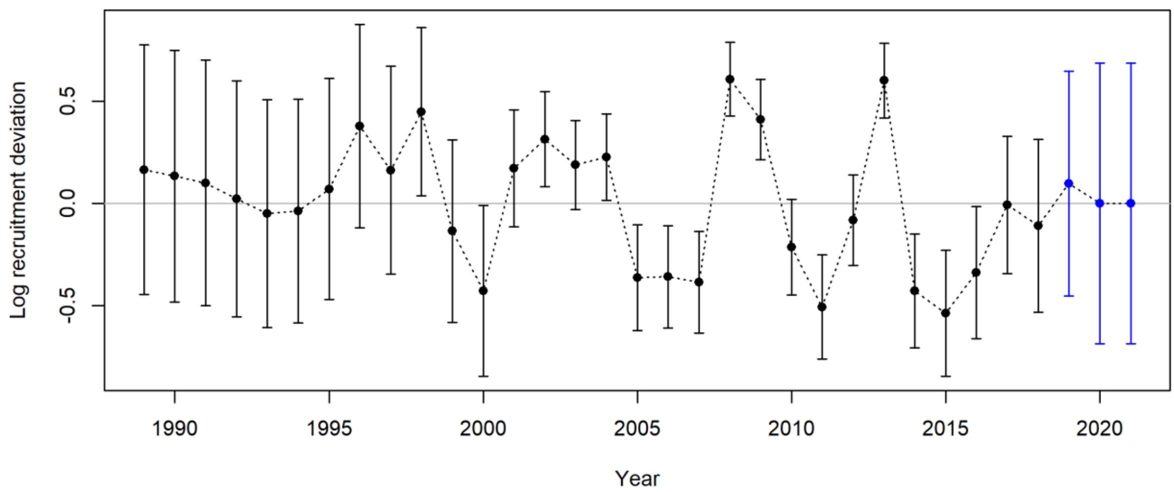

Figure 11: Recruitment deviate time series in base case model.

2.10 Model diagnostics

Likelihood profiles on the population scaling parameter are a well-established tool for exploring the influence of different data components in stock assessments (Lee et al., 2014). In the great majority of assessments there is a single optimum overall, and a single optimum for each likelihood component. In this assessment, virgin recruitment profiles for both the base case and the scenario 4 model had two optima in the total likelihood with a spike of higher likelihood in between. Multiple optima also occur for the age, length, and index likelihoods in the base case, and for all those plus the recruitment likelihood in the scenario 4 model. The likelihood space becomes relatively flat above a certain point, apart from the large spike in the middle.

In both cases the higher solutions are within less than 4 likelihood points and may be within the 95% uncertainty distribution. A likely explanation is that the fits at higher values of Rzero were unstable. We ran some alternative models to explore aspects of the model configuration and a number of these failed to converge.

This kind of pattern in likelihood profiles is not often seen, and in our experience is usually associated with errors. Errors in Stock Synthesis 3 code are unlikely, but the model configuration may cause some parameters to be estimated at unrealistic values. The best case is that the alternative stable states occur at unrealistic values of one or more parameters such as CV on length-at-age, which can be resolved by changing parameter boundaries in the SS control file. The assessment has features that may cause instability, particularly the one-way-trip of the declining CPUE series, which fails to respond to a large reduction in catch. The internal conflict between datasets is also likely to cause instability (conflict between the information in the catch, the age-age data and the historical CPUE on the one hand, and the recent CPUE and the length composition data on the other hand).

We recommend that the analysts thoroughly explore the likelihood space to better understand the model states and parameter values associated with each solution.

Data conflicts are apparent in the assessment diagnostics (e.g., the fits to the length composition, age-length data and the CPUE) that may be resolved by improving the observations supplied to the model or with alternative model parameterisations. We note that conflicts in data are usually not conflicts between the observations but rather a result of model misspecification (Francis, 2017). When there is data conflict within a stock assessment, it is best addressed via alternative models that resolve the internal conflict by the use of assumptions to explain the conflict and by excluding (or significantly downweighing) individual data sets to evaluate the effect of the conflict on model outcomes (Francis, 2017).

