BENCHMARK WORKSHOP FOR DEMERSAL SPECIES (WKDEM)

VOLUME 2 | ISSUE 31

ICES SCIENTIFIC REPORTS

RAPPORTS SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA
CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H.C. Andersens Boulevard 44-46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

ISSN number: 2618-1371 I © 2020 International Council for the Exploration of the Sea

ICES Scientific Reports

Volume 2 | Issue 31

BENCHMARK WORKSHOP FOR DEMERSAL SPECIES (WKDEM)

Recommended format for purpose of citation:

ICES. 2020. Benchmark Workshop for Demersal Species (WKDEM).
ICES Scientific Reports. 2:31. 136 pp. http://doi.org/10.17895/ices.pub.5548

Editors

Daniel Howell • Richard Nash

Authors

Robert Boenish • Olav Nikolai Breivik • Bjarte Bogstad • Anatoly Chetyrkin • Robin Cook • Gjert Endre Dingsør • Helen Dobby • Johanna Fall • Andrzej Jaworski • Edda Johannesen • Alexandros Kokkalis • Yuri Kovalev • Tanja Miethe • Anders Nielsen • Alfonso Pérez-Rodriguez • Alexey Russkikh • Amy Schueller

ICES CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

Contents

i Executive summary

A Benchmark Workshop for Demersal species (WKDEM), chaired by External Chair Richard D.M. Nash, UK and ICES Chair Daniel Howell, Norway and attended by two invited external experts Amy Schueller, US and Robert Boenish, US met at ICES, HQ, Copenhagen 9–13 December 2019 for a data evaluation meeting and at ICES HQ, Copenhagen, Denmark, for a Benchmark meeting 10–14 February 2020. In addition, two WebEx meetings were convened (23rd October 2019 and 23rd February 2020) to discuss data issues and logistics related to the work plan prior to the data evaluation meeting and the Benchmark.

Four demersal stocks were Benchmarked in WKDEM, Haddock (*Melanogrammus aeglefinus*) in subareas 1 and 2 (Northeast Arctic) (Had.27.1–2), Cod (*Gadus morhua*) in Division 6.a (West of Scotland) (Cod.27.6a), Whiting (*Merlangius merlangus*) in Division 6.a (West of Scotland) (Whg.27.6a) and Whiting in Division 3.a (Skagerrak and Kattegat) (Whg.27.3a).

The 2020 benchmark for Northeast Arctic Haddock was proposed due to the poor retrospective pattern (annually changing perception of the stock) for the assessment in 2018. The final SAM model was based on the updated dataset and a new SAM setting allowing for plus groups in surveys. The final model had a lower AIC and more parameters compared to the configuration used in the previous assessment with the same data as the final model. Some changes were made to the prediction models, mainly to reflect the current perception of the stock dynamics. The reference points were not changed.

The 2020 benchmark for 6.a cod was proposed initially due to uncertainties in the catch data and the selectivity patterns in the catch and survey data. In addition, a separate analysis of the stock estimated a different trend in fishing mortality and higher biomass, which lead to an inter-benchmark in February 2019. A number of alternative assessment methods were considered with a decision to switch from the TSA model to SAM. An improved approach to estimating area-misreported landings was implemented, and maturity and natural mortality values were modified. The survey data included in the final assessment are consistent (in terms of year and age range) with those used previously. The reference points were updated having been derived following the ICES guidelines and using EqSim. There was some concern that an alternative model, FAF model, along with an extension to allow natural mortality to be estimated by the model ("TVM" model) provided a different perception of the stock dynamics. The latter model was not presented for evaluation prior to the Benchmark and as such needs to be evaluated as to its efficacy for assessment of this stock in the future.

The 2020 benchmark for 6.a whiting was proposed because the assessment was influenced partly by incorrect reporting of landings data (species and quantity) from the past which directly affected the perception of the stock. There were newly available catch data and reworked survey indices (including the combined Irish and Scottish Q4 survey), plus reworked maturity and natural mortality estimates were available. It became clear that running TSA with the new data and changed survey configuration resulted in poorly converged optimisation runs. An age-aggregated stochastic Surplus Production in Continuous Time (SPiCT) model was tried but the modelled stock level was highly uncertain, although the trends were considered more robust. The result was to reclassify 6.a whiting to category 3 using the SPiCT model to give trends, according to the ICES guidelines for data-limited stocks.

The 2020 benchmark for 3.a whiting was proposed because it was considered a category 5 stock with no advice and more information is available (time-series of catches (landings and discards), a new fisheries independent biomass index that combines all relevant scientific surveys in the area) that were not being used but could be utilised to infer stock status and provide better catch

advice. The outcome was to reclassify the stock to category 3 and provide advice using the "2-over-3" trend based approach with the new biomass index. An attempt to provide an analytical assessment for the stock using the surplus production model (SPiCT) was unsuccessful.

ii Expert group information

Expert group name	Benchmark Workshop for Demersal Species (WKDEM)
Expert group cycle	Annual
Year cycle started	2020
Reporting year in cycle	1/1
Chairs	Daniel Howell, Norway
	Richard Nash, UK
Meeting venues and dates	10–14 February 2020, Copenhagen, Denmark, 21 participants

1 Introduction

ICES defines the purpose and aims of a benchmark for assessment as (http://ices.dk/community/advisory-process/Pages/Benchmarks.aspx, accessed February 2020):

'The aim of a benchmark is consensus agreement on an assessment methodology that is to be used in future update assessments, laid down in a stock annex.

ICES advice for fisheries is based on annual fish stock assessments. The methodology for these assessments is evaluated every three to five years in a benchmark workshop. Here, all information – ecosystem and fisheries data, stock distribution, assessment model, forecast method and reference points – is reviewed. A benchmark deals with single stock assessment methods, but it also aims to integrate ecosystem information into the assessment.

A benchmark meeting is open to experts and stakeholders and it is reviewed by external experts throughout the process. Preparing for a benchmark process takes about five to seven months, including a data compilation workshop and a final benchmark meeting. ACOM (Advisory Committee) agrees on benchmarks a year in advance.'

A Benchmark Workshop for Demersal species (WKDEM) was convened over the period 2019/2020 to evaluate the appropriateness of data and methods to determine stock status and investigate methods for short-term outlook, taking agreed or proposed management plans into account, for four demersal stocks (Haddock in divisions 1–2, Cod in Division 6a, Whiting in Division 6a and Whiting in Division 3a). The terms of reference for this Benchmark Workshop are given in Annex 3.

The four gadoid stocks are located in three very different Ecoregions; Northeast Arctic haddock in the Barents Sea (ICES, 2019a), 6a cod and whiting in the Celtic Seas Ecoregion (ICES, 2019b), and 3a whiting in the Greater North Sea Ecoregion (ICES, 2019c). Overviews of all the fisheries occurring in each the ecoregions can be found in the relevant ICES fisheries overviews (2019d, e, f). The stocks undergoing benchmark here are assessed in one of three assessment working groups which generally reflect the ecoregions in which they are located. The Northeast Arctic haddock are assessed in the Arctic Fisheries Working Group (AFWG), 6a cod and whiting in the Working Group for the Celtic Seas Ecoregion (WGCSE) and 3a whiting in the Working Group for the North Sea and Skagerrak (WGNSSK). All four stocks have a long history of exploitation; however, the quantity and quality of information and scientific evidence available for undertaking stock assessment and for understanding the ecology and dynamics varied considerably between areas and stocks.

The Working Documents (WD) related to all of the stocks covered in this benchmark are retained on the WKDEM SharePoint. If the reader wants to obtain a copy of a WD please make a request to the lead author (addresses are given in Annex 1, Participants list).

1.1 References

ICES. 2019a. Barents Sea Ecoregion – Ecosystem overview. In Report of the ICES Advisory Committee, 2019. ICES Advice 2019, Section 5.1, https://doi.org/10.17895/ices.advice.5747.

ICES. 2019b. Celtic Seas Ecoregion – Ecosystem overview. In Report of the ICES Advisory Committee, 2019. ICES Advice 2019, Section 7.1, https://doi.org/10.17895/ices.advice.5749.

ICES. 2019c. Greater North Sea Ecoregion – Ecosystem overview. In Report of the ICES Advisory Committee, 2019. ICES Advice 2019, Section 9.1, https://doi.org/10.17895/ices.advice.5750.

ICES. 2019d. Barents Sea Ecosystem – Fisheries overview. In Report of the ICES Advisory Committee, 2019. ICES Advice 2019, Section 5.2. 28 pp. https://doi.org/10.17895/ices.advice.5705.

ICES. 2019e. Celtic Seas Ecosystem – Fisheries Overview. In Report of the ICES Advisory Committee, 2019. ICES Advice 2019, Section 7.2. 40 pp. https://doi.org/10.17895/ices.advice.5708.

ICES. 2019. Greater North Sea Ecosystem – Fisheries Overview. In Report of the ICES Advisory Committee, 2019. ICES Advice 2019, section 9.2. 42 pp. https://doi.org/10.17895/ices.advice.5710.

2 Haddock (*Melanogrammus aeglefinus*) in subareas 1 and 2 (Northeast Arctic) (Had.27.1–2)

2.1 Why a benchmark?

The benchmark in 2020 for Northeast Arctic Haddock was proposed due to the poor retrospective pattern for the assessment in 2018 (ICES, 2018). The last benchmark was in 2015 (ICES, 2015), which also justified having a new benchmark now.

2.2 Summary of final model (assessment and prediction) and reference point investigations

2.2.1 SAM model

The final SAM model was based on the updated dataset (3.3 below and working documents 3.1–8), and a new SAM setting allowing for plus groups in surveys. The final model had a lower AIC (a reduction from 2413 to 2055) and more parameters (38 vs 43) compared to the configuration used in the previous assessment and the same data as the final model. NB at the benchmark meeting a developer version of SAM was used to run the models, when the prediction – variance link (see below) from the developer version was made available at stockassessment.org (27 March 2020), then mode instead of the mean is used as the prediction in the prediction–variance link to reduce computation time. The results were the same, but the AIC was slightly lower (2052).

Adding a plus group had the strongest impact on the retrospective pattern and improved it greatly (compare ICES, 2019 and Figure 2.1 below, see also WD8), and the retrospective pattern is now good. Mohns Rho (five years): fishing mortality =-0.028; SSB=0.029; TSB=0.004.

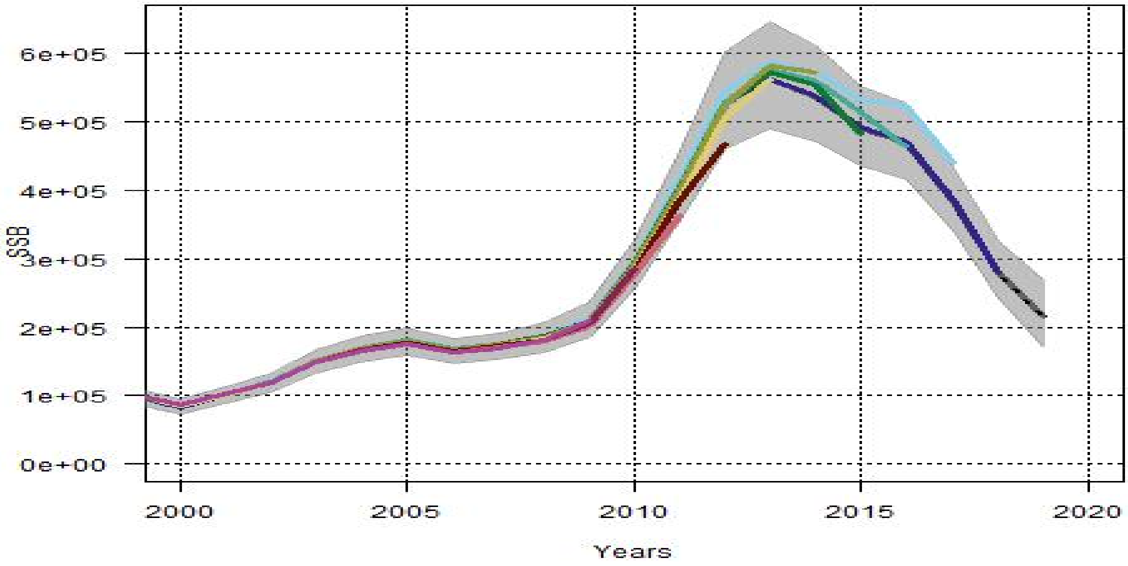


Figure 2.1a. Retrospective pattern (10 year peel) of SSB of final model.

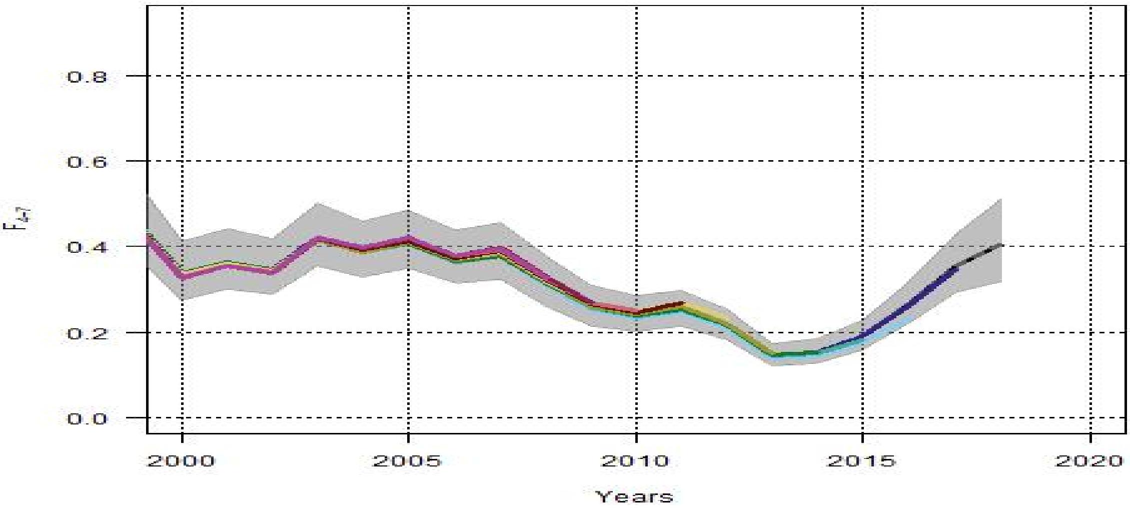


Figure 2.1b. Retrospective pattern (ten year peel) of F_{bar} of final model.

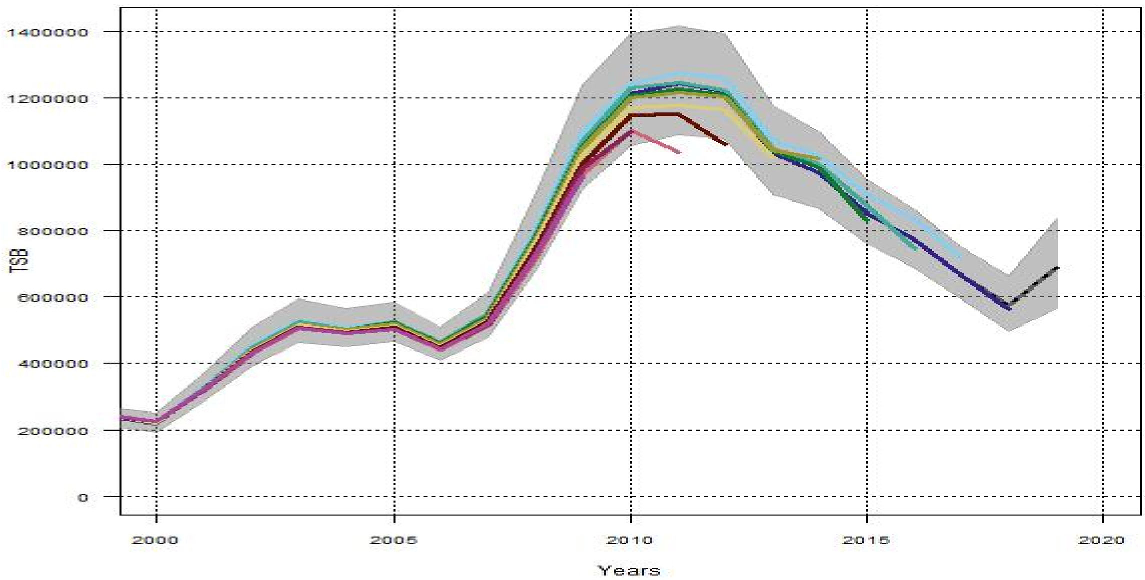


Figure 2.1c. Retrospective pattern (ten year peel) of TSB of final model.

The final model includes a new SAM option which at the moment is available in a development version (https://github.com/fishfollower/SAM/tree/FprocVarMeanLink), and will soon be available at stockassessment.org. The option includes a link between the variance and the prediction of the observations (Breivik, Nielsen and Berg, In prep). This procedure is inspired by Taylor (1961), which observed the same link between the mean and the variance in several surveys. The procedure is further inspired by XSAM (Aanes, 2016) which use a similar link to smooth external variance estimates, which are further used as input in the assessment model.

The prediction–variance link is given on the following functional form:

$$\sigma_a^2 = \log(\alpha_a \mu_{a,y}^{\beta_a - 2} + 1)$$

where $\mu_{a,y}$ is the predicted observation on natural scale, and α_a and β_a are parameters estimated internally in SAM. Note that α_a and β_a are assumed constant across years, and can vary between fleets. By including this relationship, we include flexibility in SAM to assume time-varying observation variance depending on how large the observation is predicted to be.

This change improved the fit of the modelled catches to the observed catches greatly (compare Figure 2.2 to ICES, 2019). The model now predicts the catches well, except for a high catch in 1973. This catch was highly unusual in that it consisted of a large proportion of haddock aged 3–4.

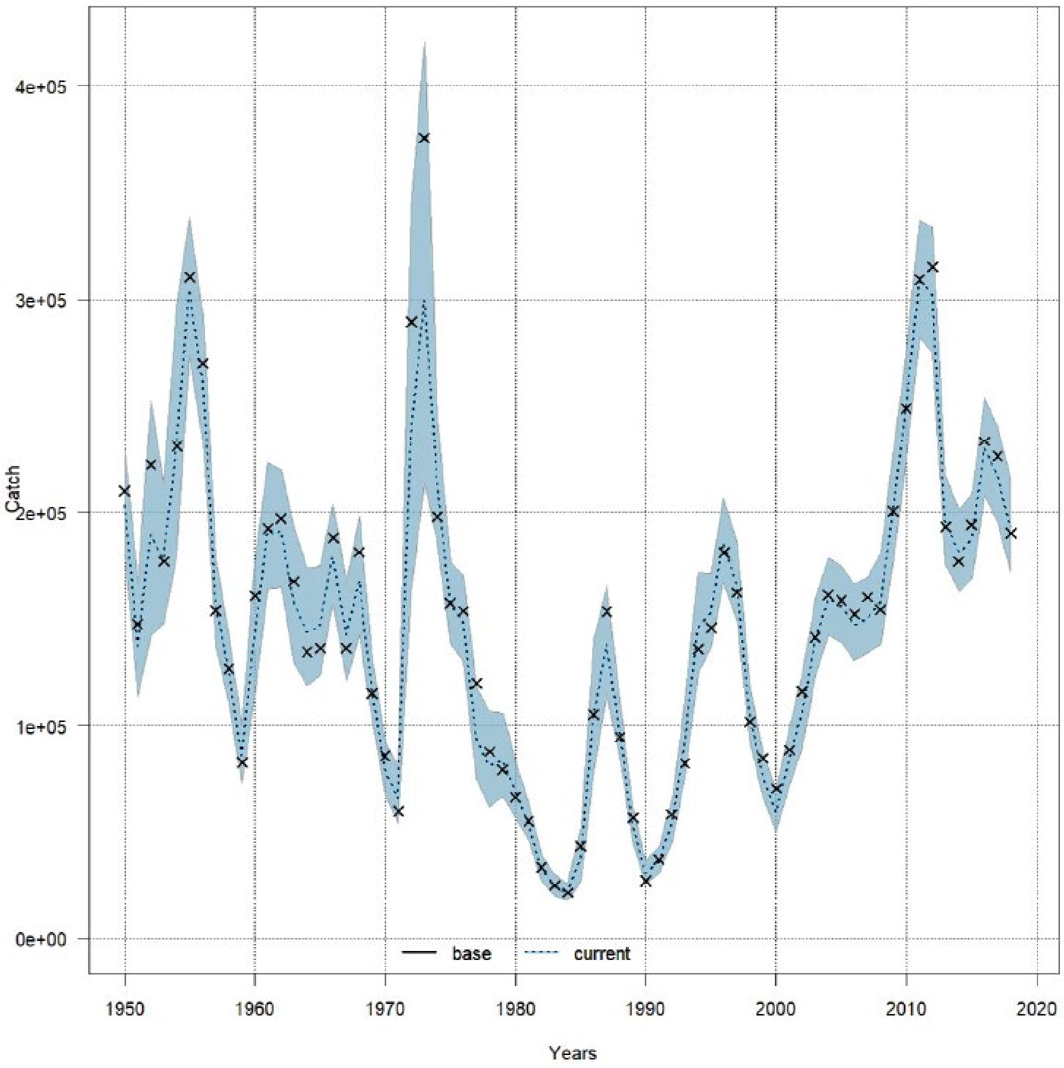


Figure 2.2. Observed (dotted) vs fitted catches (x).

The final model includes correlation structure for the observation variation (Berg and Nielsen, 2016). The coupling of the fishing mortalities and the catchability and configurations linking catchabilities to stock size was also modified. The one-step-ahead (OSA) residuals or process errors did not show any obvious patterns (Figure 2.3). The leave one out diagnostics show that the model is strongly driven by the catch data (Figure 2.4). The results of the final model are shown in Figure 2.5. The results are compared with last years' assessment in Figure 2.6.

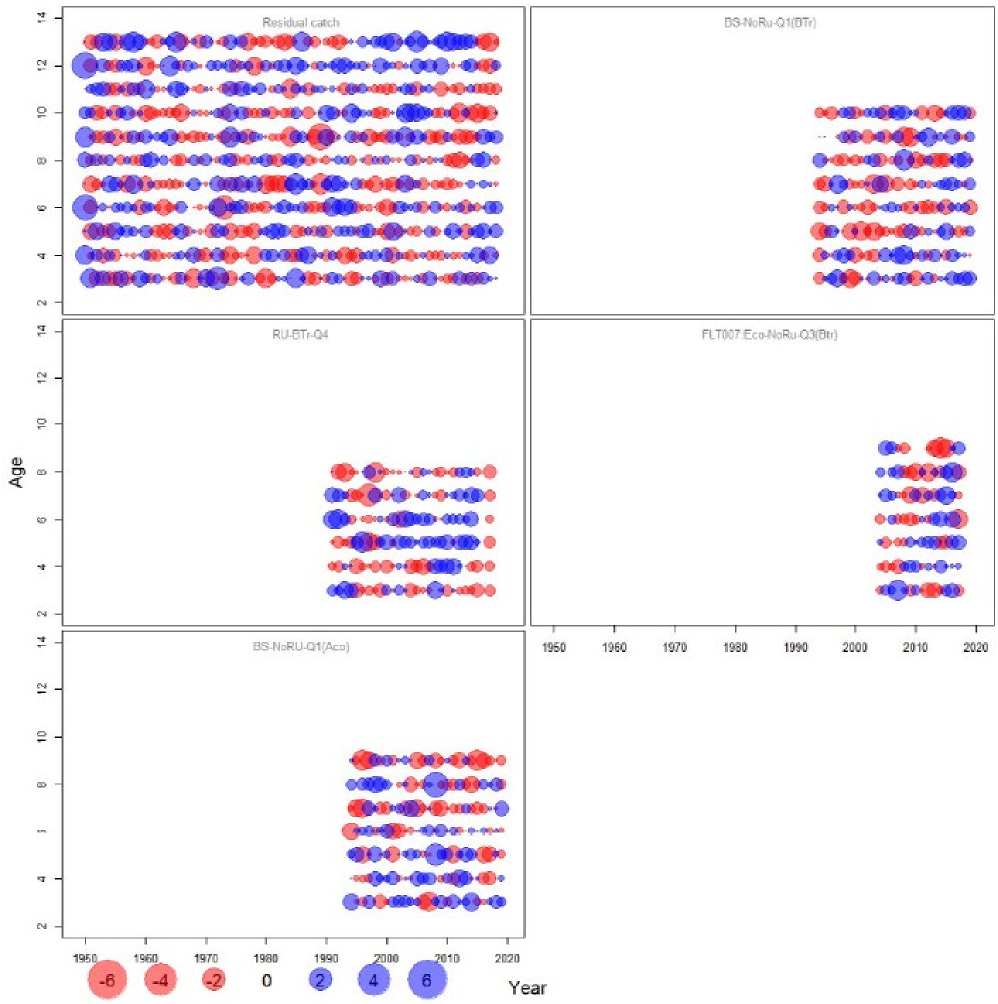


Figure 2.3a. OSA residuals from final model.

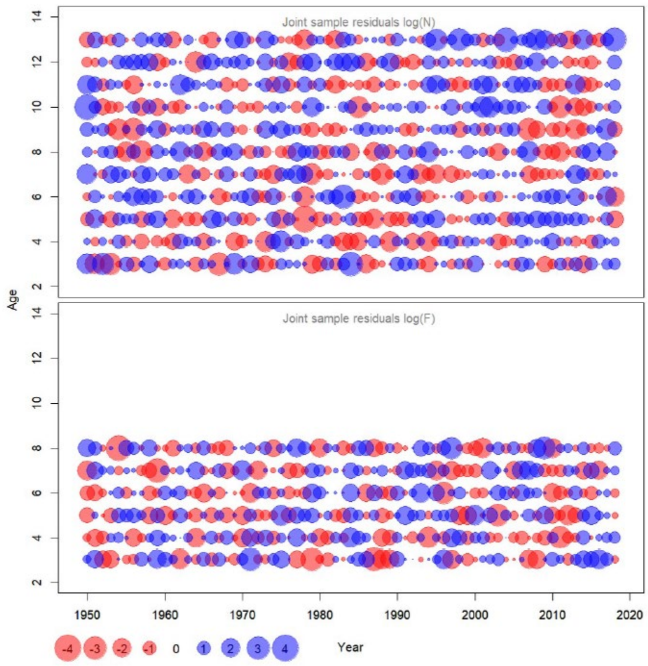
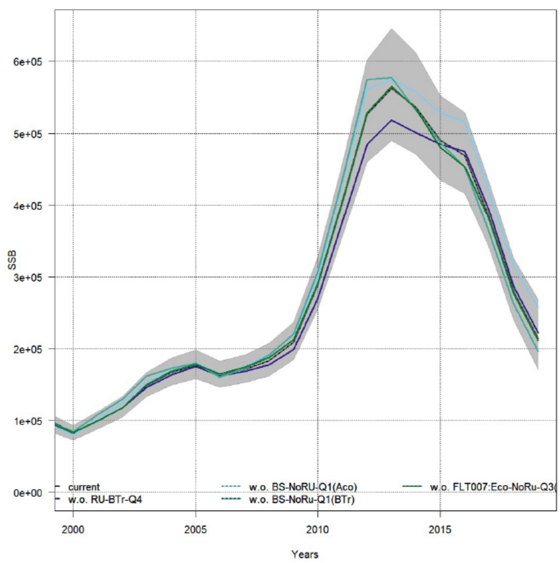
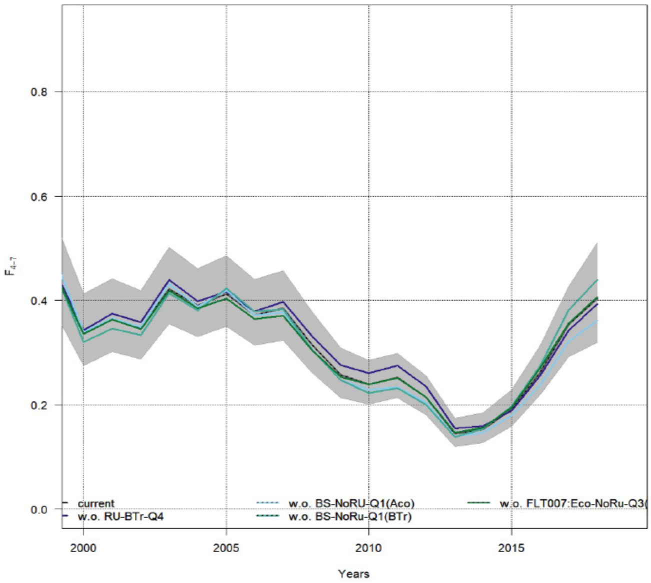


Figure 2.3b. Process errors final model.





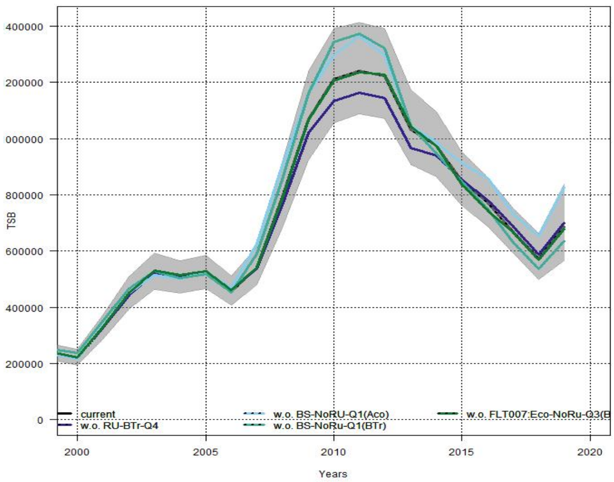


Figure 2.4c. Leave one out results, Top:SSB, middle F, bottom: TSB.

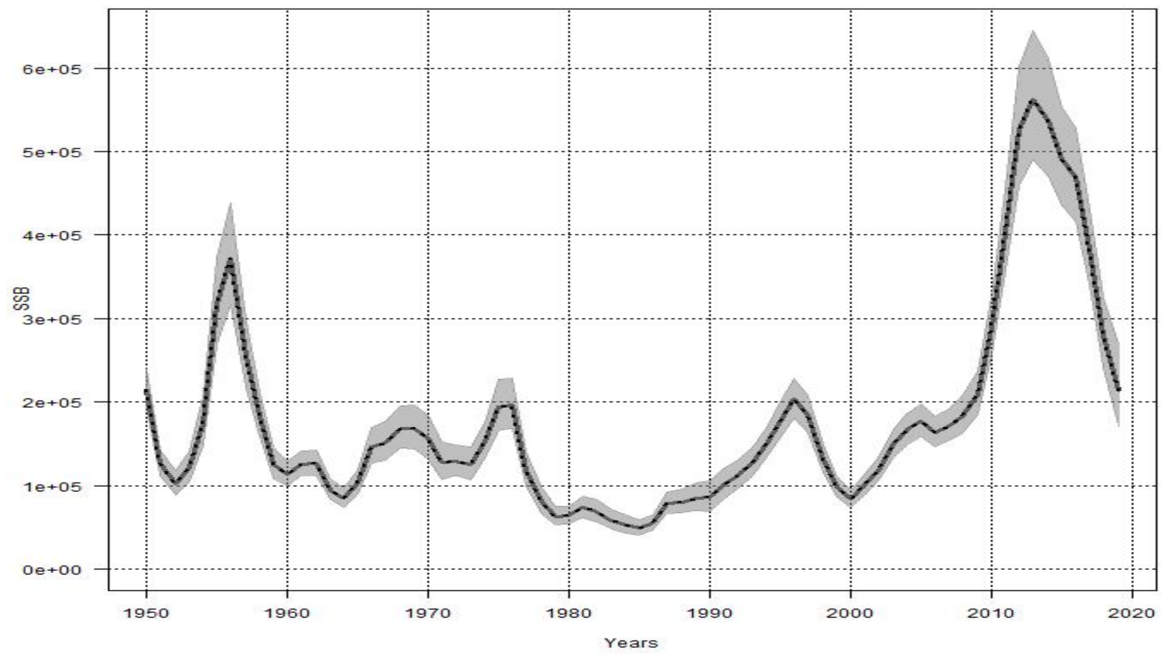


Figure 2.5a) final model result SSB.

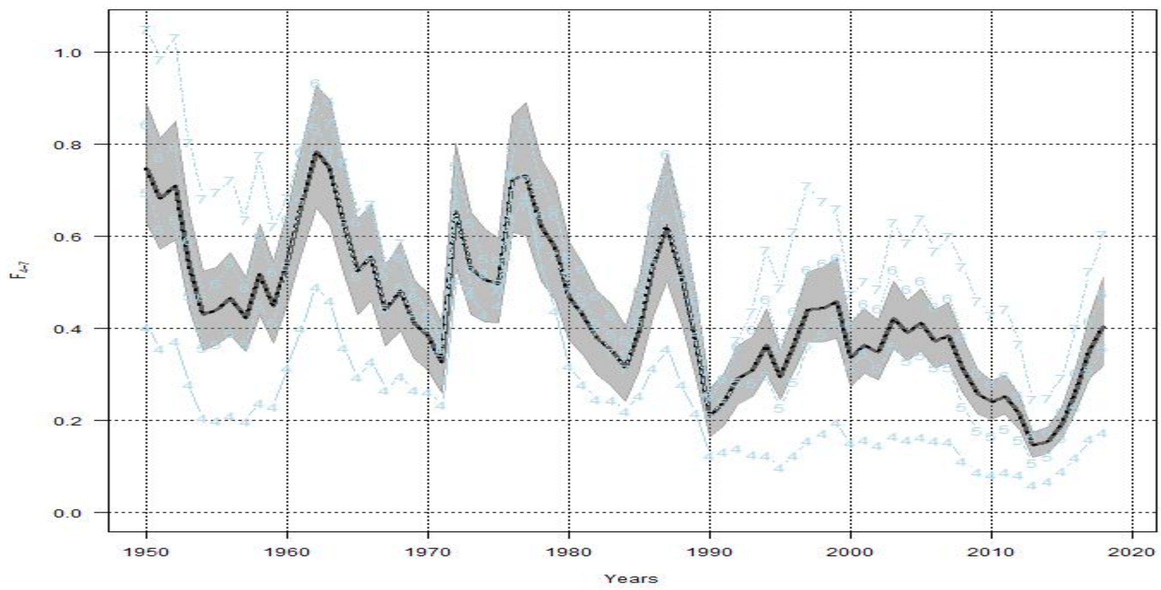


Figure 2.5b. Final model results F_{bar}. The thin lines show the Fs for age 4 and age 7.

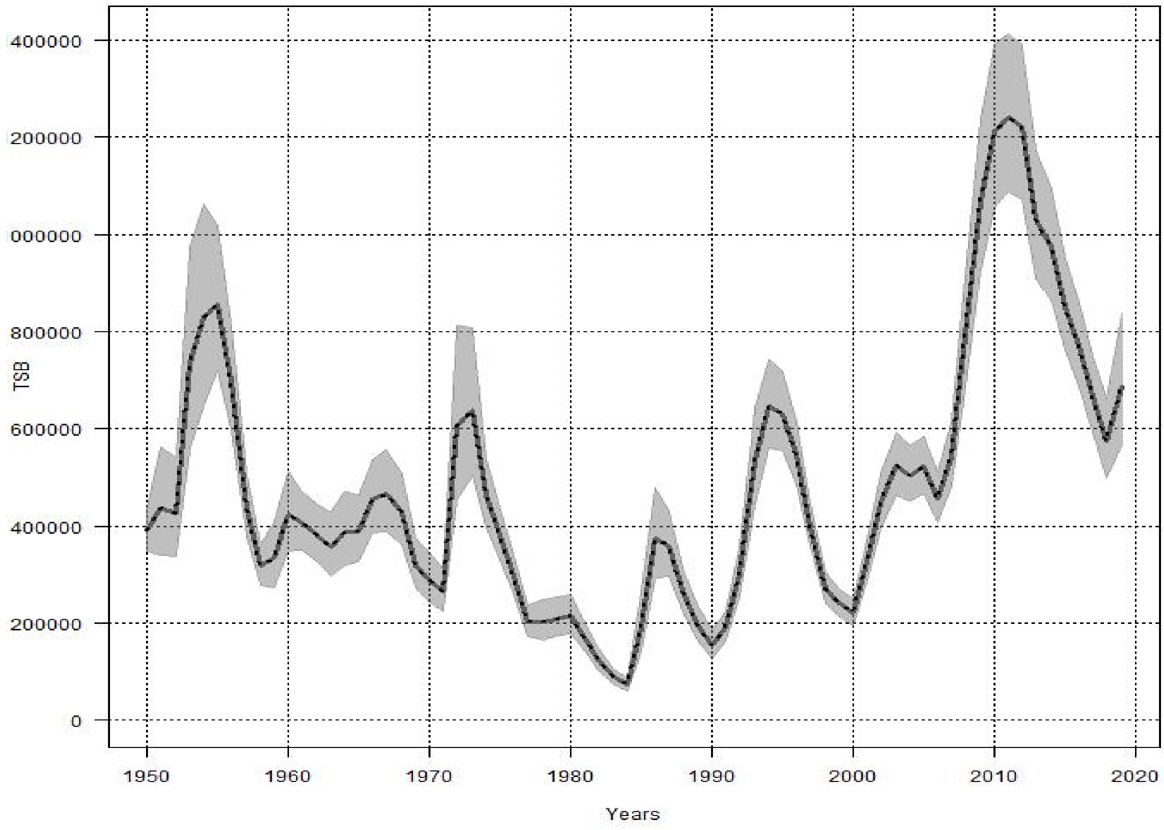


Figure 2.5c. Final model results TSB.

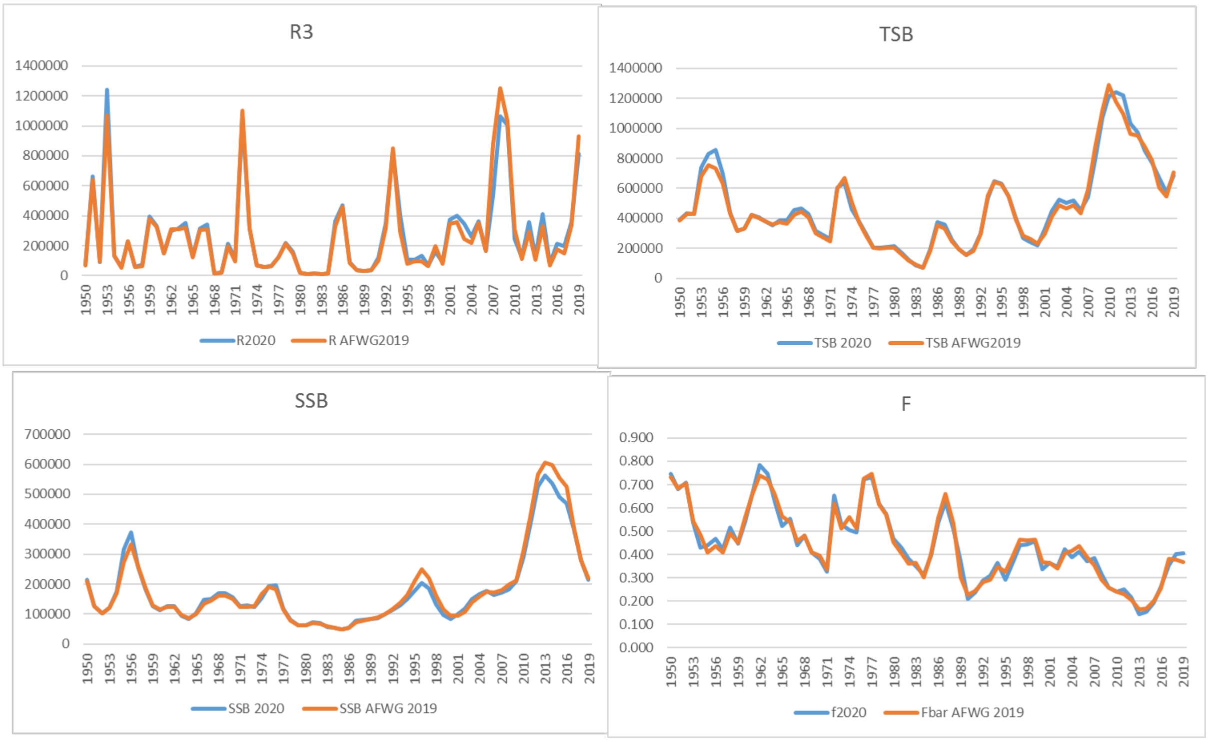


Figure 2.6. The orange line is last year assessment and blue line is the result of the final assessment model (disregard F in 2019). The difference in SSB is mainly due to a revision of the maturity data (WD5).

2.2.2 Prediction

There were some changes in the prediction models from the previous benchmark (for details see the stock annex and WDs 3.9 and 3.10):

Recruitment of age 3 in intermediate year now is from SAM estimates.

Prediction of weight-at-age 3 is now based on the average recruitment-at-age 3 in the same year and two previous years. This applies both to short- and long-term predictions.

Also, weight-at-age in catch predictions are now linked to weight-at-age in stock predictions, both for short- and long-term predictions. Previously this was only implemented for long-term predictions. The linking between catch and stock weight for the oldest age groups was also changed compared to what was previously used in long-term predictions.

Finally, the maturation model used in the long-term prediction was improved by a considerable reduction of the number of parameters.

2.2.3 Reference points

NEA haddock is classified as a Type 1 stock, a spasmodic stock with occasional large year classes (Figure 2.7). B_{lim} is then based on the lowest SSB where large recruitment is observed (ICES, 2017).

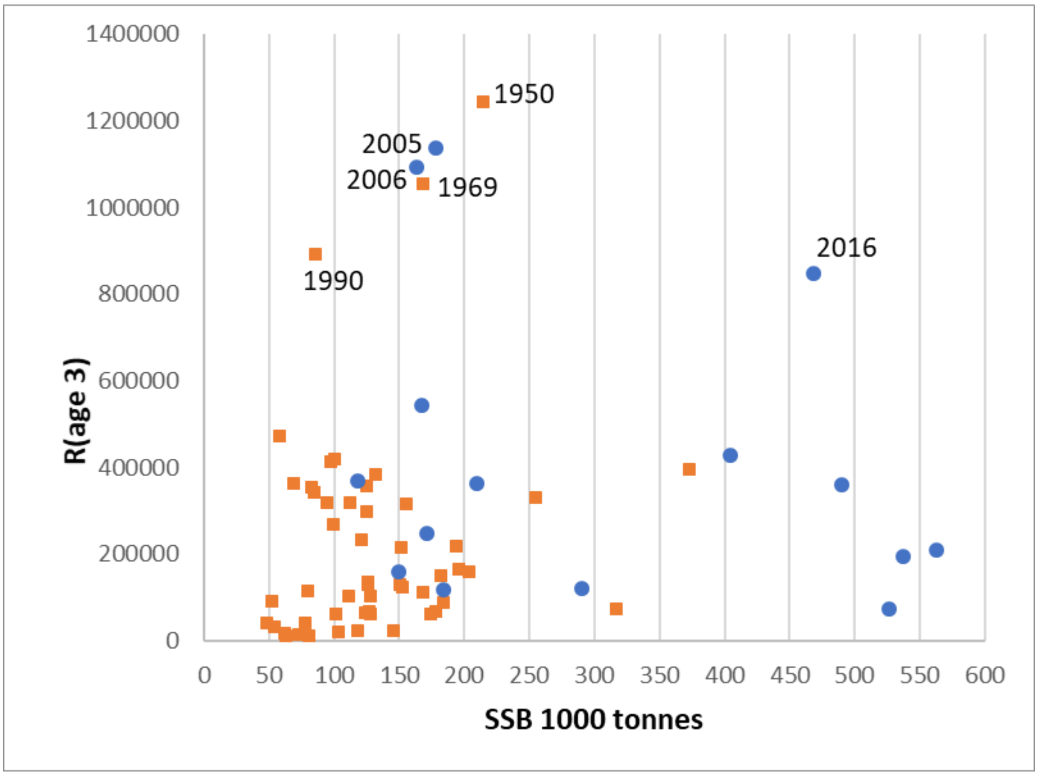


Figure 2.7. SSB recruitment plot, year classes 1950–2001 are shown in orange, the year classes 2002–2016 are shown in blue. B_{lim} is 50 000 tonnes.