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.

3. Discussion

We have considered many aspects of the stock assessment for Spanish mackerel. It is a difficult stock to assess, with a complex spatial domain, likelihood of hyperstability in the CPUE, and uncertainties about fish behaviour. 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.

In conclusion, we find that the model shows signs of misspecification, with residual trends in the decadal CPUE time series, age structure and length composition data, potential bias in the estimated growth curve, and instability in model fits and the likelihood profiles. Some of these issues relate to the way in which observational data were determined and supplied to the assessment model, and some to the model assumptions and structure. We think these issues should be resolved before the model is used to develop management advice.

Overall, the diagnostics would suggest that the misspecifications in the model would likely lead it to overestimate total mortality and therefore underestimate total biomass. These problems may largely be driven by the lack of recovery in the estimated CPUE after the large reduction in estimated catch from 2005. Nominal CPUE does recover after 2005, but after the processes of CPUE standardization and adjustment for fishing power, the trend becomes a steady decline. The methods currently used for CPUE analysis are likely to bias results towards a more depleted outcome, unless the rate of mixing of the entire east coast Spanish mackerel population was high. The methods used for composition sampling are affected by the same issue. Studies of Spanish mackerel in northern Australia suggest that population mixing is likely to be relatively low north of Townsville, and rapid random mixing appears unlikely south of Townsville. We recommend changes to the CPUE standardisation process in several areas, including the probability model, the spatial weighting process, and the adjustments for fishing power.

4. Recommendations

1. We note that the model show signs of misspecification and recommend that these issues should be resolved before the model is used in the development of management advice.
2. We recommend that analysts resolve indications of model instability and poor fit to the data: residual trends in the age structure, historical CPUE data, and length composition data which suggest data conflict; a growth curve that estimates unrealistically low length at age; an unusual pattern of multiple optima in likelihood profiles; and instability in model fitting.
3. Reducing uncertainty about the degree of mixing among areas is a high priority. Available information about Spanish mackerel behaviour should be used to develop hypotheses about population structure, and to design a research programme. Analysts should allow for the implications of incomplete mixing in their data preparation and selection of stock assessment methods.
4. CPUE: review and update approach to the development of the CPUE index, including but not limited to the following:
    a. Explore and characterize the catch and effort data, including factors associated with targeting, to identify patterns likely to affect CPUE.
    b. Change approach to weighting by latband, to weight by the product of relative habitat area (or an appropriate proxy) and density rather than by catch or the number of records.
    c. Review the fishing power estimates and their relevance for the fishing methods included in the CPUE index before using them in the base case assessment model. Consider whether search-related factors should be included in the catch rate or the probability model.
    d. Produce indices without data from the Latitude 19 area where spawning schools concentrate and hyperstability is most likely to affect CPUE.
    e. Allow for the effects of targeting on catch rates. Run models that standardize the CPUE of vessels that exclusively target Spanish mackerel.
    f. Drop runs that include the probability model until issues with it have been addressed.
    g. Produce a separate document with a comprehensive overview of CPUE methods, results, and diagnostics.
5. Auxiliary analyses:
    a. Explore the long-term monitoring length and age data set to identify spatial, seasonal, and sectoral effects on age and length composition of the catch. Check for evidence of spatial patterns in length-at-age, and spatial and temporal patterns in age-at-length.
6. Catch and related issues:
    a. Review approaches used to develop time series of recreational catch; and explore sensitivity of assessment outcomes to alternative time series.
    b. When preliminary estimates of depredation rates become available, we recommend testing their effects in the assessment.
7. Input parameters:
    a. For steepness, we recommend applying a range of values from 0.45 to 0.95, with a median value close to the 0.69 FishLife estimate for Scombridae.
    b. For recruitment variability (σ_R), the commonly assumed value of 0.6 should be used when fitting the model to estimate recruitments, instead of 0.35.

c. For natural mortality we recommend (i) updating the prior mean to M=5.4/Amax, based on Hamel and Cope (2022), who identified a modelling error in the approach of Then et al (2015); (ii) replacing constant M at age with the biologically well-justified Lorenzen approach of setting M inversely proportional to body length (Lorenzen, 2022); and (iii) fixing M in the model at a range of values across the prior, rather than estimating it in the model.
d. We recommend applying the Schnute-Richards (1990) growth curve rather than the von Bertalanffy.