Biological and fisheries reference points for NEA haddock were last set following a thorough analysis as part of the WKNEAMP-2 (ICES, 2016) Harvest Control Rule evaluation in 2016. As described above, the revised model developed during WKDEM 2020 produced better fits to the data but only a small change in the reconstructed stock. A brief analysis at WKDEM indicated that the reference points from the current model are very similar to the previously estimated values (WD 3.10). Given the more thorough analysis at WKNEAMP-2 (ICES, 2016). This is taken as indicating that there is no evidence to deviate from the existing reference points. We therefore keep the reference points unchanged, at B_{lim}=50 000 B_{pa}=80 000, F_{MSY}=0.35, F_{pa} =0.47, F_{lim}=0.77.

2.3 Investigations undertaken (summary)

2.3.1 Dataset

At the WKDEM 2020 data evaluation meeting held in December 2019, we updated weight-at-age and maturity-at-age, and included new strata covered from 2014 onwards in winter survey–a survey that two tuning series are based on (WD 3.4–3.6). This reduced the estimate of the SSB, especially for the peak years (WD8). After the data meeting, we also added a plus group to the survey data and included age 3 for the winter survey tuning series (WD 3.8). This improved the retrospective pattern (3.2 above and WD 3.8).

2.3.2 Model settings

At the meeting we tried out several SAM model configurations, intermediate steps and results plots can be found at SharePoint in the WKDEM 2020 folder: “02. Background documents”, sub-folder: “nea haddock models”.

Here we briefly summarise the steps made at the meeting to arrive at the final model. We used AIC values, model results and retrospective patterns to guide us through the various steps.

1. Observation variances

Prediction–variance link (see 2.2 above). This option was tried first for each survey and the catches separately, and then together and modified by age. The results were evaluated with Akaike Information Criterion (AIC), a reduction in AIC is considered an improvement of how well the model fits the data. We found that this option was most important for the catches (largest drop in AIC), followed by the winter survey series. For the two other surveys, the AIC did not differ, but we decided to use the option also for these surveys for consistency. The retrospective pattern was still good, and the model fitted the observed catches much better, and the AIC was much lower than when using the base configuration.

2. Correlation structure in the variance

The next step was to test for correlation structure in the variance for each survey and for the catch. First without blocking the ages, and then after looking at the result, block the ages into reasonable groups. There was evidence for a correlation structure in the surveys, and it was included with as few parameters as possible.

The AIC was greatly improved, and the retro was still good. We compared AIC and retro for several groupings. In the end we ended up with a version that seemed stable and adding or reducing groups had little effect.

3. Coupling fishing mortalities

First, we let all the fishing mortality vary by age and by looking at the plots, it was decided to try different options: first blocking age 6 and above, then 7, 8, 9 and 10.

The AIC was used to choose, and 7 and 8 had equal support. This was also tried excluding the last ten years of the data, and the results was similar. We ended up with coupling from age 8 and above, which is a slight change from the model used at AFWG 2019 which had coupling from age 9 and above. There was a moderate improvement in AIC with the new coupling, and the retro was still good.

4. Catchabilities

There are two settings related to catchabilities, and they are highly dependent. One is the catchabilities ($keyLogFpar). The other is the setting linking the indices to the size of the stock ($keyQpow). If this parameter is set to 1 or the option is not used, then a linear relationship between stock size and indices is assumed. If the parameter is >1 then the indices will increase faster than the stock, suggesting that catchability increases with stock size/density. The $keyQpow is an "exotic" parameter (A. Nielsen, Personal communication), and should be used with caution, not to overfit the model and make it unstable. On that basis we decided to keep this parameter constant by age but varying between surveys.

Then the $keyLogFpar was allowed to vary freely by age and between surveys, the results were inspected, and then the ages were grouped from the results of the free run. Different groupings were explored using AIC. The (pre-)final configuration was fairly simple.

This had a low AIC compared to the start (base model /old configuration), also a good retrospective pattern and a much better fit of the modelled catches to the observed catches.

When the result was presented to the group, two concerns were raised:

- The catchability did not vary much over age – that is the model assumed an almost constant catchability over age. This appeared unusual according to some comments.
- Furthermore, the model did not fit some very low observations of ages 8–10. This was the two indices from the winter surveys (bottom trawl and acoustic) from around 2000.

We therefore tried a new configuration, letting the catchability of the oldest ages (>7) of the two winter survey indices differ from the rest.

This model gave a lower catchability for older ages at smaller population size in comparison to the intermediate ages for the winter surveys (trawl and acoustic), while it has a higher catchability for older ages when the abundance of the stock is higher.

This model had a lower AIC compare to the (pre-)final model, and a slightly better retro. The results were similar, but this model gave a higher estimate of the peak abundance.

Even though there is a danger that this model could be more unstable, since it has more parameters for the $keyQpow, the retro seemed as good (or better than) for the (pre-)final model, and AIC was lower, so we decided to stick with this model as the final model. Note that the number of catchability/power parameters are reduced compared to the configurations decided during the 2015 benchmark, when it was included four more power parameters.

2.4 Future considerations

Alternative methods for estimating stock weight and maturity-at-age should be investigated. New methods for estimating survey indices should be explored, particularly for the ecosystem survey.

2.5 Reviewers' comments

The analytical team noted a retrospective during the vetting of the model prior to the review workshop. A plus group was added to the model configuration, which eliminated that retrospective pattern, and allowed the model to accumulate fish at the larger sizes for the surveys.

Additionally, during the analytical vetting the team noticed that the model was not tracking high catches well. A feature was added to SAM called the 'prediction-variance link'. This feature allowed for sample sizes to come into play when fitting large catches or large survey points. The analytical team demonstrated improved fits to extreme values, particularly larger catch values. For species like haddock, large intermittent year classes require substantial flexibility. During the benchmark meeting, the team looked at multiple configurations and sensitivity runs to configure the mean–variance link for the catches and surveys.

The addition of the mean–variance link improved fits to the catches. Half of the surveys had stronger mean–variance links, while the other two surveys did not. Those with the stronger link had poorer fits to the older age classes, which is because those age classes had relatively few fish and the mean–variance link allows for worse fits to those data. Work was completed in order to improve the fits to those two surveys for the older age classes.

Some additional sensitivity runs were requested, but they did not change the final decisions regarding the base run. First, the survey catchability was flat after the first age; thus, a sensitivity run was requested to address this assumption. Second, we requested a run dropping the two surveys that are in the same area and time but are acoustic and trawl-based.

The AIC values, residual fits, retrospective analyses, and biology were used to choose the final base run for the assessment. These are good practices for model selection. While the decision framework for model choice was consistent and sound; we stress caution regarding chasing noise in the data. Strong fits to lower quality data provides the model with more certain information than are actually available. The analytical team needs to be sure to balance the fits to high quality data with fits to lower quality data, regardless of AIC values.

Both the hindcast and forecast are consistent with respect to consistency in life-history information. The continuation of assumptions regarding density-dependence and time-varying life-history components from the SAM model into the projections provides the best information for projecting forward into the future. In addition, the additional work that was done to look at the robustness of the projections was sufficient to make final decisions.

2.6 References

Aanes, S. 2016. A statistical model for estimating stock parameters accounting for errors in data: Applications to data for Norwegian Spring-Spawning Herring. Working Document for WKPELA 2016.

Breivik, O. N., Nielsen A., Berg C. Prediction-variance link in state–space assessment. In prep.

ICES. 2015. Report of the Benchmark Workshop on Arctic Stocks (WKARCT), Copenhagen 26–30 January 2015. ICES C. M. 2015/ACOM:31, 126 pp.

ICES. 2016. Report of the second workshop on Management Plan Evaluation on Northeast Arctic cod and haddock and Barents Sea capelin (WKNEAMP-2), 25–28 January 2016, Kirkenes, Norway. ICES CM 2016/ACOM:47. 76 pp.

ICES. 2017. ICES technical guidelines. ICES fisheries management reference points for category 1 and 2 stocks. DOI: 10.17895/ices.pub.3036.

ICES. 2018. Report of the Arctic Fisheries Working Group, Ispra, Italy, 16–24 April 2018. ICES C.M. 2018/ACOM:06,857 pp.

ICES. 2019. Arctic Fisheries Working Group (AFWG). ICES Scientific Reports. 1:30. 930 pp. http://doi.org/10.17895/ices.pub.5292

Taylor, L. R. 1961. Aggregation, variance and the mean. Nature, 189(4766):732–735.

3 Cod (*Gadus morhua*) in Division 6.a (West of Scotland) (Cod.27.6a)

3.1 Why a benchmark?

The last benchmark for this stock was carried out in February 2012 (ICES, 2012) with subsequent inter-benchmark (IBP) in February 2015 (ICES, 2015). Analysis of the stock reported in Cook (2019a) that estimated a different trend in fishing mortality and higher biomass lead to an inter-benchmark in February 2019 (ICES, 2019a).

The uncertainties in the catch data for this stock have long been acknowledged (ICES, 1997). From the 1990s to mid-2000s there was known to be significant under-reporting of landings and total catch data have been excluded from the assessment for these years. Since the benchmark in 2012 estimates of total cod landings from 6.a (for 2006 onwards) have made use of area misreporting estimates obtained from Marine Scotland Compliance accounting for over half of the total landings. In recent years concern has been raised regarding the reliability of these estimates for the purposes of stock assessment (See WD 4.4 and ICES, 2015). Therefore, one of the issues to be addressed at this benchmark was developing and implementing a more objective approach to the estimation.

The TSA assessment of cod in Division 6.a (Gudmundson, 1994; Fryer, 2001; 2011) agreed at the IBP in 2019 was found to be robust to assumptions about fishery selectivity, survey catchability, the time-series of data included and the relative weight of survey and landings data. However, following that IBP, there remained significant uncertainties. As a result of the particular uncertainties in the input data (both survey indices and commercial data) for this stock, it appears that these data can be interpreted in different ways by different assessment methods (ICES, 2019a; Cook, 2019a). The differences in the perception of stock trends appear to be associated largely with differences in the estimated fishery selection pattern. Both the TSA model and the Cook model estimated a large reduction in fishing mortality, although there are differences in the timing and magnitude of decline. The TSA estimated a decline in mean F of around 50%, which is commensurate with (although occurs several years later than) the decline in reported effort from the main fleets operating in the fishery (STECF effort data). Trends in fishing mortality estimated by the Cook model are also reflected in the large reductions in effort resulting from the decommissioning of active vessels and fishing mortality trends in other geographically similar stocks (whiting in 6a, haddock in 46a and cod in 347d) although the decline in estimated fishing mortality in this model begins before the major decommissioning schemes occurred.

Furthermore, the fishing mortality estimated by the Cook model is much lower than in the TSA and is also substantially lower than in other cod stocks around the British Isles (cod in 347d and cod 27.7e–k). In addition to the differences in fishing mortality, the TSA model estimated recent stock biomass as significantly below the 7600 tonnes of cod consumed by seals (in 2010/2011) as estimated by the Sea Mammal Research Unit (Hammond and Wilson, 2016). In contrast, the Cook model cod biomass estimate is double the seal consumption for the same year. There is debate about the degree of overlap between the cod populations exploited by the fishery (Russell *et al.*, 2017) and by seals and it remains a major source of uncertainty.

One of the recommendations from the IBP was the exploration of alternative assessment methods including SAM (Nielsen and Berg, 2014), a4a (Jardim *et al.*, 2015) and the model described in Cook (2019a) with the aim of identifying the reasons for these different model interpretations.

Stock structure remains an issue for cod in Division 6.a. The latest evidence (WD 4.1) suggests that there are at least three substocks which remain largely geographically isolated throughout the year with the northern offshore component (currently responsible for the majority of the landings) more closely linked to cod in the northern North Sea than the rest of Division 6.a. Given the current lack of catch data disaggregated at the appropriate spatial scale, the issue of multiple stocks has not been addressed at this benchmark. The current assessment therefore remains an assessment of multiple substocks.

3.2 Summary of final model

3.2.1 Catch data

One of the main issues to address as part of this benchmark was an improved approach to estimating area-misreported landings by the Scottish demersal fleet; that is, landings which are taken in Division 6.a but declared as being caught in other areas (most commonly Division 4.a). Previously, the estimates have been provided by Marine Scotland Compliance, based on intelligence and expert judgement. The approach agreed at this benchmark is a more objective approach based on an analysis of VMS data linked to daily logbook landings and documented in WD 4.4. The revised estimates of area-misreported landings are provided for 2006 onwards and given in Table 3.2.1.

A data call for national landings and discards age compositions was issued for this stock ahead of the benchmark. Revisions were provided for 2003 onwards. Following an analysis of Scottish catch sampling data, it was agreed that for the purposes of allocated age compositions and discard rates, the area-misreported landings should be treated as part of the Scottish demersal trawl fleet. A full description of available data and the assumptions applied in the preparation of the total landings and discards age compositions in InterCatch is provided in WD 4.5.

The final estimates of total landings and discards and associated age compositions and mean weights are given in Tables 3.2.1 to 3.2.7.

3.2.2 Biological parameters

Maturity and natural mortality were also modified as part of this benchmark process. The maturity ogive had previously not been updated and the basis was unknown. An analysis of Scottish survey data (following the approach advocated by ICES, 2008) indicated a proportion of individuals at-age 1 to be mature, but no temporal trend in maturity. A new ogive was therefore used for the full time-series (WD 4.2).

Age	1	2	3	4	5	6	7+
WGCSE 2019	0	0.52	0.86	1.0	1.0	1.0	1.0
WKDEM	0.27	0.53	0.48	0.91	0.97	0.99	1.0

Previously, a time-invariant natural mortality, dependent on weight (mean weight-at-age over the time period of the assessment data) had been used (Lorenzen, 1996). Given the trends in observed mean weights, it was agreed that a temporally varying natural mortality would be more appropriate. It was agreed to use smoothed catch weights as stock weights, and then use these with the Lorenzen (1996) function with the 'natural ecosystems' parameters to obtain natural mortality (WD 4.3). Updated stock weights and natural mortality are given in Tables 3.2.8 and 3.2.9.

3.2.3 Assessment model

Although the data from a number of additional surveys were worked up ahead of this benchmark, the resulting biomass indices are very short with extremely wide confidence intervals (over much of the time-series) and therefore these were not considered further for inclusion in the stock assessment (WD 4.6 and WD 4.7). The survey data included in the final assessment are therefore consistent (in terms of both year range and age range) with those used at the last assessment WG meeting (ICES, 2019b). There is reasonable internal consistency in the early surveys and some limited between survey consistency (between Irish and Scottish surveys) for certain age groups. On the whole however, the survey indices are noisy, particularly in the recent time period. Exploratory analysis of survey data are presented in detail elsewhere (ICES, 2019a, b) and not discussed further in this report.

A summary of the assessment input data is provided in the table below:

Data TYPE	Year RANGE	AGE RANGE	NOTE
Catch numbers-at-age	1981–2018	1–7+	1995–2006: uses age compositions only and estimates an annual scaling factor on total catch
Catch weights-at-age	1981–2018	1–7+	
WCIBTS.Q1 (survey)	1985–2010	1–6	
SCO.Q1 (survey)	2011–2019	1–6	
WCIBTS.Q4 (survey)	1996–2009	1–4	
SCO.Q4 (survey)	2011–2018	1–6	Data excluded in2013 due to vessel breakdown
IRGFS.Q4 (survey)	2003–2018	1–3	
Proportion mature at-age	Fixed ogive over all assessment years		
Natural mortality	Time varying derived from mean stock weight-at-age (which are modelled mean catch weights-at-age) & Lorenzen (1996)		

The final SAM model configuration was chosen by consideration a combination of model residuals, AIC and retrospective patterns. The configuration file is given in Table 3.2.10. To summarise the main features:

- Fishing mortality at ages 4 and above are assumed equal (See # Coupling of the fishing mortality states, Table 3.2.10).
- Survey catchabilities are mostly freely estimated for each age with the exception of the two oldest ages (i.e. no survey catchability plateau assumed). The exception to this is the WIBTS.Q1 for which all catchabilities are independently estimated.
- Catch observation variance parameters are allowed to differ for age 1 and age 7+ while other age groups are coupled (# Coupling of the variance parameters for the observations). To allow for greater uncertainty in the catch data for 2006 onwards (when the fishery changes from being a landings fishery to largely discards), the estimated catch observation error standard deviation is doubled for 2006 onwards (based on inspection of the one step ahead residuals).
- Survey observation variance parameters differ between surveys but are coupled for all age groups within a survey.
- Recruitment is modelled as a random walk.
- A catch scaling factor is estimated for 1995–2006 when underreporting of landings was considered significant.
- Fishing mortality across ages is modelled with AR(1) and process variance parameters coupled across all ages with the exception of age 1. Process variance in stock numbers-at-age were assumed coupled with the exception of age 1 (the age at recruitment).

The final assessment results are shown in Figure 3.2.1 (in comparison with a TSA assessment using the same data, configured as per WGCSE 2019) and the estimated model parameters from the assessment are given in Table 3.2.11. The estimated mean F increases through the time-series until the mid-2000s, declines from 2010 onwards with a subsequent increase in recent years. Following the decline in estimated SSB in the 1980s and 1990s, there is a small increase in SSB between 2006 and 2016 (although this declines again in recent years). The main difference between the SAM and the TSA assessment appears to be in the estimates of the catch scaling factor during the period when under-reporting of landings was believed to have occurred (1995–2006). During this period, the SAM assessment estimates much lower total catches (closer to the actual data) than the TSA assessment although the assessments are very similar since 2006. The other clear difference between the SAM and TSA assessments is the much smoother trend in estimated mean F with the decline to 2016 occurring more gradually in SAM.

The standardised one step ahead residuals are shown in Figure 3.2.2. The model fit to the catch data looks reasonable with no obvious patterns or significant outliers in the residuals (most lying within ±2). With the exception of changes in the assumptions/coupling of the catch observation variance, other model runs conducted as part of the sensitivity testing showed little observable impact on the quality of the fit of the model to the catch-at-age data (in terms of inspection of residuals). There are a few patterns apparent in the survey residuals which are rather similar to those previously observed in TSA assessments (ICES, 2019a and b): most notably some evidence of a tendency to more positive residuals in the latter half of the WIBTS.Q1 (at-age 1) and WIBTS.Q4 (at-age 2) and some year effects in most of the surveys (years with mostly positive or mostly negative residuals). The sensitivity analysis in which survey catchability is forced to be more flat-topped (with freed-up fishery selectivity) results in generally poorer survey residuals (i.e. more patterns), particularly in WIBTS.Q1, SCO.Q1 and SCO.Q4, with a preponderance of negative residuals at-age 3 and positive residuals at-ages 5 and 6 (over the whole time-series) and having a limited impact on the observed year effects. The standardized residuals from the exploratory TVM assessment (see Section 3.3) suggest that there are similar (to both TSA and

SAM) systematic issues with lack of fit at particular years/ages in the survey data and additionally show bias in fitting the discard data and increasing landings residuals across a number of age classes.

The retrospective analysis is also shown in Figure 3.2.3. Although the Mohn's rho value is relatively small, there is a tendency to over-estimate F when the underlying fishing mortality starts declining, although this is not consistent over all retrospective peels. The estimates of mean F appear to be somewhat noisier than either recruitment or SSB. The Mohn's rho values (as %) are as follows:

SSB	Mean F	Recruitment
-9.5	8.5	-2.8

The model runs which leave out each survey index in turn are shown in Figure 3.2.4. With the exception of the period when total catches are excluded from the assessment (catch-scaling factor estimated for 1995–2006), the estimates of SSB and recruitment are robust to the exclusion of the different survey series. Excluding the early Scottish Q4 survey (WCIBTS.Q4) results in higher estimates of SSB, recruitment and catch than the baseline run during this period (when catches area excluded) and excluding the early Scottish Q1 survey much lower estimates. When the WCIBTS.Q4 is excluded, estimates of mean F are lower than the baseline during the first part of this period (to 2000) and higher than the baseline after 2000 while excluding the WCIBTS.Q1 shows the opposite effect. The relative magnitude of the changes when each of these surveys are excluded suggests the WCIBTS.Q1 to be much more influential in the overall assessment of stock trends. There are small differences in estimated mean F in recent years depending when either the SCO.Q1 or SCO.Q4 survey series are excluding while excluding the Irish survey index appears to have little impact on the assessment results.

3.2.4 Reference points

The final agreed reference points (compared to the previous reference points) derived following the ICES guidelines and using EqSim are given below. Further details of their derivation can be found in Section 3.3 and in WD 4.9.

Reference Point	Value (previous value in brackets)	Technical Basis
F_{MSY}	0.30 (0.29)	F that provides maximum yield (calculated from EqSim using Segmented regression stock–recruitment relationships including full time-series of stock–recruit data)
MSY $B_{trigger}$	20 126 t (20 000 t)	B_{PA}
B_{lim}	14 376 t (14 000t)	SSB avoiding low recruitment (SSB in 1992 as estimated by WKDEM)
B_{PA}	20 126 t (20 000 t)	B_{lim} x 1.4
F_{lim}	0.73 (0.77)	Based on simulation using segmented regression with B_{lim} as the breakpoint (EqSim): F such that 50% probability of SSB <B_{lim}
F_{PA}	0.52 (0.55)	F_{lim}/1.4
$F_{MSY\ lower}$ (without ICES AR)	0.18 (0.20)	F at 95% MSY (below F_{MSY})
$F_{MSY\ upper}$ (without ICES AR)	0.49 (0.41)	F at 95 % MSY (above F_{MSY})
$F_{p.05}$ (with ICES AR)	0.57 (0.64)	F that gives a 5 % probability of SSB <B_{lim} when the ICES advice rule is applied

Table 3.2.1. Cod.27.6a. Annual reported landings (international), estimated area-misreported landings (Scottish), total estimated discards and catch (all in tonnes) as estimated by WKDEM. No revisions were made to pre-2003 data.

Year	Reported landings	Misreported landings	Discards	Total Catch
2003	1292.3		60.4	1352.7
2004	572.8		77.9	650.7
2005	516.1		54.3	570.4
2006	469.8	34.2	461.0	930.9
2007	485.1	29.6	1650.9	2136.0
2008	459.7	101.6	1036.8	1496.5
2009	230.8	53.5	1287.4	1518.2
2010	239.3	118.8	1574.7	1814.0
2011	211.4	129.7	3866.5	4077.9
2012	162.1	64.5	1914.2	2076.3
2013	172.4	93.4	1870.4	2042.8
2014	160.6	233.8	3369.2	3529.8
2015	258.4	269.8	2498.2	2756.7
2016	336.3	272.3	1499.4	1835.7
2017	355.2	320.2	3519.4	3874.6
2018	377.8	612.6	2428.6	2806.4

Table 3.2.2. Cod.27.6a. Landings numbers-at-age (thousands). No revisions were made to pre-2003 data.

	1	2	3	4	5	6	7+
2003	53	487	93	120	7	2	2
2004	45	99	90	12	27	3	1
2005	37	124	46	40	7	6	0
2006	18	97	78	23	14	2	1
2007	7	170	53	28	2	3	2
2008	0	20	106	21	13	1	2
2009	1	9	10	40	6	1	0
2010	6	80	26	20	11	1	1
2011	0	29	51	18	4	6	1
2012	1	1	18	24	3	2	2
2013	0	8	7	39	9	2	1
2014	0	5	73	34	25	2	0
2015	0	44	40	29	21	19	1
2016	1	17	82	52	17	9	11
2017	0	13	52	47	46	13	3
2018	2	10	28	78	51	32	11

Table 3.2.3. Cod.27.6a. Discards numbers-at-age (thousands). No revisions were made to pre-2003 data.

	1	2	3	4	5	6	7+
2003	124	27	7	0	0	0	0
2004	238	23	0	0	0	0	0
2005	127	22	0	0	0	0	0
2006	1058	45	25	2	3	1	0
2007	283	1321	46	35	2	3	0
2008	64	151	416	3	1	0	0
2009	590	157	116	146	8	7	0
2010	410	810	150	17	7	0	0
2011	303	579	1255	102	1	4	0
2012	1029	180	605	78	0	0	0
2013	2175	346	220	167	24	0	3
2014	913	948	644	116	45	2	0
2015	264	571	620	72	18	2	0
2016	1253	377	189	94	13	0	0
2017	240	429	912	223	43	5	0
2018	87	447	206	300	54	18	6

Table 3.2.4. Cod.27.6a. Landings mean weights-at-age (kg). No revisions were made to pre-2003 data.

	1	2	3	4	5	6	7+
2003	0.659	1.046	2.272	3.820	5.932	8.022	8.681
2004	0.605	1.026	2.191	4.398	6.033	8.242	9.840
2005	0.750	1.109	2.425	3.969	4.775	6.616	10.214
2006	0.659	1.176	2.239	3.813	6.160	7.759	11.041
2007	0.728	1.127	2.592	4.322	6.503	7.738	8.830
2008	0.556	1.157	3.067	4.843	6.283	7.964	8.487
2009	0.974	2.038	2.861	4.781	6.004	8.327	9.137
2010	0.936	1.468	2.918	4.064	5.785	9.158	10.275
2011	NA	1.804	2.811	4.510	5.842	6.528	9.837
2012	0.661	1.797	3.118	5.331	6.428	7.617	8.695
2013	0.957	1.368	2.933	4.075	6.135	7.144	9.842
2014	1.028	1.600	2.097	3.051	4.693	5.503	7.207
2015	0.914	2.406	2.958	3.844	5.455	5.558	9.158
2016	0.713	1.429	2.367	3.917	5.137	6.596	7.622
2017	0.902	1.229	2.063	4.533	5.616	5.081	9.243
2018	0.871	1.686	2.761	4.163	5.427	6.427	8.575

Table 3.2.5. Cod.27.6a. Discards mean weights-at-age (kg). No revisions were made to pre-2003 data.

	1	2	3	4	5	6	7+
2003	0.311	0.600	0.388	NA	NA	NA	NA
2004	0.261	0.576	NA	NA	NA	NA	NA
2005	0.242	0.483	0.803	NA	NA	NA	NA
2006	0.276	1.346	2.786	3.501	6.242	5.581	11.151
2007	0.196	0.948	3.014	4.457	4.985	10.635	NA
2008	0.224	0.999	2.049	3.853	5.216	NA	NA
2009	0.264	1.333	2.296	3.834	6.051	6.985	9.119
2010	0.273	1.274	2.268	3.218	3.245	NA	NA
2011	0.266	1.072	2.213	2.993	4.891	4.168	NA
2012	0.142	1.118	2.179	3.222	NA	NA	NA
2013	0.125	1.155	2.110	3.050	5.029	0.000	6.269
2014	0.150	1.210	2.390	3.066	3.998	4.349	NA
2015	0.404	1.063	2.330	3.428	4.414	6.103	NA
2016	0.205	1.096	2.212	3.759	4.435	NA	NA
2017	0.262	1.048	2.183	3.473	4.397	7.714	NA
2018	0.217	1.046	2.219	3.649	5.300	4.980	2.117

Table 3.2.6. Cod.27.6a. Catch numbers-at-age (thousands). No revisions were made to pre-2003 data.

	1	2	3	4	5	6	7+
2003	176	514	100	120	7	2	2
2004	282	122	90	12	27	3	1
2005	163	146	46	40	7	6	0
2006	1076	143	104	25	17	3	1
2007	290	1492	100	64	5	6	2
2008	64	171	522	24	15	1	2
2009	591	166	126	186	14	8	1
2010	416	889	175	37	17	1	1
2011	303	608	1307	120	5	10	1
2012	1030	181	623	101	3	2	2
2013	2175	355	228	206	33	2	4
2014	913	953	717	149	70	4	0
2015	264	615	660	102	39	21	1
2016	1254	394	271	146	30	9	11
2017	240	442	963	270	89	18	3
2018	88	457	235	378	105	49	16

Table 3.2.7. Cod.27.6a. Catch mean weights-at-age. No revisions were made to pre-2003 data.

	1	2	3	4	5	6	7+
2003	0.415	1.023	2.140	3.820	5.932	8.022	8.681
2004	0.316	0.943	2.191	4.398	6.033	8.242	9.840
2005	0.356	1.014	2.425	3.969	4.775	6.616	10.214
2006	0.282	1.230	2.373	3.789	6.175	7.002	11.046
2007	0.209	0.969	2.788	4.397	5.726	9.174	8.830
2008	0.224	1.018	2.256	4.715	6.189	7.964	8.487
2009	0.266	1.372	2.342	4.039	6.030	7.222	9.111
2010	0.282	1.291	2.363	3.683	4.784	9.158	10.275
2011	0.266	1.107	2.237	3.221	5.722	5.507	9.837
2012	0.142	1.120	2.205	3.713	6.428	7.617	8.695
2013	0.125	1.160	2.137	3.243	5.336	7.144	7.145
2014	0.150	1.212	2.360	3.063	4.245	4.984	7.207
2015	0.405	1.159	2.368	3.548	4.964	5.612	9.158
2016	0.206	1.110	2.259	3.815	4.834	6.596	7.622
2017	0.263	1.053	2.177	3.656	5.032	5.746	9.243
2018	0.229	1.060	2.285	3.755	5.362	5.909	6.304

Table 3.2.8. Cod.27.6a. Stock mean weights-at-age (kg) derived as smoothed catch weights-at-age.

	1	2	3	4	5	6	7+
1981	0.511	1.304	2.942	4.971	6.859	8.292	9.896
1982	0.503	1.293	2.918	4.943	6.831	8.281	9.899
1983	0.495	1.282	2.894	4.915	6.804	8.27	9.9
1984	0.486	1.271	2.87	4.887	6.776	8.26	9.9
1985	0.478	1.26	2.847	4.86	6.749	8.25	9.899
1986	0.469	1.249	2.823	4.833	6.722	8.241	9.895
1987	0.461	1.238	2.8	4.806	6.695	8.231	9.889
1988	0.453	1.228	2.777	4.779	6.667	8.221	9.88
1989	0.444	1.217	2.754	4.753	6.638	8.211	9.868
1990	0.436	1.206	2.731	4.726	6.608	8.2	9.853
1991	0.428	1.196	2.708	4.7	6.577	8.188	9.834
1992	0.419	1.186	2.685	4.673	6.544	8.174	9.813
1993	0.411	1.175	2.663	4.646	6.51	8.158	9.789
1994	0.402	1.165	2.641	4.617	6.473	8.139	9.762
1995	0.394	1.155	2.618	4.587	6.433	8.118	9.734
1996	0.386	1.146	2.596	4.555	6.392	8.093	9.703
1997	0.377	1.137	2.574	4.521	6.347	8.063	9.671
1998	0.369	1.129	2.553	4.484	6.3	8.03	9.638
1999	0.361	1.122	2.531	4.444	6.249	7.991	9.602
2000	0.352	1.115	2.51	4.402	6.196	7.947	9.564
2001	0.344	1.111	2.489	4.357	6.14	7.898	9.523
2002	0.335	1.107	2.469	4.31	6.082	7.844	9.478
2003	0.327	1.105	2.449	4.261	6.022	7.783	9.429
2004	0.319	1.103	2.429	4.21	5.961	7.716	9.374
2005	0.31	1.103	2.41	4.158	5.898	7.642	9.311
2006	0.302	1.104	2.391	4.104	5.833	7.559	9.242
2007	0.294	1.106	2.372	4.049	5.768	7.466	9.164
2008	0.285	1.108	2.353	3.993	5.701	7.362	9.078
2009	0.277	1.11	2.335	3.936	5.633	7.247	8.984

	1	2	3	4	5	6	7+
2010	0.268	1.112	2.316	3.879	5.564	7.121	8.882
2011	0.26	1.115	2.298	3.822	5.494	6.986	8.773
2012	0.252	1.116	2.28	3.766	5.423	6.842	8.658
2013	0.243	1.118	2.262	3.71	5.351	6.692	8.538
2014	0.235	1.119	2.245	3.656	5.279	6.539	8.414
2015	0.227	1.12	2.227	3.602	5.207	6.384	8.289
2016	0.218	1.121	2.209	3.549	5.136	6.228	8.161
2017	0.21	1.121	2.191	3.497	5.065	6.073	8.033
2018	0.202	1.122	2.174	3.445	4.993	5.917	7.904

Table 3.2.9. Cod.27.6a. Natural mortality-at-age.

	1	2	3	4	5	6	7+
1981	0.492	0.375	0.296	0.254	0.232	0.219	0.208
1982	0.494	0.376	0.297	0.255	0.232	0.219	0.208
1983	0.496	0.377	0.297	0.255	0.232	0.219	0.208
1984	0.499	0.378	0.298	0.255	0.232	0.219	0.208
1985	0.501	0.378	0.299	0.256	0.233	0.219	0.208
1986	0.504	0.379	0.300	0.256	0.233	0.220	0.208
1987	0.507	0.380	0.300	0.257	0.233	0.220	0.208
1988	0.509	0.381	0.301	0.257	0.233	0.220	0.208
1989	0.512	0.382	0.302	0.258	0.234	0.220	0.208
1990	0.515	0.383	0.302	0.258	0.234	0.220	0.208
1991	0.518	0.384	0.303	0.258	0.234	0.220	0.209
1992	0.521	0.385	0.304	0.259	0.235	0.220	0.209
1993	0.524	0.386	0.305	0.259	0.235	0.220	0.209
1994	0.527	0.387	0.305	0.260	0.235	0.220	0.209
1995	0.530	0.388	0.306	0.260	0.236	0.220	0.209
1996	0.533	0.389	0.307	0.261	0.236	0.221	0.209
1997	0.537	0.390	0.308	0.261	0.237	0.221	0.210
1998	0.540	0.391	0.308	0.262	0.237	0.221	0.210
1999	0.544	0.391	0.309	0.263	0.238	0.221	0.210
2000	0.548	0.392	0.310	0.263	0.238	0.222	0.210
2001	0.552	0.393	0.311	0.264	0.239	0.222	0.211
2002	0.555	0.393	0.311	0.265	0.240	0.223	0.211
2003	0.560	0.393	0.312	0.266	0.240	0.223	0.211
2004	0.564	0.393	0.313	0.267	0.241	0.224	0.211
2005	0.568	0.393	0.314	0.268	0.242	0.224	0.212
2006	0.573	0.393	0.314	0.269	0.243	0.225	0.212
2007	0.577	0.393	0.315	0.270	0.243	0.226	0.213
2008	0.582	0.393	0.316	0.271	0.244	0.227	0.213
2009	0.587	0.393	0.316	0.272	0.245	0.228	0.214

	1	2	3	4	5	6	7+
2010	0.593	0.392	0.317	0.273	0.246	0.229	0.215
2011	0.598	0.392	0.318	0.274	0.247	0.230	0.216
2012	0.604	0.392	0.319	0.276	0.248	0.232	0.216
2013	0.610	0.392	0.319	0.277	0.249	0.233	0.217
2014	0.616	0.392	0.320	0.278	0.250	0.235	0.218
2015	0.622	0.392	0.321	0.279	0.251	0.236	0.219
2016	0.629	0.392	0.322	0.280	0.252	0.238	0.220
2017	0.636	0.391	0.322	0.281	0.253	0.240	0.221
2018	0.644	0.391	0.323	0.283	0.254	0.242	0.222

Table 3.2.10. Cod.27.6a. SAM model configuration file used in the final assessment run.

```
 # Where a matrix is specified rows corresponds to fleets and columns to ages.
# Same number indicates same parameter used
# Numbers (integers) starts from zero and must be consecutive
#
$minAge
# The minimium age class in the assessment
 1

$maxAge
# The maximum age class in the assessment
 7

$maxAgePlusGroup
# Is last age group considered a plus group for each fleet (1 yes, or 0 no).
 1 0 0 0 0 0

$keyLogFsta
# Coupling of the fishing mortality states (nomally only first row is used).
   0  1  2  3  3  3  3
  -1 -1 -1 -1 -1 -1 -1
  -1 -1 -1 -1 -1 -1 -1
  -1 -1 -1 -1 -1 -1 -1
  -1 -1 -1 -1 -1 -1 -1
  -1 -1 -1 -1 -1 -1 -1

$corFlag
# Correlation of fishing mortality across ages (0 independent, 1 compound symmetry, 2 AR(1), 3 separable AR(1).
 2

$keyLogFpar
# Coupling of the survey catchability parameters (normally first row is not used, as that is covered by fishing mortality).
  -1 -1 -1 -1 -1 -1 -1
   0  1  2  3  4  5 -1
   6  7  7 -1 -1 -1 -1
   8  9 10 10 -1 -1 -1
  11 12 13 14 15 15 -1
  16 17 18 19 20 20 -1

$keyQpow
# Density dependent catchability power parameters (if any).
  -1 -1 -1 -1 -1 -1 -1
  -1 -1 -1 -1 -1 -1 -1
  -1 -1 -1 -1 -1 -1 -1
  -1 -1 -1 -1 -1 -1 -1
  -1 -1 -1 -1 -1 -1 -1
  -1 -1 -1 -1 -1 -1 -1

$keyVarF
```

```
# Coupling of process variance parameters for log(F)-process (nomally only first row is used)
  0  1  1  1  1  1  1
 -1 -1 -1 -1 -1 -1 -1
 -1 -1 -1 -1 -1 -1 -1
 -1 -1 -1 -1 -1 -1 -1
 -1 -1 -1 -1 -1 -1 -1
 -1 -1 -1 -1 -1 -1 -1

$keyVarLogN
# Coupling of process variance parameters for log(N)-process
 0 1 1 1 1 1 1

$keyVarObs
# Coupling of the variance parameters for the observations.
  0  1  1  1  1  1  2
  3  3  3  3  3  3 -1
  4  4  4 -1 -1 -1 -1
  5  5  5  5 -1 -1 -1
  6  6  6  6  6  6 -1
  7  7  7  7  7  7 -1

$obsCorStruct
# Covariance structure for each fleet ("ID" independent, "AR" AR(1), or "US" for unstructured). |
Possible values are: "ID" "AR" "US"
 "ID" "ID" "ID" "ID" "ID" "ID"

$keyCorObs
# Coupling of correlation parameters can only be specified if the AR(1) structure is chosen above.
# NA's indicate where correlation parameters can be specified (-1 where they cannot).
#1-2 2-3 3-4 4-5 5-6 6-7
 NA  NA  NA  NA  NA  NA
 NA  NA  NA  NA  NA  -1
 NA  NA  -1  -1  -1  -1
 NA  NA  NA  -1  -1  -1
 NA  NA  NA  NA  NA  -1
 NA  NA  NA  NA  NA  -1

$stockRecruitmentModelCode
# Stock recruitment code (0 for plain random walk, 1 for Ricker, 2 for Beverton-Holt, and 3 piece-
wise constant).
 0

$noScaledYears
# Number of years where catch scaling is applied.
 12

$keyScaledYears
# A vector of the years where catch scaling is applied.
 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006

$keyParScaledYA
```

A matrix specifying the couplings of scale parameters (nrow = no scaled years, ncols = no ages).

0	0	0	0	0	0	0
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
10	10	10	10	10	10	10
11	11	11	11	11	11	11

$fbarRange
lowest and higest age included in Fbar
2 5

$keyBiomassTreat
To be defined only if a biomass survey is used (0 SSB index, 1 catch index, 2 FSB index, 3 total catch, 4 total landings and 5 TSB index).
-1 -1 -1 -1 -1 -1

$obsLikelihoodFlag
Option for observational likelihood | Possible values are: "LN" "ALN"
"LN" "LN" "LN" "LN" "LN" "LN"

$fixVarToWeight
If weight attribute is supplied for observations this option sets the treatment (0 relative weight, 1 fix variance to weight).
0

$fracMixF
The fraction of t(3) distribution used in logF increment distribution
0

$fracMixN
The fraction of t(3) distribution used in logN increment distribution
0

$fracMixObs
A vector with same length as number of fleets, where each element is the fraction of t(3) distribution used in the distribution of that fleet
0 0 0 0 0 0

$constRecBreaks
Vector of break years between which recruitment is at constant level. The break year is included in the left interval. (This option is only used in combination with stock-recruitment code 3)

Table 3.2.11. Cod.27.6a. SAM estimated model parameters.

	par	sd(par)	exp(par)	Low	High
logFpar_0	-9.82198	0.15854	0.00005	0.00004	0.00007
logFpar_1	-7.97051	0.15181	0.00035	0.00026	0.00047
logFpar_2	-7.08696	0.15184	0.00084	0.00062	0.00113
logFpar_3	-6.63981	0.15400	0.00131	0.00096	0.00178
logFpar_4	-6.17857	0.16600	0.00207	0.00149	0.00289
logFpar_5	-5.77270	0.17501	0.00311	0.00219	0.00442
logFpar_6	-10.99971	0.21230	0.00002	0.00001	0.00003
logFpar_7	-11.34045	0.17027	0.00001	0.00001	0.00002
logFpar_8	-8.21823	0.23195	0.00027	0.00017	0.00043
logFpar_9	-7.16892	0.23229	0.00077	0.00048	0.00123
logFpar_10	-6.85880	0.18424	0.00105	0.00073	0.00152
logFpar_11	-8.67263	0.32628	0.00017	0.00009	0.00033
logFpar_12	-6.28518	0.31150	0.00186	0.00100	0.00347
logFpar_13	-5.73623	0.31058	0.00323	0.00173	0.00601
logFpar_14	-5.21387	0.31091	0.00544	0.00292	0.01013
logFpar_15	-4.27832	0.24428	0.01387	0.00851	0.02260
logFpar_16	-7.13570	0.24880	0.00080	0.00048	0.00131
logFpar_17	-6.12370	0.24571	0.00219	0.00134	0.00358
logFpar_18	-5.32078	0.24462	0.00489	0.00300	0.00797
logFpar_19	-4.17854	0.24700	0.01532	0.00935	0.02511
logFpar_20	-3.28732	0.23191	0.03735	0.02349	0.05940
logSdLogFsta_0	-2.20065	0.84913	0.11073	0.02026	0.60509
logSdLogFsta_1	-2.40944	0.23054	0.08987	0.05667	0.14251
logSdLogN_0	-0.14520	0.13072	0.86485	0.66588	1.12327
logSdLogN_1	-2.21469	0.34948	0.10919	0.05428	0.21965
logSdLogObs_0	-0.53038	0.14289	0.58838	0.44213	0.78302
logSdLogObs_1	-1.58510	0.08740	0.20493	0.17206	0.24407
logSdLogObs_2	-0.81995	0.13556	0.44045	0.33586	0.57763
logSdLogObs_3	-0.35495	0.06702	0.70121	0.61325	0.80178

	par	sd(par)	exp(par)	Low	High
logSdLogObs_4	-0.22718	0.11930	0.79678	0.62764	1.01149
logSdLogObs_5	-0.24855	0.11200	0.77993	0.62341	0.97575
logSdLogObs_6	-0.11763	0.10359	0.88903	0.72266	1.09369
logSdLogObs_7	-0.53339	0.12039	0.58661	0.46108	0.74632
itrans_rho_0	0.84298	0.39299	2.32329	1.05865	5.09863
logScale_0	0.01925	0.14815	1.01944	0.75802	1.37101
logScale_1	-0.16768	0.16966	0.84562	0.60230	1.18724
logScale_2	-0.11470	0.18214	0.89163	0.61941	1.28348
logScale_3	0.12800	0.18806	1.13655	0.78026	1.65552
logScale_4	0.18321	0.19106	1.20107	0.81963	1.76003
logScale_5	0.35964	0.19499	1.43281	0.97010	2.11620
logScale_6	0.69873	0.19951	2.01120	1.34948	2.99739
logScale_7	0.60431	0.19787	1.82999	1.23191	2.71844
logScale_8	1.11045	0.19351	3.03573	2.06152	4.47033
logScale_9	1.36295	0.18365	3.90772	2.70651	5.64206
logScale_10	1.18261	0.17340	3.26288	2.30667	4.61548
logScale_11	0.66893	0.22411	1.95214	1.24695	3.05614

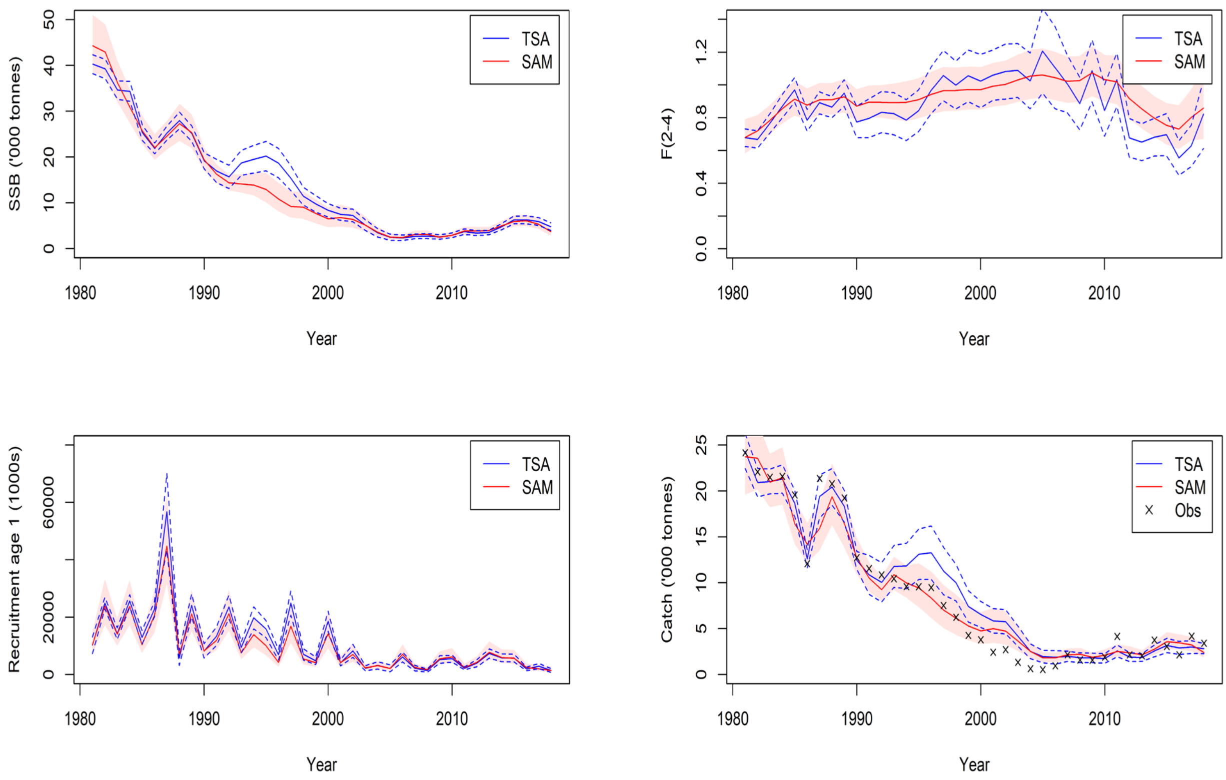


Figure 3.2.1. Cod.27.6a. Final SAM model run compared to TSA (the latter with data agreed at WKDEM and same model configuration as WGCSE 2019). Shaded areas/dashed lines represent 95% confidence intervals.