8. Model structure:
    a. Update the approach to modelling selectivity to better fit the composition data: separate the commercial and recreational fisheries; consider spatial and/or seasonal structure; non-asymptotic selectivity, particularly for the commercial fishery.

5. References

Ballagh, A.C.; Begg, G.A.; Mapleston, A.; Tobin, A. Growth trends of Queensland east coast Spanish mackerel (Scomberomorus commerson) from otolith back-calculations. Marine and Freshwater Research. 57:383-393; 2006.

Buckley, S. Historical change in the marine ecology and fisheries of Australia and Kenya. School of Biological Sciences: The University of Queensland; 2016

Buckley, S.M.; Thurstan, R.H.; Tobin, A.; Pandolfi, J.M. Historical spatial reconstruction of a spawning-aggregation fishery. Conservation Biology. 31:1322-1332; 2017.

Buckworth, R.; Newman, S.; Ovenden, J.; Lester, R.; McPherson, G. eds. The stock structure of northern and western Australian Spanish mackerel. Final Report, Fisheries Research & Development Corporation Project 1998/159: Department of Primary Industry, Fisheries and Mines, Northern Territory Government, Australia; 2007

Campbell, R.A. Constructing stock abundance indices from catch and effort data: Some nuts and bolts. Fisheries Research. 161:109-130; 2015. 10.1016/j.fishres.2014.07.004

Carruthers, T.R.; McAllister, M.K.; Ahrens, R.N.M. Simulating spatial dynamics to evaluate methods of deriving abundance indices for tropical tunas. Canadian Journal of Fisheries and Aquatic Sciences. 67:1409-1427; 2010.

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.

Chapman, B.; Hulthén, K.; Brodersen, J.; Nilsson, P.A.; Skov, C.; Hansson, L.A.; Brönmark, C. Partial migration in fishes: causes and consequences. Journal of fish biology. 81:456-478; 2012.

Francis, R.C. Revisiting data weighting in fisheries stock assessment models. Fisheries Research. 192:5-15; 2017.

Francis, R.I.C.C. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences. 68:1124-1138; 2011.

Gilman, E.; Clarke, S.; Brothers, N.; Alfaro-Shigueto, J.; Mandelman, J.; Mangel, J.; Petersen, S.; Piovano, S.; Thomson, N.; Dalzell, P.; Donoso, M.; Goren, M.; Werner, T. Shark interactions in pelagic longline fisheries. Marine Policy. 32:1-18; 2008. 10.1016/j.marpol.2007.05.001

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.

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.

Hoyle, S.D.; Lee, S.I.; Kim, D.N. CPUE standardization for southern bluefin tuna (Thunnus maccoyii) in the Korean tuna longline fishery, accounting for spatiotemporal variation in targeting through data exploration and clustering. PeerJ. 10:e13951; 2022.

Klaer, N. Independent review of the stock assessment of east coast Spanish mackerel (*Scomberomorus commerson*) in Queensland, Australia. Prepared for Queensland Department of Agriculture and Fisheries; 2021

Lawson, A. An investigative analysis of Queensland's statewide recreational fishing surveys: Identifying bias in self reported catch and effort data. 2015.

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.; Piner, K.R.; Taylor, I.G.; Kitakado, T. On the use of conditional age at length data as a likelihood component in integrated population dynamics models. Fisheries Research. 216:204-211; 2019.

Lee, H.; Thomas, L.; Piner, K.; Maunder, M. Effects of age-based movement on the estimation of growth assuming random-at-age or random-at-length data. Journal of Fish Biology. 90:222-235; 2017.