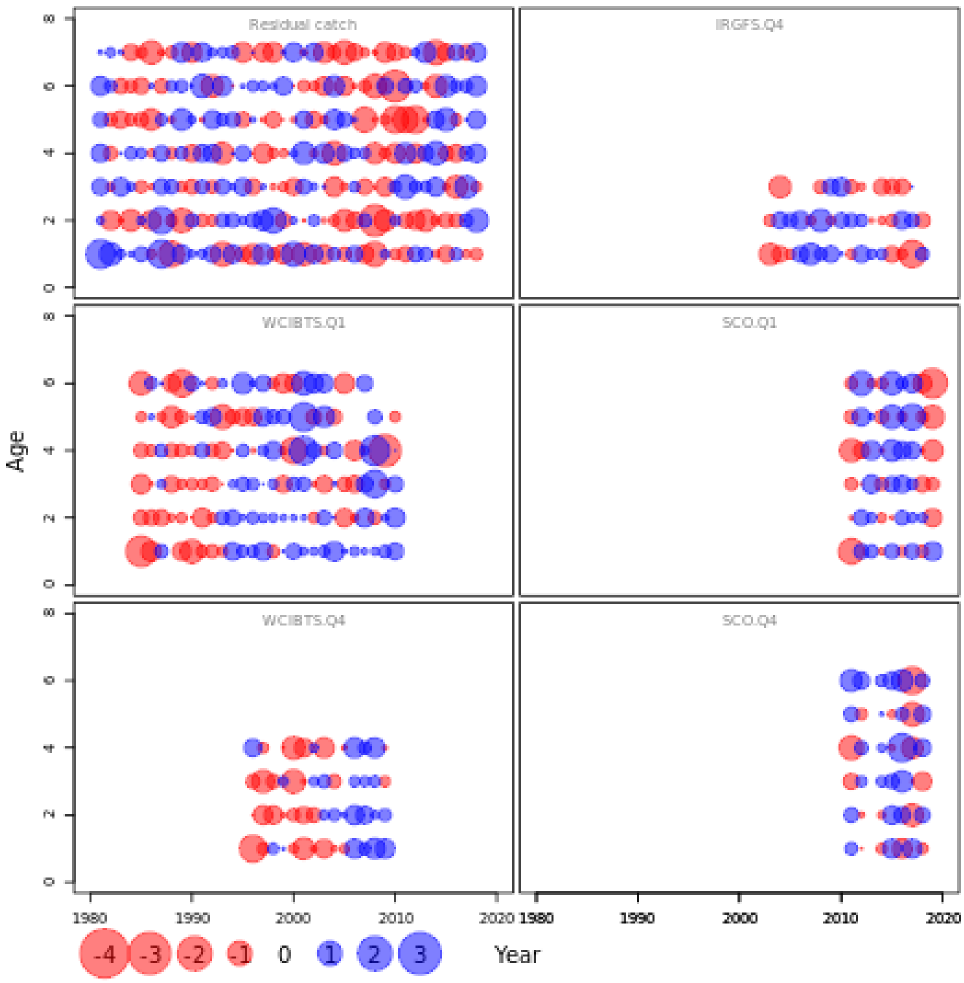


Figure 3.2.2. Cod.27.6a. One step ahead residuals from the final assessment model run.

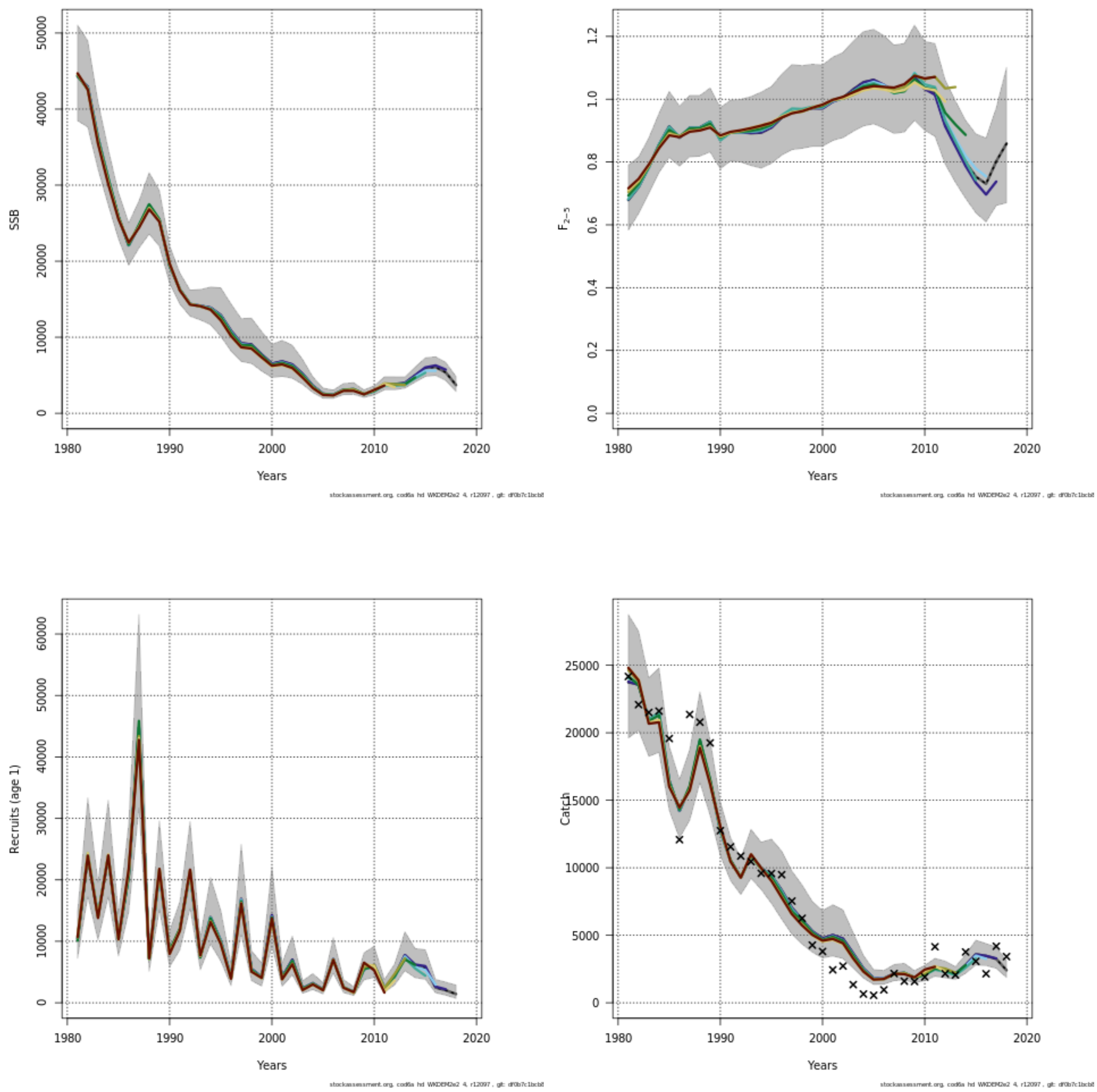


Figure 3.2.3. Cod.27.6a. Final SAM model run including retrospective analysis. Shaded areas represent 95% confidence intervals.

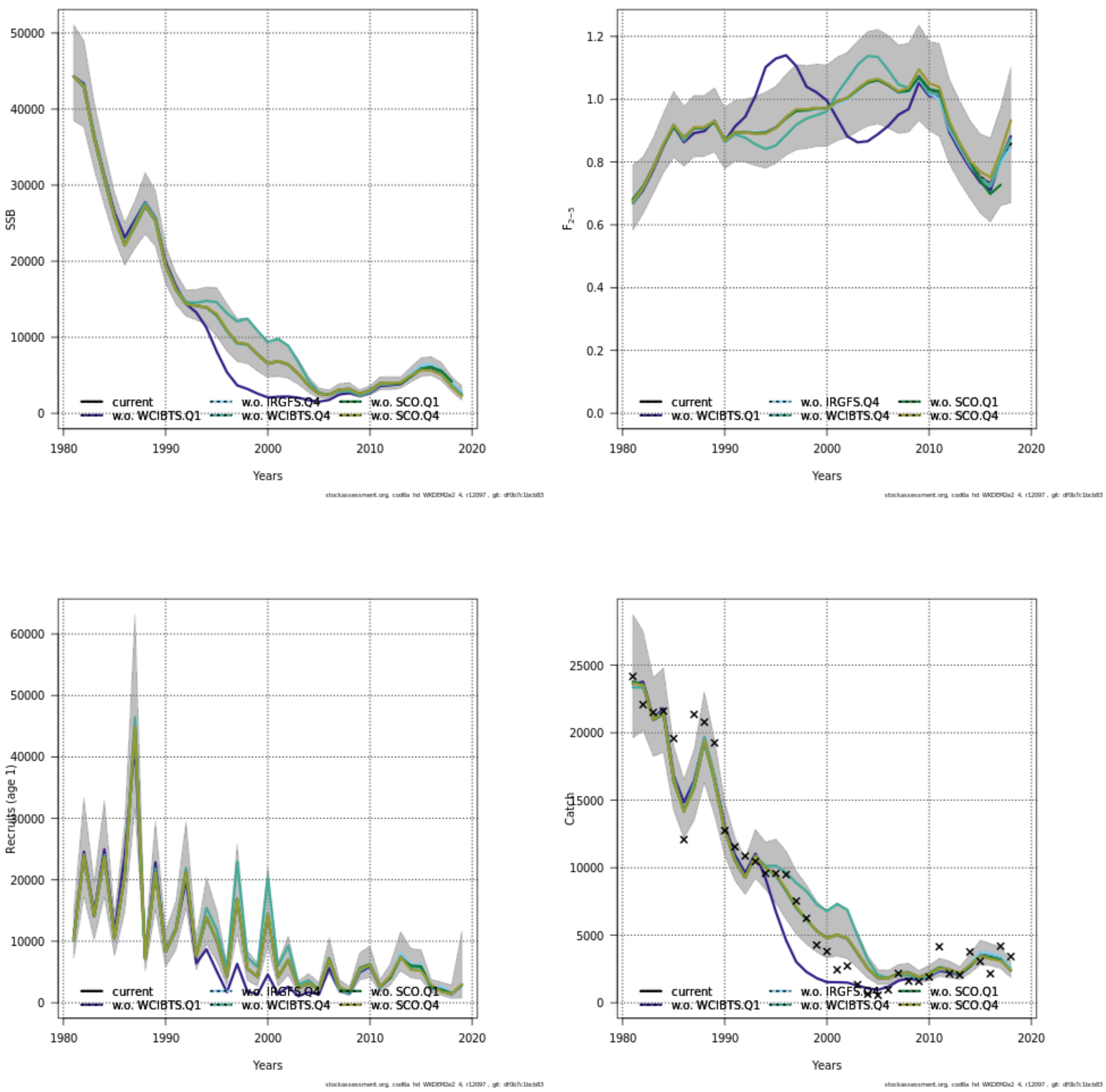


Figure 3.2.4. Cod.27.6a. Final SAM model run with leave one out (surveys) runs. Shaded areas represent 95% confidence intervals.

3.3 Investigations undertaken (summary)

3.3.1 Assessment model

TSA has previously been used as the assessment model for this stock. The current implementation of TSA is slow to converge and difficult to optimise and therefore for practical reasons there was a motivation for moving to an alternative assessment method at this benchmark. In addition, given the current resources, there is likely to be limited future development support for TSA within Marine Scotland Science (MSS) where the main expertise lies. A number of widely available assessment models (SAM, a4a) were explored as potential alternatives, with the main focus being on SAM, primarily due to limitations with the way a4a handles missing (or at least unreliable) catch data. An additional model which estimates time varying natural mortality (TVM, WD 4.8) was presented at the meeting as an exploratory assessment.

The final agreed model settings are described in Section 3.2. However, a number of sensitivity analyses were conducted before agreeing those settings and those are described in further detail here.

3.3.2 Catch-scaling period

From around the mid-1990s to 2006, under-reporting of landings is considered to have been an issue in the Scottish demersal fishery, and is likely to have impacted the quality of the catch data during this period. For this reason, previous TSA assessments have made use of only the landings and discards age composition data during this period and estimated an annual total catch scaling parameter to account for the under-reporting. The TSA assessment agreed at the 2012 benchmark meeting excluded total catches from the assessment from 1991, despite under-reporting apparently not being considered as an issue until the mid-1990s (ICES, 2012) the reason cited being to provide model overlap with a period in which catch data were considered to be of adequate quality (although it is not clear why this would be required except to check that catch scaling estimates were ~1 during this period).

A sensitivity analysis of the SAM model results to the first year of catch scaling estimation was carried out (Figure 3.3.1). Minor differences in estimated catch, SSB, mean F and recruitment are apparent during the early and mid-1990s with model estimates converging from the late 1990s onwards. In all model runs, catches are underestimated in the early 1990s and when a catch scaling factor is estimated from 1990 onwards, it estimates a value below 1 (i.e. catches have been under-reported) in the early 1990s (the same feature is apparent in the TSA, but to a lesser extent). Given there is no reason to suspect that under-reporting of landings was a major issue, the period of catch scaling estimation in the baseline model run was taken to begin in 1995. A comparison of the estimated catch scaling factors is given in Figure 3.3.2.

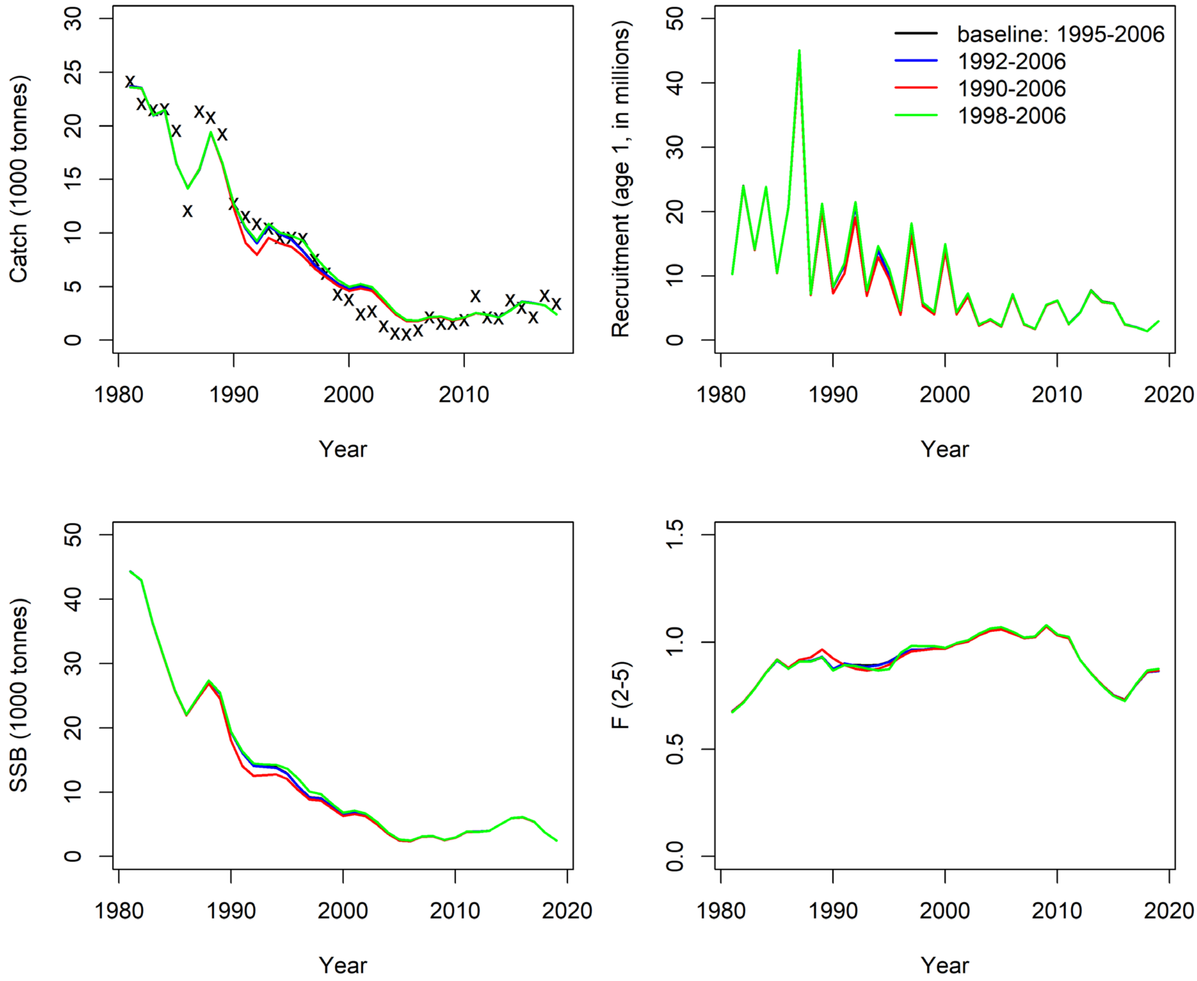


Figure 3.3.1. Cod.27.6a. Comparison of results from SAM model runs with different periods for which a catch scaling factor is estimated.

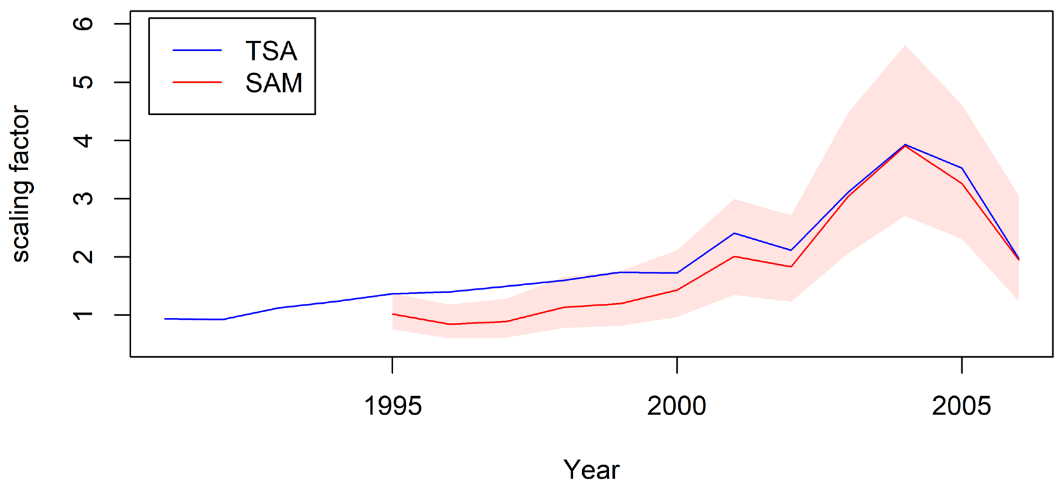


Figure 3.3.2. Cod.27.6a. Comparison of catch scaling factor estimated by the final SAM run (including confidence intervals) and by TSA.

3.4 Catch observation error

Initial SAM model runs in which catch observation variance was assumed to be the same over all ages resulted in some heterogeneity in residuals, with larger residuals apparent for ages one and seven (Figure 3.3.3). External estimates of CVs associated with the Scottish landings and discards data (derived from market and observer sampling data as part of the assessment input data estimation process) also suggest greater uncertainty at younger and older ages (see ICES, 2019a). In the final assessment model run the catch observation variance parameters are therefore allowed to differ for age 1 and age 7+ while other age groups are coupled (# Coupling of the variance parameters for the observations) which improves the catch residual patterns.

The residuals shown in Figure 3.3.3 (initial model run) also suggest an increase in residuals towards the end of the assessment time period. The timing of this change appears to be consistent with the point at which the change in reporting regulations resulted in the fishery changing from one largely dominated by landings to one largely dominated by discards (the regulation made it much harder to make unreported landings). Given typically poorer sampling levels for discards (than landings) and the increased importance of discards in the total catch since 2006 it is unsurprising that the catch data are more uncertain during this period. (This is potentially exacerbated by the large reduction in landings and hence a reduction in landings sampling opportunities resulting in greater uncertainty in the landings age composition data during this period as well). A doubling of the standard deviation in the catch data from 2006 onwards (when the regulation came into force), results in less heterogeneity in the catch residuals over time.

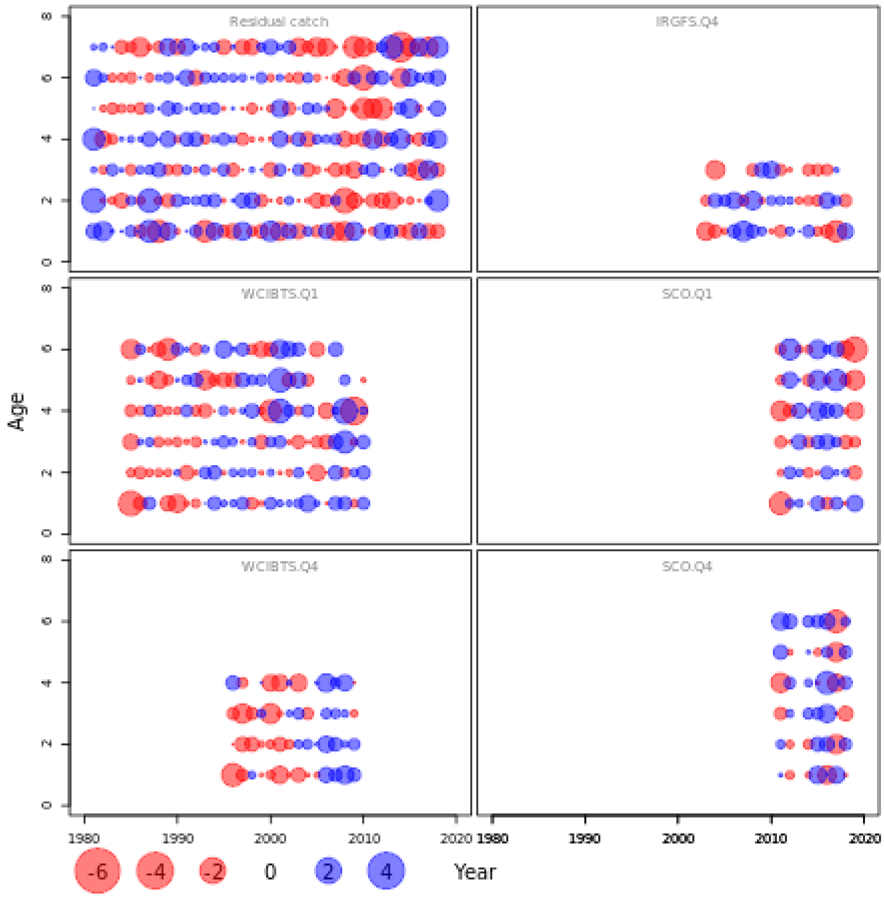


Figure 3.3.3. Cod.27.6a. One step ahead residuals from SAM model run with coupled catch observation across all ages and assuming no change in observation error over time.

3.4.1 Fishery selectivity/survey catchability

At the inter-benchmark in 2019, differences in the estimated selection pattern between the TSA assessment and that presented in Cook (2019a) were highlighted as being central to the differences in the perception of stock status in the two approaches. The Cook (2019a) model estimates a very dome-shaped fishery selectivity pattern while TSA assessment estimates it to be relatively flat-topped even when the shape is allowed to be freely estimated. A significant amount of time at this benchmark was therefore spent exploring this issue (and the related issue of estimated survey catchability) using both the SAM model and an exploratory a4a assessment.

The robustness of the SAM model results to different assumptions about fishery selectivity (with limited constraints on the survey catchability) was explored. Figure 3.3.4 compares model outputs from SAM runs with fishery selectivity flat-topped above age 3, age 4 (the final assessment) and age 6 (the latter option allowing the pattern to be more freely estimated). The change in model assumptions results in very little difference in estimates of stock metrics. Despite the fishery-selectivity pattern being allowed greater flexibility in the third model run presented here (F-at-age 6 = F-at-age 7), the estimated pattern either increases over ages or remains virtually flat-topped over the whole time period. (Consistent with previous TSA results; ICES, 2019a). Across the three model runs, survey catchability estimates show a generally increasing trend with increasing age (with the exception of those constraints described in Section 3.2), with little difference between models.

There is very little to choose between the diagnostics of these three models (diagnostic plots available on stockassessment.org). The model in which fishery selectivity was coupled for ages 4 and above was chosen as the final SAM assessment due to showing slightly better retrospective patterns (despite have a slightly higher AIC).

RunID on stockassessment.org	Fishing mortality	loglikelihood	Num params	AIC
Cod6a_hd_WKDEM2e2	Flat-topped age 3 and above	-586.41	46	1264.82
Cod6a_hd_WKDEM2e2_4	Flat-topped age 4 and above	-588.96	46	1269.92
Cod6a_hd_WKDEM3e1	Equal at age 6 and 7	-594.93	46	1281.85

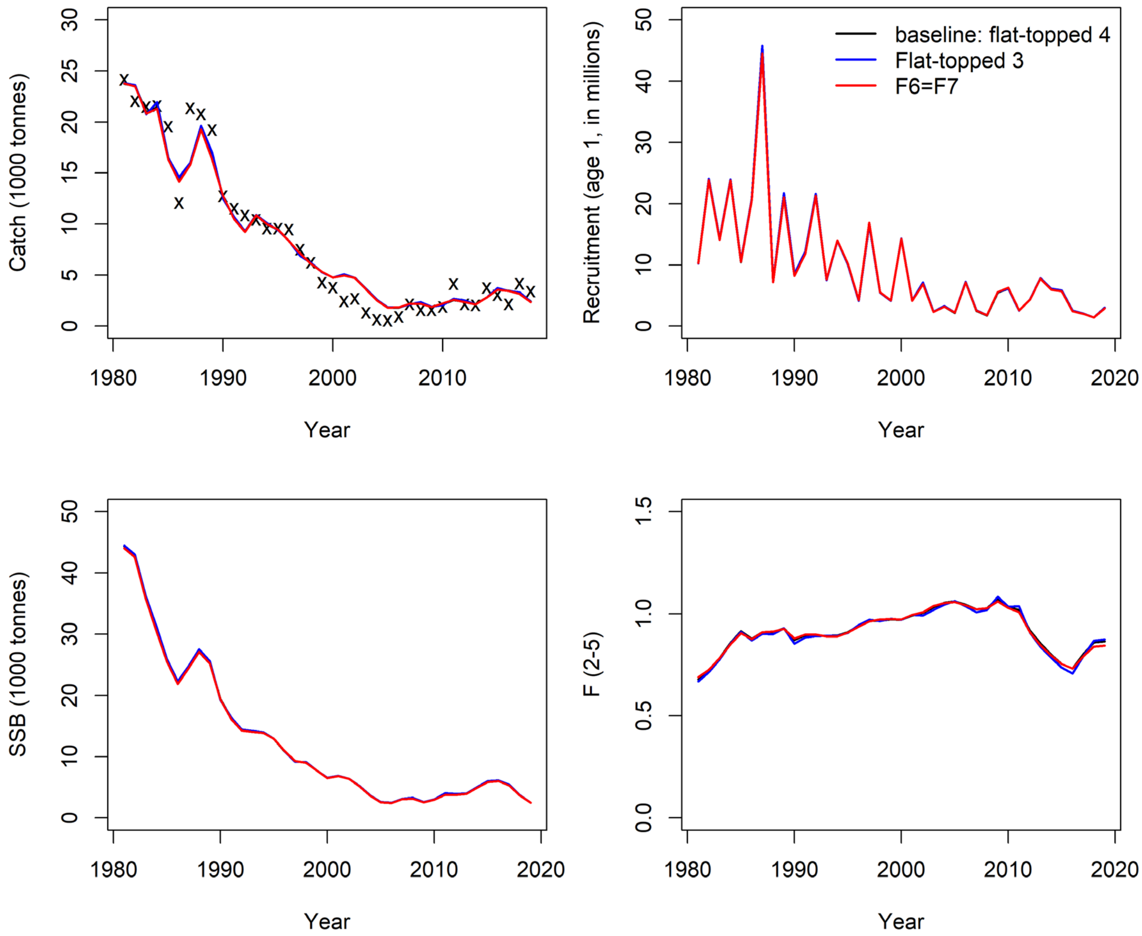


Figure 3.3.4. Cod.27.6a. Comparison of SAM model results with different assumptions about fishery selectivity. Black: final SAM assessment with coupled estimates at-age 4 and above, blue: coupled estimates age 3 and above, Red: coupled estimates age 6 and above (i.e. shape allowed to be more freely estimated).

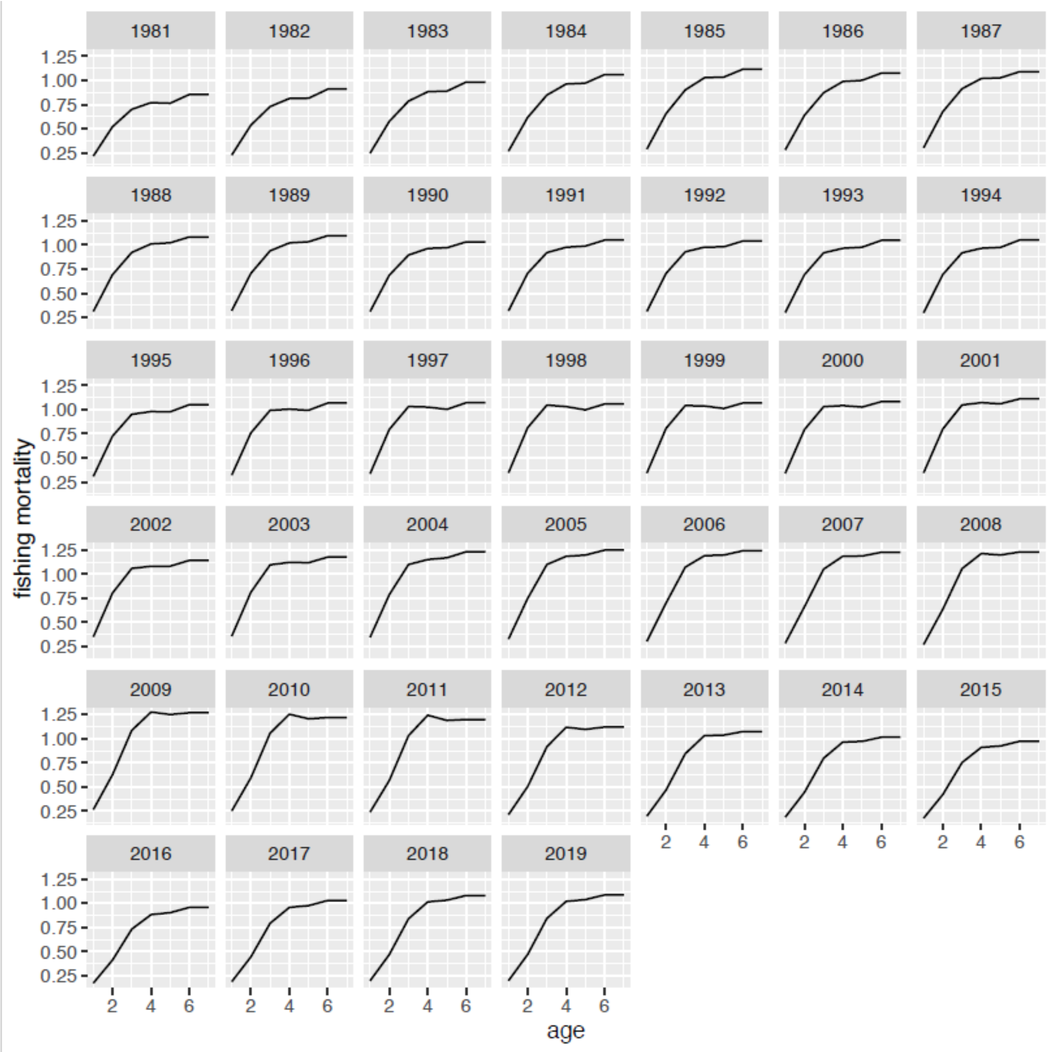


Figure 3.3.5. Cod.27.6a. Fishery selectivity estimated in the SAM model run with estimates coupled at-age 6 and 7 and other ages estimated separately (with limited constraints on survey catchability, see Section 3.2).

To explore the sensitivity of the results to the survey catchability estimates, a further series of SAM model runs was conducted in which the survey catchability (all surveys) was constrained to be flat-topped. In these sensitivity runs, the fishery selectivity pattern was also freed up (as above) to explore the impact on estimates of fishing mortality.

RunID on stockassessment.org	Fishing mortality	Survey catchability	loglikelihood	Num params	AIC
Cod6a_hd_WKDEM2e2_4	Flat-topped age 4 and above	Constraints as per section 4.2	-588.96	46	1269.92
Cod6a_hd_WKDEM3e1	Equal at-age 6 and 7	Constraints as per section 4.2	-594.93	46	1281.85
Cod6a_hd_WKDEM7e1	Equal at-age 6 and 7	Coupled age 4 and above (all surveys)	-606.96	44	1301.92
Cod6a_hd_WKDEM8e1	Equal at-age 6 and 7	Coupled age 3 and above (all surveys).	-629.5	40	1338.99

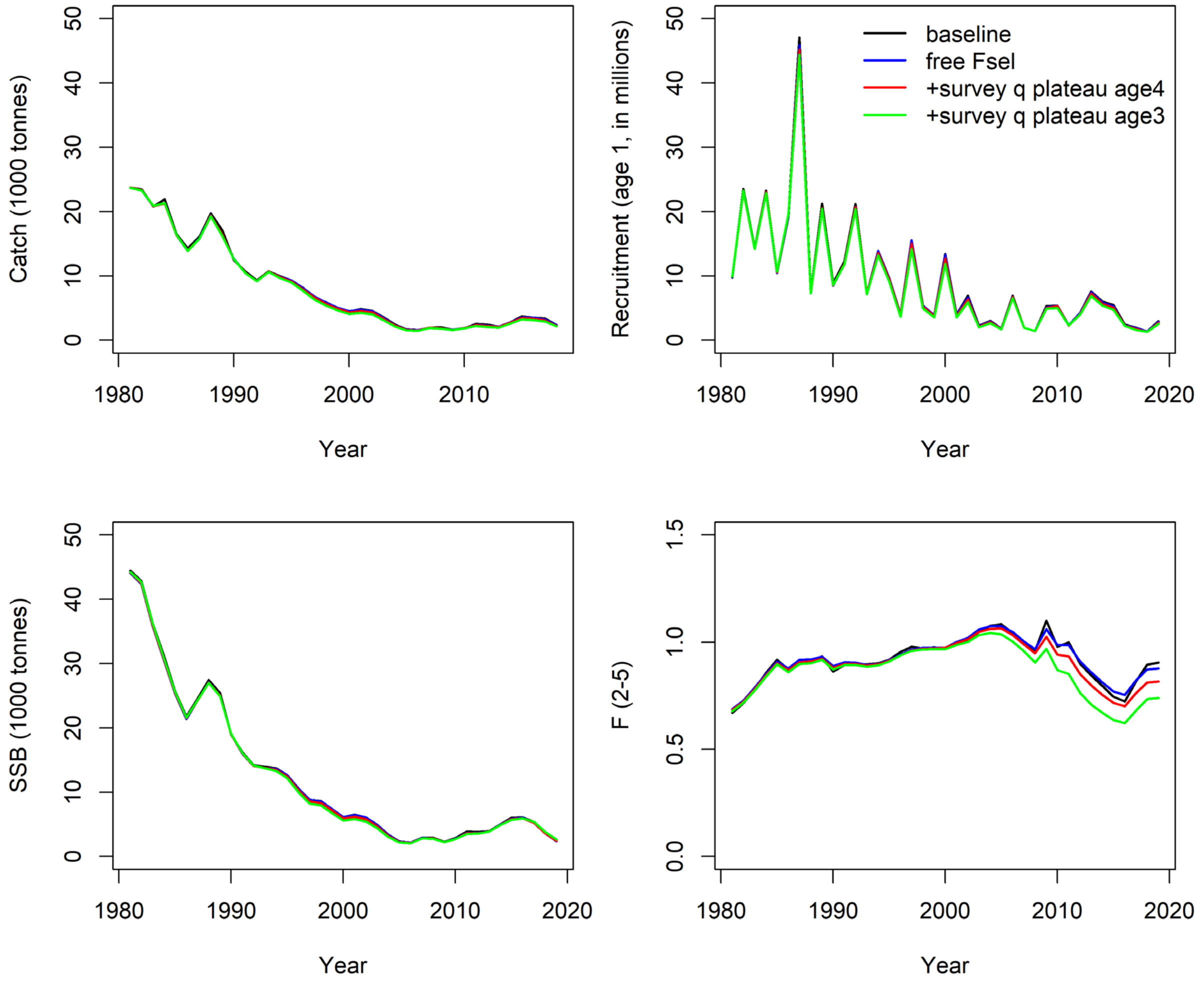


Figure 3.3.6. Cod.27.6a. Comparison of SAM model results with different assumptions about survey catchability. Blue: The final SAM assessment is also included for comparison (black line).

The model results (other parameter estimates and stock summary, the latter is shown in Figure 3.3.6) appear relatively insensitive to these changes in survey catchability. The estimates of recruitment and SSB from the model runs are almost indistinguishable (Figure 3.3.6) while there are small differences in the estimate of mean F with the model runs in which the plateau in survey catchability is assumed to include younger ages (red and green lines in Figure 3.3.6) showing a slightly greater decline in mean F.

The estimated fishery selectivities from the model run with survey catchability plateau at-age 3 and above are shown in Figure 3.3.7. During the early part of the time-series the estimates of fishing mortality increase with age and then become generally flat-topped in the 1990s. In this assessment (with survey catchability forced to be flat-topped), there is some evidence of a limited 'doming' of the selectivity pattern (particularly from the mid-2000s onwards) with highest fishing mortality estimated for ages 3 and 4 and slightly lower values at older ages. Note that this model run results in a worse fit to the data (in terms of AIC) and worse survey residual patterns, particularly for the WCIBTS.Q1 (persistent underestimation at-age 3 and overestimation at-ages 5 and 6).

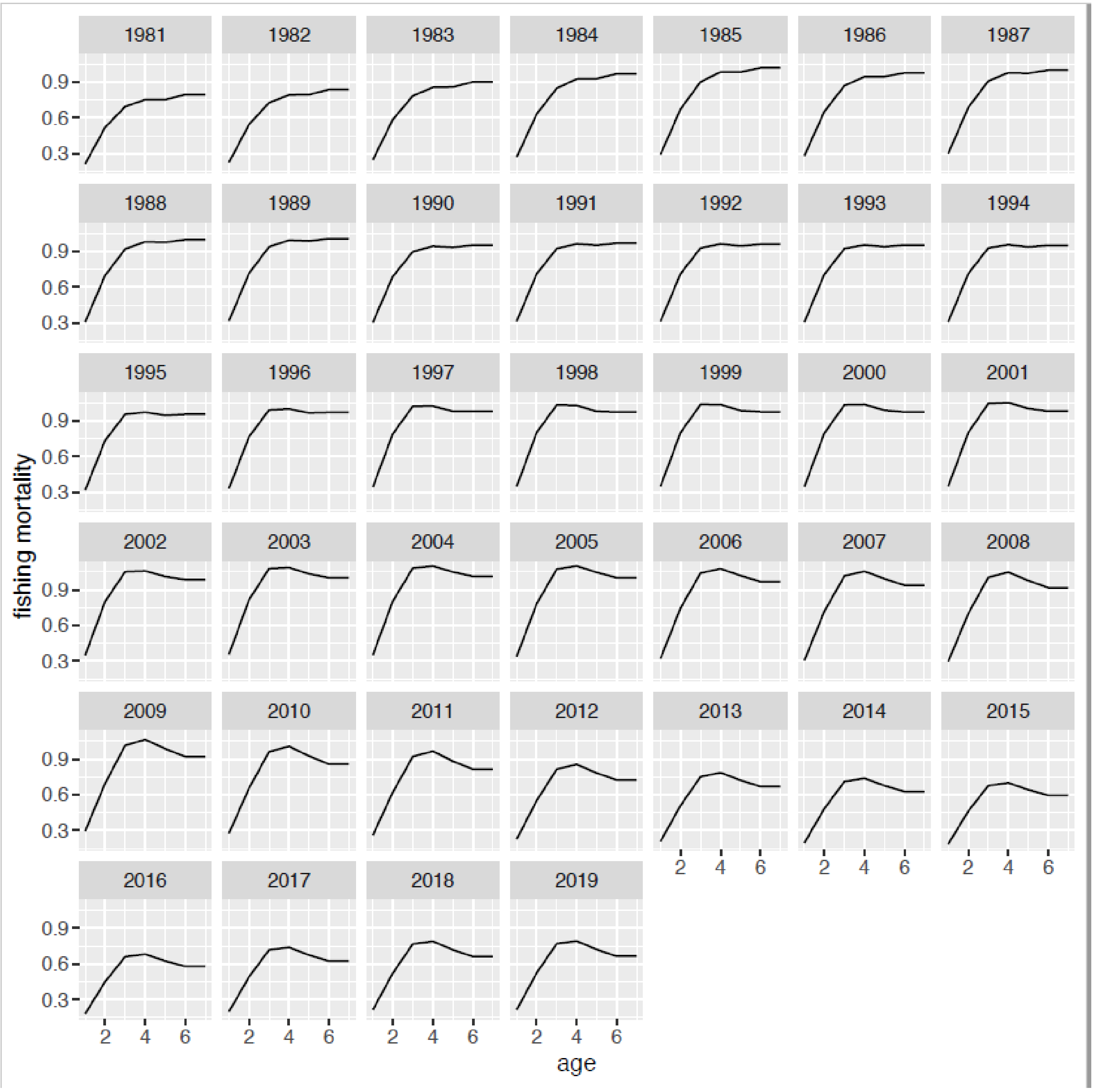


Figure 3.3.7. Cod.27.6a. Fishery selectivity estimated in the SAM model run with estimates coupled at-age 6 and 7 and other ages estimated separately, and survey catchability constrained to a plateau at-age 3 and above.