Leigh, G.M.; O'Neill, M.F.; Stewart, J. Stock assessment of the Australian east coast tailor (Pomatomus saltatrix) fishery. 2017.
Lo, N.C.H.; Jacobson, L.D.; Squire, J.L. Indices of relative abundance from fish spotter data based on delta-lognormal models. Canadian Journal of Fisheries and Aquatic Sciences. 49:2515-2526; 1992.
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.
Mackie, M.C.; Lewis, P.D.; Gaughan, D.J.; Newman, S.J. Variability in spawning frequency and reproductive development of the narrow-barred Spanish mackerel (Scomberomorus commerson) along the west coast of Australia. Fishery Bulletin. 103:344-354; 2005.
Magnusson, A.; Magnusson, M.; Morison, J. Economic and Social Indicators for the Queensland East Coast Spanish Mackerel Fishery, 2017/18 and 2018/19. Department of Agriculture and Fisheries (Queensland); 2020
Mangel, M.; MacCall, A.D.; Brodziak, J.; Dick, E.J.; Forrest, R.E.; Pourzand, R.; Ralston, S.; Rose, K. A perspective on steepness, reference points, and stock assessment. Canadian Journal of Fisheries and Aquatic Sciences. 70:930-940; 2013. 10.1139/cjfas-2012-0372
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.; 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
McGarvey, R.; Feenstra, J.E.; Ye, Q. Modeling fish numbers dynamically by age and length: partitioning cohorts into "slices". Canadian Journal of Fisheries and Aquatic Sciences. 64:1157-1173; 2007. 10.1139/f07-080
McKechnie, S.; Hoyle, S.; Harley, S. Longline CPUE series that account for changes in the spatial extent of fisheries. 2013
McPherson, G. Historical stock definition research on *Scomberomorus commerson* in Queensland waters. in: Buckworth R., Newman S., Ovenden J., Lester R., McPherson G., eds. The stock structure of northern and western Australian Spanish mackerel Final Report, Fisheries Research & Development Corporation Project 1998/159: Department of Primary Industry, Fisheries and Mines, Northern Territory Government, Australia; 2007
McPherson, G.R. Age and growth of the Narrowed-barred Spanish Mackerel (Scomberomorus commerson Lacepede, 1800) in North-eastern Queensland Waters. Australian Journal of Marine and Freshwater Research. 43:1269-1282; 1992.
Merino, G.; Urtizberea, A.; Fu, D.; Winker, H.; Cardinale, M.; Lauretta, M.V.; Murua, H.; Kitakado, T.; Arrizabalaga, H.; Scott, R. Investigating trends in process error as a diagnostic for integrated fisheries stock assessments. Fisheries Research. 256:106478; 2022.
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.
Methot, R.D.; Wetzel, C.R.; Taylor, I.G.; Doering, K.L.; Johnson, K.F. Stock Synthesis User Manual: Version 3.30.17. Seattle, WA: NOAA Fisheries; 2021
Minte-Vera, C.V.; Maunder, M.N.; Aires-da-Silva, A.M. Auxiliary diagnostic analyses used to detect model misspecification and highlight potential solutions in stock assessments: application to yellowfin tuna in the eastern Pacific Ocean. ICES Journal of Marine Science. 78:3521-3537; 2021.