Like the TSA sensitivity analysis conducted at the IBP which took place in 2019, none of the sensitivity runs presented here suggest a significant dome-shaped fishery selectivity. This in contrast to the TVM model which estimates fishing mortality-at-ages 5 and 6 to be lower than ages 2, 3 and 4 throughout the time-series.

Some further exploration of estimated fishery selectivity (and stock trends) and sensitivity of these associated with different survey catchability assumptions was carried out using a4a (Jardim *et al.*, 2015). a4a is a flexible stock assessment framework based on a statistical catch-at-age model implemented in R (making use of FLR) and using ADMB for optimisation purposes. The current implementation of a4a is unable to make use of catch age composition data alone (i.e. without including absolute level of catches) and therefore it was not possible to run the assessment for the full time-series i.e. including the period when under-reporting of landings was believed to be significant. Assessments were therefore conducted based on different time periods of data: i) the recent time period (2007 onwards) - the time period for which the landings are considered to be unbiased and during which time the fishery has been dominated by discards and ii) the early time period (pre-1995) when catch data are considered to be unbiased and discards much less significant.

In both cases a separable fishing mortality model is initially implemented, and then models including various types of smoothing: the use of splines to smoothly model selectivity (F-at-age) and F over years independently, and then consider a tensor product of cubic splines over age and year (correlation along age and year, but also cross correlation). In all cases, survey catchability is age-dependent (without smoothing over ages and no year effect) and recruitment is estimated as a parameter for each year (i.e. without including a stock–recruitment model).

In both the early and late period stock assessments, using a separable model for fishing mortality (without smoothing) results in an almost monotonic increasing fishery selectivity. In the versions which include smoothing there is a flattening off of the selectivity with small decline at older ages and in the non-separable case a more significant decline at older ages in some years (although not persistent over time and not to the extent described in Cook (2019a)). For each assessment period, the different fishing mortality models result in very similar stock summaries and these show good consistency with the SAM model in the recent period (Figure 3.3.8). The associated estimates of survey catchability are also similar to those estimated by SAM, showing generally increasing estimates with age.

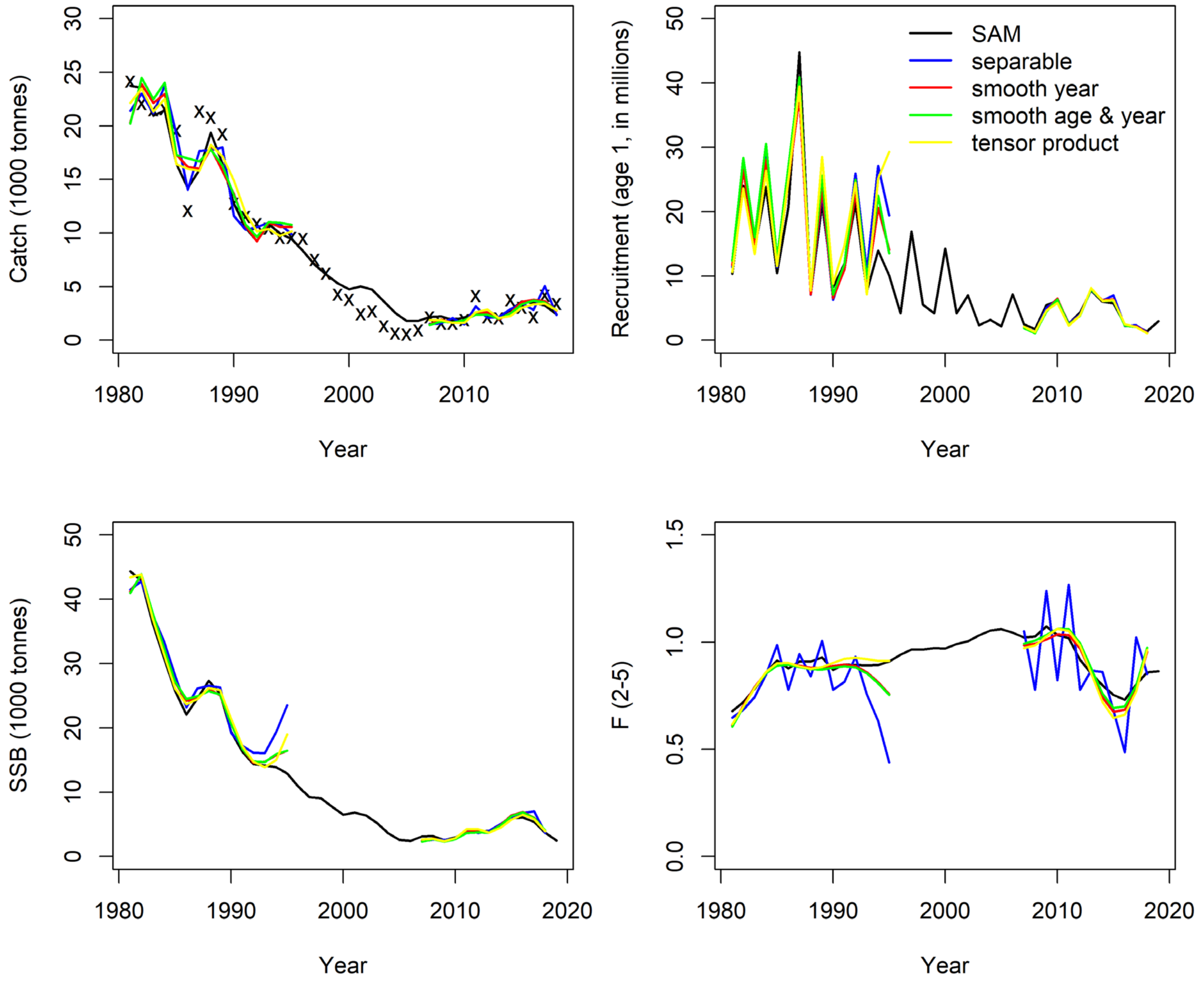


Figure 3.3.8. Cod.27.6a. Comparison of a4a assessments with different fishing mortality model assumptions with the final SAM assessment.

A further set of sensitivity tests was carried out using a4a in which the impact of constraining the survey catchability is investigated. A separable fishing mortality model is used, and four model runs conducted in which increasing constraints are placed on the estimates of survey catchability: from independently estimated at all ages to a plateau in the estimates at-age 3 and above. In the early period assessments, the estimated fishery selectivity pattern is robust to the assumptions about survey catchability i.e. it remains increasing/flat-topped for all alternative runs (Figure 3.3.10). In contrast, for the recent period assessments, estimates of fishing mortality at older ages (5 to 7) are lower than at ages 3 and 4 when survey catchability is fixed for age 4 and above, and the estimates become severely dome-shaped when survey catchability is flat-topped for age 3 and above (the latter being similar to the pattern estimated in the TVM model). The a4a assessment model results are robust to these differences with the exception of the run with survey catchability plateau at age 3 (in the late period assessment) (Figure 4.3.9). Estimates of mean F in this latter model run are much lower and show a general decline (although estimates are noisy) from the start of the time period (2007) which is in contrast to the runs with other catchability assumptions where the decline in mean F does not occur until around 2010 (more similar to the SAM assessment). The lower mean F is reflected in the increase in SSB since 2007 which is much greater in the other model runs.

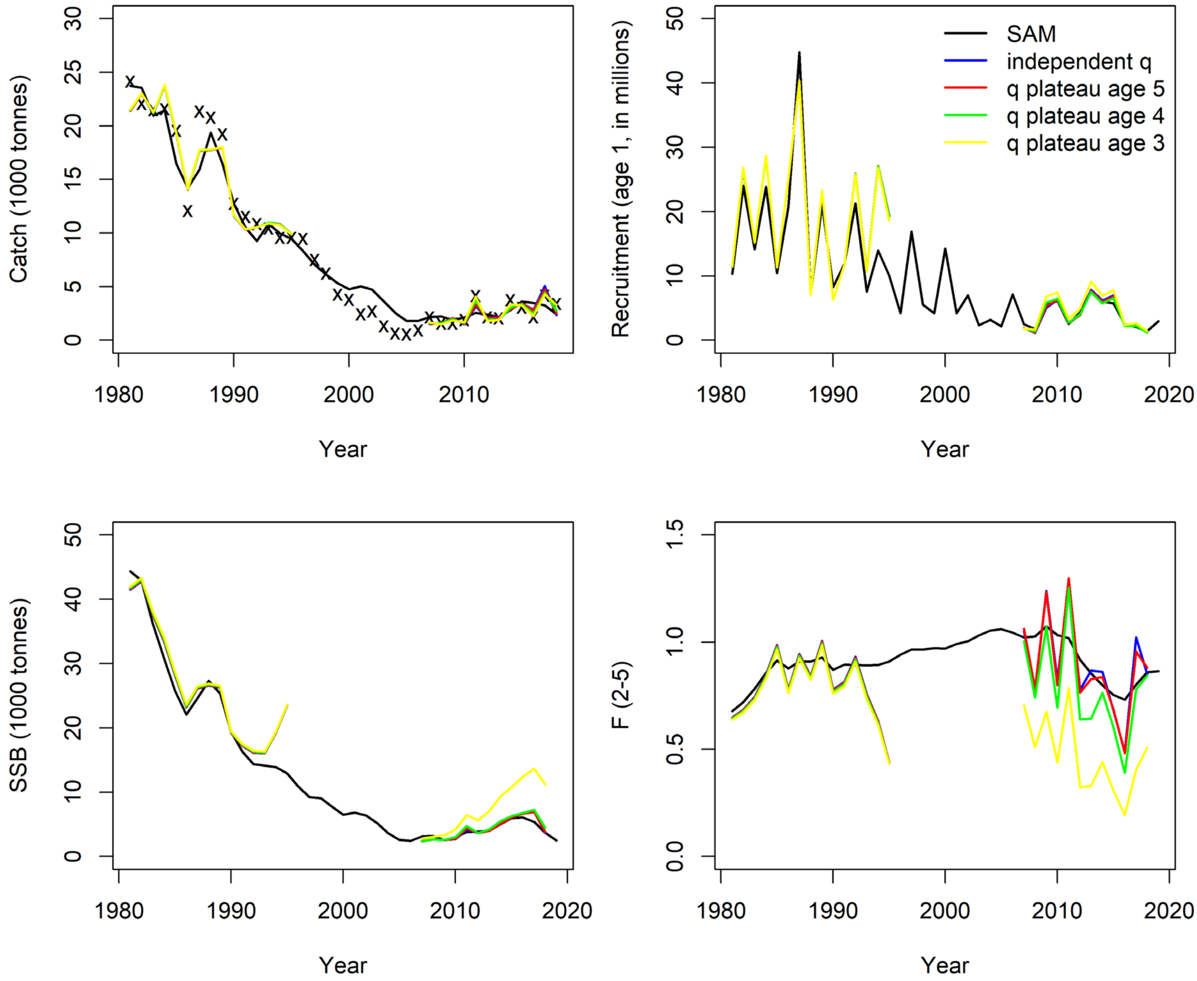


Figure 3.3.9. Cod.27.6a. Comparison of a4a assessments (early and later period assessments shown on same figure) with different survey catchability assumptions with the final SAM assessment (survey catchability assumptions apply to all surveys included in the assessment).

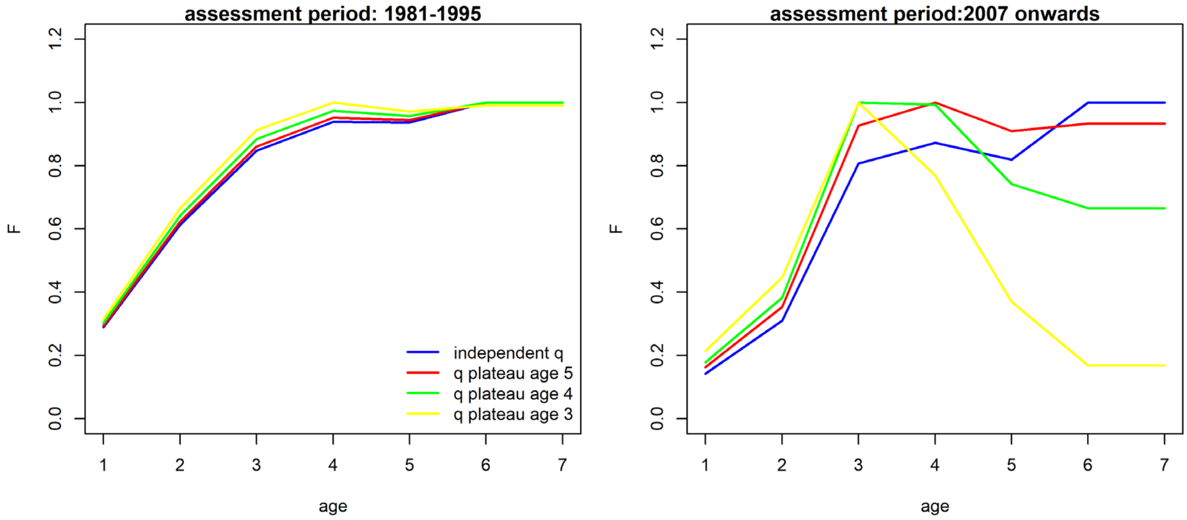


Figure 3.3.10. Cod.27.6a. Comparison of estimated fishery selectivity patterns from a4a model runs with different assumptions about survey catchability (applies to all surveys included in the assessment).

To summarise, the SAM assessment model finds limited evidence of a significantly dome-shaped fishery selectivity pattern even when survey catchabilities are constrained to be flat-topped, although there is some evidence of a change in selectivity in more recent years. This is consistent with the findings from the exploratory a4a assessments which have been conducted in two time blocks: only in the recent period does the model estimate any doming in the fishery selectivity and this only results in significant differences to the assessment results when the survey catchability (all relevant surveys) is forced to plateau at-age 3. This is in contrast to the TVM assessment which estimates a significantly dome-shaped fishery selectivity throughout the time period of the stock assessment and seems less likely.

3.4.2 Natural mortality

Weight-dependent natural mortalities-at-age were adopted for the first time at the benchmark meeting in 2012 to take account of increased mortality at younger ages (rather than fixed across all ages). At this benchmark meeting, it was agreed that assumed natural mortalities should be time varying to reflect the generally decreasing trend in mean weights. Therefore, natural mortalities were derived as a function (Lorenzen, 1996) of stock weight-at-age (which are in turn are derived as modelled catch weights-at-age), resulting in a smooth time-series of natural mortalities (Figure 3.3.11). The natural mortality estimates from the TVM exploratory assessment model are also shown for comparison in Figure 3.3.11.

The sensitivity of the SAM model results was explored in relation to the assumed natural mortality by conducting a model run making use of the natural mortalities estimated by TVM. The SAM model failed to converge when the raw estimates were used as input data (most likely due to the high variability) and therefore smoothed values were used. The results are shown in Figure 3.3.12 and show that the run using the TVM estimates of M result in a significant rescaling of the recruitment and a small rescaling of the SSB estimates while the estimated mean F shows a steady decline since the mid-1980s.

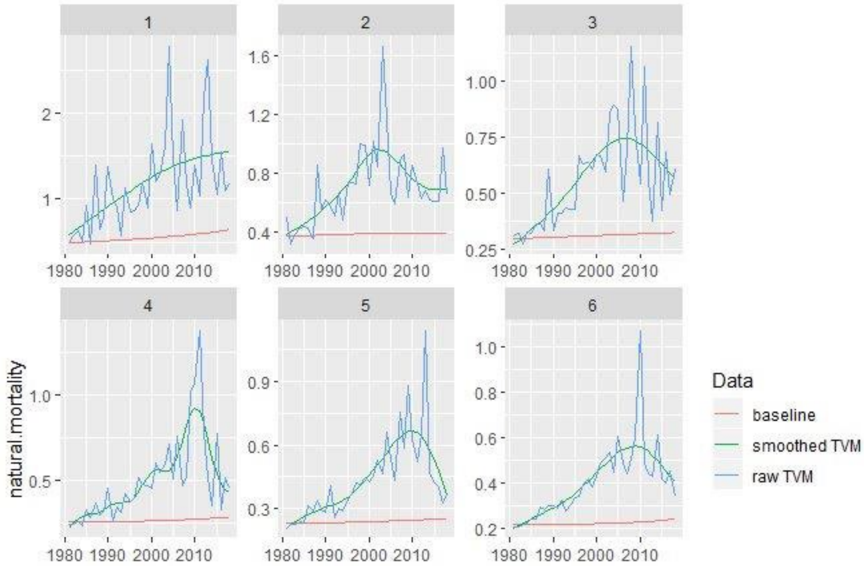


Figure 3.3.11. Cod.27.6a. Comparison of natural mortality values used in SAM model runs. Baseline (black): Time-varying, modelled weight-dependent values (Lorenzen, 1996); Raw TVM (blue): estimates from the exploratory TVM assessment model (WD 4.8); Smoothed TVM (green): derived using a gam fitted to the raw TVM values.

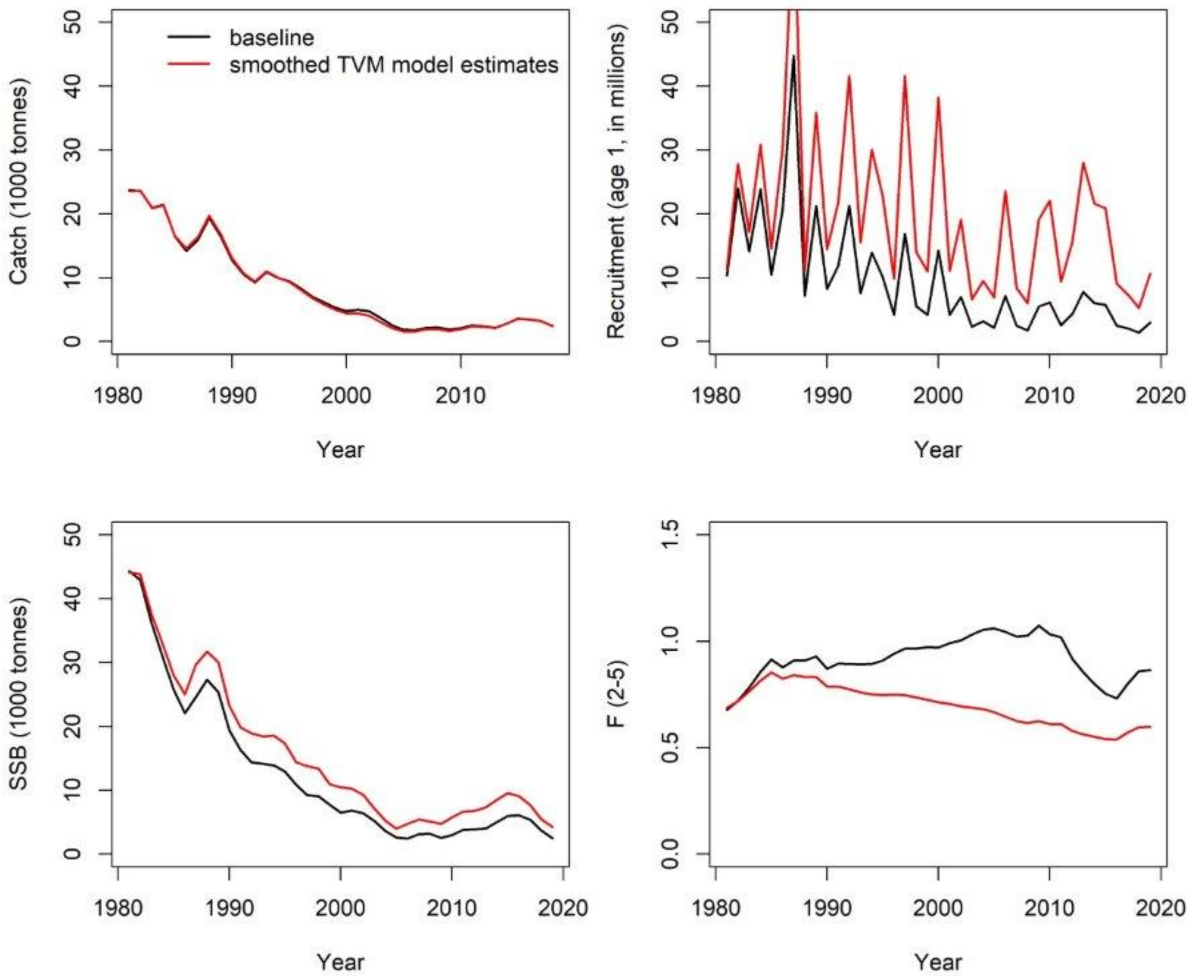


Figure 3.3.12. Cod.27.6a. Comparison of SAM assessment model runs with different natural mortality assumptions: baseline model (black) uses time-varying weight-dependent estimates (Lorenzen, 1996); red: smoothed TVM estimates.

3.4.3 Assessments with alternative models (TVM and FAF)

The models used to assess the stock are based on the peer reviewed model in Cook (2019b), referred to as the FAF model, and with an extension to allow natural mortality to be estimated by the model (the “TVM” model). The TVM model is described and tested in the working document (Cook, 2020, WD 4.8). In addition, a version of the FAF model was run where seal predation was estimated directly using estimates of cod consumption by seals made by the Sea Mammal Research Unit (SMRU). Sensitivity tests were conducted that included constraints on the commercial fishery-exploitation pattern and survey selectivity. A summary of the main results is presented here with further details in the working document. For simplicity, model fit diagnostics are only shown for the TVM model.

Diagnostics: Figure 3.3.13 shows the residuals for the surveys, landings and discard data. Patterns of negatives and positives are evident in the quarter 4 surveys but less so in the quarter 1 survey and landings data. The pattern in the discard data is the result of the change in the pattern of discarding that occurred from 2006 onwards were quota restrictions caused catches to be dumped regardless of fish size. Retrospective runs in Figure 3.3.14 show consistency for both F and SSB but with some change of scale that results from the re-estimation of M for each run.