Mitchell, J.; McLean, D.; Collin, S.; Langlois, T. Shark depredation in commercial and recreational fisheries. Reviews in Fish Biology and Fisheries. 28:715-748; 2018a.
Mitchell, J.; McLean, D.; Collin, S.; Taylor, S.; Jackson, G.; Fisher, R.; Langlois, T. Quantifying shark depredation in a recreational fishery in the Ningaloo Marine Park and Exmouth Gulf, Western Australia. Marine Ecology Progress Series. 587:141-157; 2018b.
Munyandorero, J. Inferring prior distributions of recruitment compensation metrics from life-history parameters and allometries. Canadian Journal of Fisheries and Aquatic Sciences. 77:295-313; 2020.
O'Neill, M.F.; Langstreth, J.; Buckley, S.M.; Stewart, J. Stock assessment of Australian east coast Spanish mackerel: Predictions of stock status and reference points. 2018.
Parsa, M.; Emery, T.J.; Williams, A.J.; Nicol, S. A robust métier-based approach to classifying fishing practices within commercial fisheries. Frontiers in Marine Science. 7:552391; 2020.
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.
Sampson, D.B. Fishery selection and its relevance to stock assessment and fishery management. Fisheries Research. 158:5-14; 2014.
Schnute, J.T.; Richards, L.J. A unified approach to the analysis of fish growth, maturity, and survivorship data. Canadian Journal of Fisheries and Aquatic Sciences. 47:24-40; 1990.
Szuwalski, C.S.; Vert-Pre, K.A.; Punt, A.E.; Branch, T.A.; Hilborn, R. Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish and Fisheries. 16:633-648; 2015.
Tanimoto, M.; Fox, A.; O'Neill, M.F.; Langstreth, J. Authors' response to east coast Spanish mackerel assessment review. 2021a.
Tanimoto, M.; Fox, A.; O'Neill, M.F.; Langstreth, J. Stock assessment of Australian east coast Spanish mackerel (Scomberomorus commerson). 2021b.
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.
Taylor, I.G.; Methot, R.D. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fisheries Research. 142:75-85; 2013. 10.1016/j.fishres.2012.08.021
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.
Thorson, J.T. Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. Fish and Fisheries. 21:237-251; 2020.
Thurstan, R.H.; Buckley, S.M.; Ortiz, J.C.; Pandolfi, J.M. Setting the record straight: assessing the reliability of retrospective accounts of change. Conservation Letters. 9:98-105; 2016a.
Thurstan, R.H.; Buckley, S.M.; Pandolfi, J.M. Using commercial and recreational fisher knowledge to reconstruct historical catch rates for Queensland snapper (Chrysophrys auratus), Spanish mackerel (Scomberomorus commerson) and coral trout (Plectropomus spp.): long-term data for incorporation into future stock assessments. 2016b.
Tobin, A.; Heupel, M.; Simpfendorfer, C.; Pandolfi, J.; Thurstan, R.; Buckley, S. Utilising innovative technology to better understand Spanish mackerel spawning aggregations and the protection offered by marine protected areas. 2014.
Tobin, A.; Mapleston, A. Exploitation dynamics and biological characteristics of the Queensland east coast Spanish mackerel (Scomberomorus commerson) fishery: CRC Reef Research Centre; 2004
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.; Saunders, T.; Robins, J.; Harry, A.; Johnson, J.; Maynard, J.; Saunders, R.; Pecl, G.; Sawynok, B.; Tobin, A. Implications of climate change impacts on fisheries resources of northern Australia. Part 1: Vulnerability assessment and adaptation options. FRDC Project No: 2010/565. Australia: James Cook University; 2014

Williams, R.; Lester, R. Stock structure of Spanish mackerel Scomberomorus commerson along the Australian east coast deduced from parasite data. Journal of Fish Biology. 68:1707-1712; 2006.

Wood, S.; Wood, M. Package 'mgcv'. R package version 1.8-32. 2020

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: Project objectives

The independent review project objectives are listed below.

Project objective: To provide a report reviewing the data and assumptions in the 2021 east coast Spanish mackerel stock assessment, including peer review of the assessment.

Specific objectives:

1. Review the 2021 east coast Spanish mackerel assessment, considering the assessment document, the 2021 desktop review, and the assessment authors' responses to the peer review.
2. Comment on model inputs, the assessment model structure and implementation, model diagnostics, and model outputs, and the adequacy of these to achieve the assessment objectives.
3. Comment on the accuracy and reliability of key statements in the report summary and conclusion.
4. Comment on the recommendations for management and monitoring and inclusion of additional data in future assessments.
5. Make recommendations for additional analyses to support future assessments.