Stock trends: Figure 3.3.15 shows trends in F and SSB for a range of models. With one exception, all show F declining with current values below that in 1981. When the selection pattern is constrained to be asymptotic and M is fixed at conventional externally derived values, the model produces very high values of F with a small decline in recent years (blue line) but with current F higher than 1981. However, when the same model is run but M is internally estimated (the TVM model), the F trend reverts to a long-term decline. Of the models considered, the FAF model with fixed M (light green) gives results that mirror the reduction in effort (dotted line) in the early 2000s, however, this started prior to the reduction in effort and the change in F seen in the adjacent North Sea cod fishery. Clearly, one of the problems with the data is that there is insufficient information for models to allocate total mortality between F, M and the selection pattern and the emergent properties of the assessment are therefore heavily dependent on modelling assumptions. The changes in SSB are reflections of the F trends where the models producing lower F imply higher biomass. Here, the trend in SSB is similar to the adjacent North Sea cod SSB (Figure 3.3.16).

Survey catchability: As described above, the model that fixes M and constrains the selection pattern to be flat topped produces high F and low biomass. The corollary of this is that the assessment estimates high survey catchability on ages 4 and older. This can be seen in Figure 3.3.17 which shows log catchability estimated from three models for the Scottish quarter 1 surveys in 6.a and the North Sea IBTS (for comparison) estimated from the ICES assessment. FAF1 and FAF2 show the estimated log q for the two surveys when fishery selection is unconstrained, and M is fixed. These catchabilities are more or less asymptotic. When the fishery selection is constrained to be flat topped (FAF1s and FAF2s) the survey catchability rises over all ages and for the oldest fish is about and order of magnitude higher. There is value in further exploration as to potential causes of this variability in perception of catchabilities for the surveys.

Natural mortality: The conventional values of M used in the benchmark assessment are derived from the Lorenzen relationship that relates weight to mortality. The values for 6.a cod are calculated from weights-at-age that are derived from modelling the stock weights and effectively reduces the annual variability in M. The stock weights therefore differ from the weights used to calculate M. An alternative to externally deriving M is to estimate natural mortality from within the assessment model. This was done using the TVM model. The estimated Ms from this model differ from the conventional values and in the case of the youngest ages (1—3) the values happen to be similar to the values used for North Sea cod. The latter are derived from diet studies and predation modelling and hence have some support from direct observations. The values of natural mortality will have a direct effect on the estimated exploitation pattern in the assessment. The exploitation pattern estimated from the TVM model for 6.a resembles that for the adjacent North Sea and is dome-shaped (Figure 3.3.18).

Seal predation: This issue was investigated with the FAF, TVM and FAF+seal predation models. These models estimate substantially higher stock biomass than the benchmark assessment (Figure 3.3.15). There is debate as to the quantity of cod consumed by the seal population and raises questions as to the potential absolute size of the cod stock. This needs further investigations, including spatial and temporal dynamics of both predator and prey populations. Figure 3.3.19 shows a map of seal "usage" from Jones *et al*. (2015) that overlaps in area with all four population components identified in Figure 6 of Wright *et al*. (WD 4.1, this meeting). This includes the Northern Offshore population from which most of the commercial catch is taken and comprises most of the assessed biomass.

3.4.4 Discussion of alternative models

The models discussed here illustrate an uncertainty in perceived mortality rate trends, exploitation patterns and natural mortality. They suggest that there may be similarities between the cod fishery in 6.a and 4 along with F trends and exploitation pattern. The decline in F has a similar trend to the fishing effort, which also occurs in the related whiting and haddock assessments.

It appears that the stock trends (F and SSB) can be interpreted differently by a range of plausible models that fit the data equally well. These differences are likely to be due to the large uncertainty surrounding the catch estimates between 1995–2005, when underreporting is believed to be large, large sampling error in the age composition data that in recent years are heavily dependent on a small number of observer trips and the change in survey design that occurred after 2010. It indicates that an assessment based on a single model may be subject to “type II” errors and that a more thorough exploration is needed to quantify uncertainty in order to inform management advice.

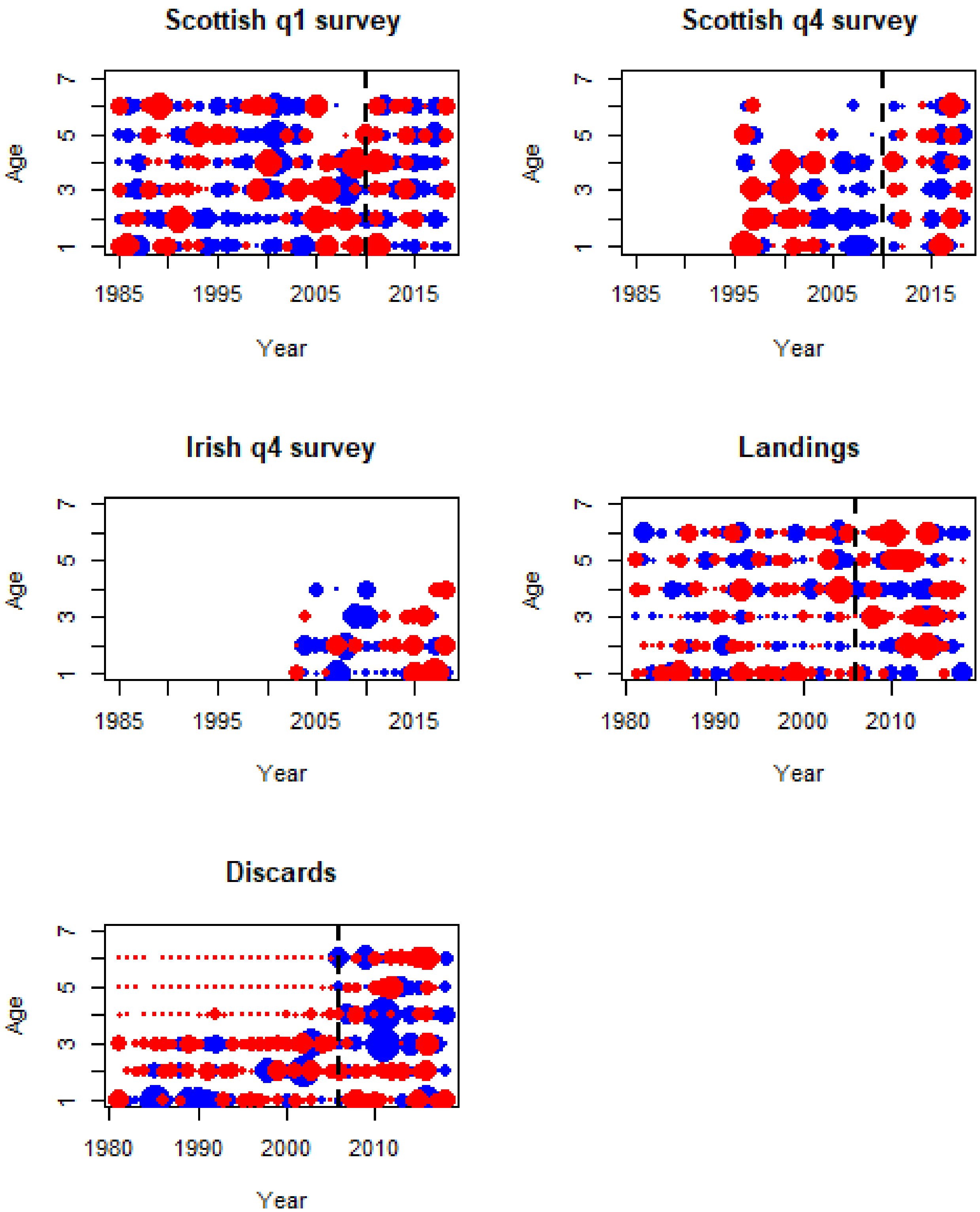


Figure 3.3.13. Cod 27.6a. Residuals for surveys and catch data from the TVM model. Blue=positive, red=negative. Vertical dashed line for surveys indicates the year when the design changed. The vertical dashed line for the landings and discard data indicated the year when the discarding pattern changed when new legislation on landings was introduced. For the discard data the observed values were zero for nearly all years up to 2005 for age 4 and above because fish were overwhelmingly discarded by size rather that quota limits which took effect after this year.

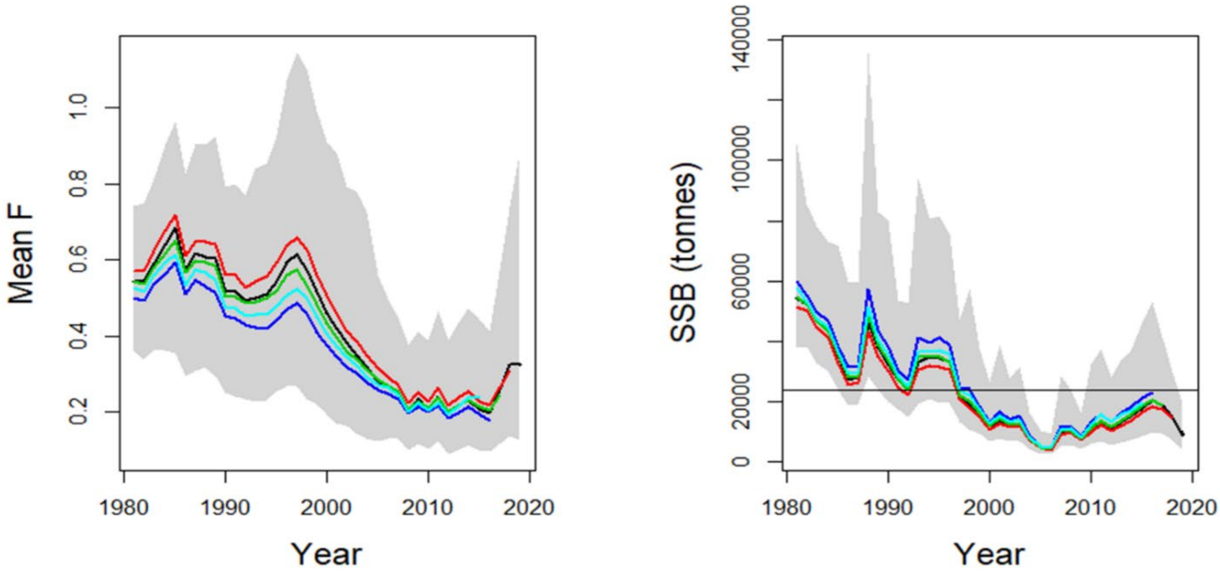


Figure 3.3.14. Cod 27.6a. Retrospective runs for the TVM model. The changes in scale are the result of re-estimating M for each "peel". The horizontal line in the SSB plot shows B_{lim} corresponding to the 1992 SSB. Shaded areas show the 95% credible intervals for the full time period (black line).

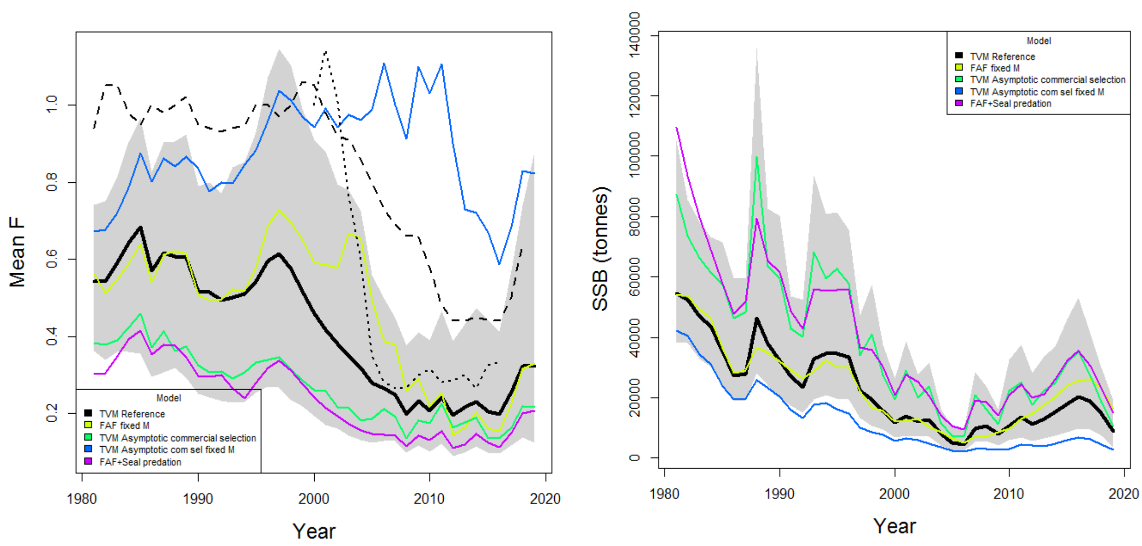


Figure 3.3.15. Cod 27.6a. Summary of mean F and SSB for a range of models (solid lines). Dotted line shows Scottish TR1 effort in kilowatt-days scaled to the 2001 value. Dashed line shows mean F for North Sea cod from the 2019 ICES assessment. The shaded area shows the 95% credible interval for the TVM reference model. All model results fall within the credible interval except for the model with asymptotic fishery selection and fixed natural mortality (blue line).

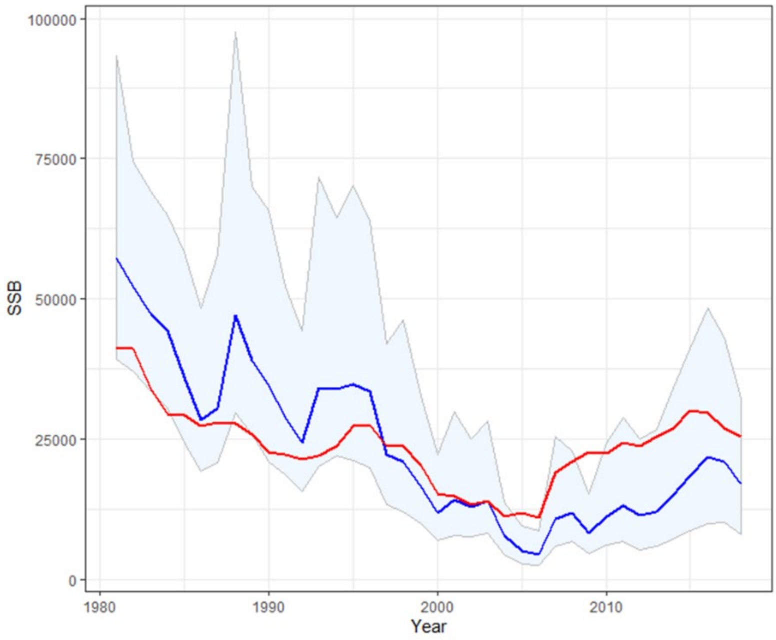


Figure 3.3.16. Cod 27.6a. SSB trend from the TVM model (blue) compared to the ICES North Sea cod assessment (red). The North Sea values are scaled to the mean of the 6a SSB values so that trends can be compared.

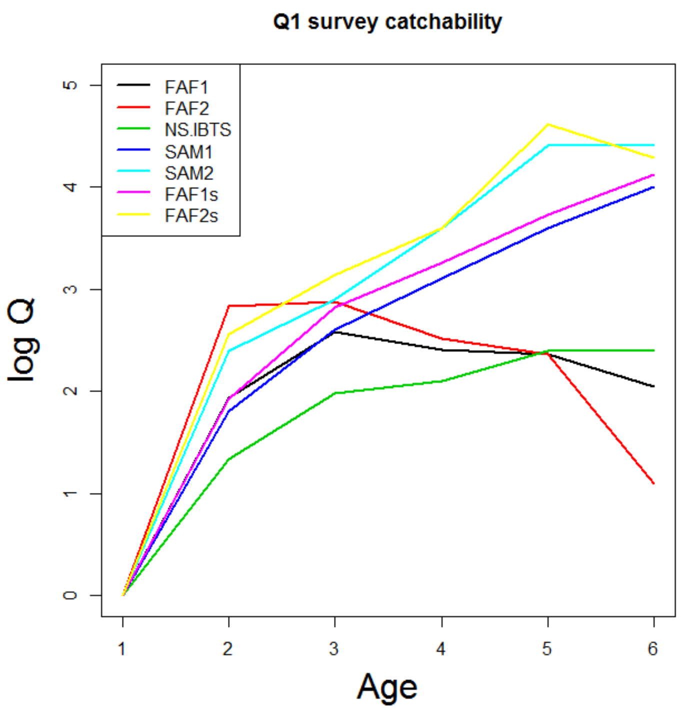


Figure 3.3.17. Cod 27.6a. Estimates of log catchability for quarter 1 surveys expressed relative to age 1. FAF1 and FAF2 are results from the FAF model where M is fixed but fishery selectivity is estimated. FAF1s and FAF2s are results when the fishery selection is set to flat-topped from age 4 and above. SAM1 and SAM2 are the estimates from the current benchmark assessment. NS.IBTS is the pattern estimated from the North Sea assessment for the IBTS. The quarter 1 surveys in 6.a are an extension of the North Sea surveys and use the same vessel and sampling protocol.

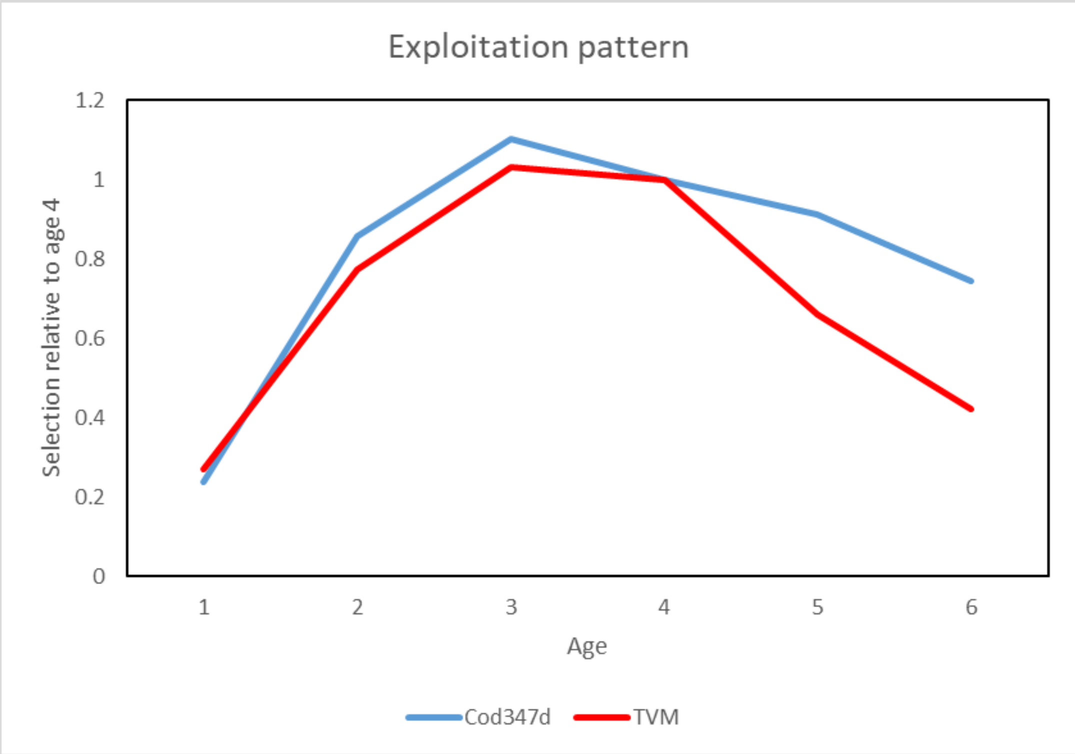


Figure 3.3.18. The mean selection pattern (1981–2018) for 6a cod estimated from the TVM model (red) and cod in 347d from the 2019 ICES assessment expressed relative to age 4.

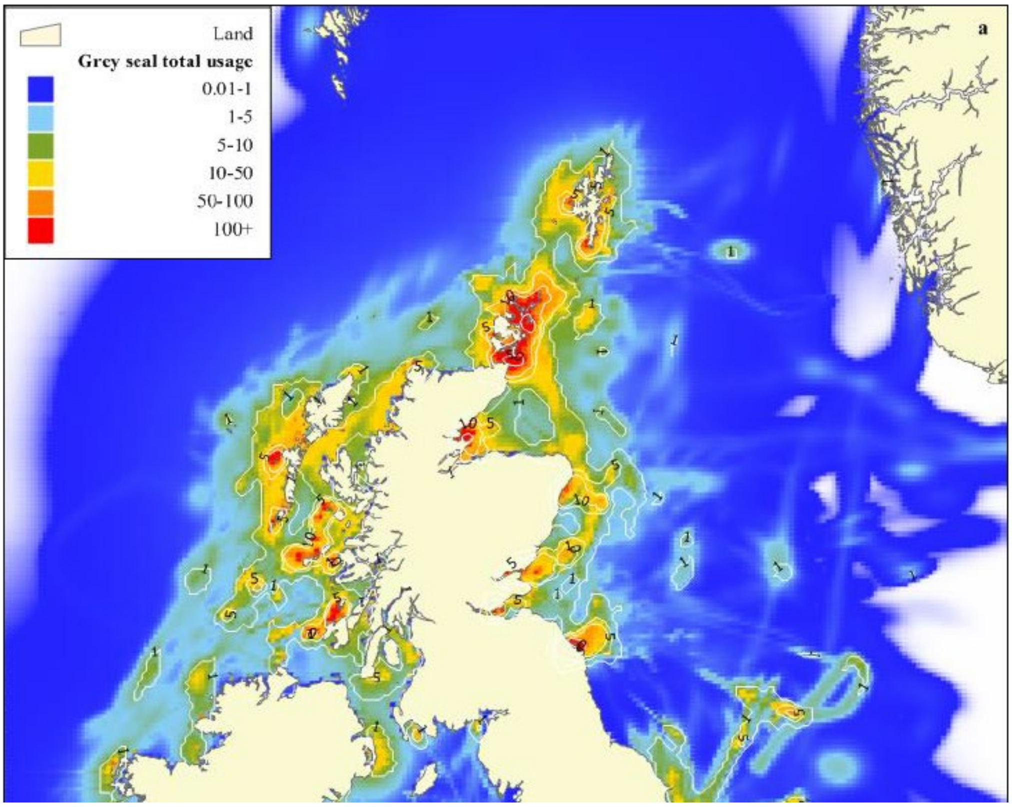


Figure 3.3.19. Areal usage by grey seals around the Scottish coast from Jones *et al*. (2015).

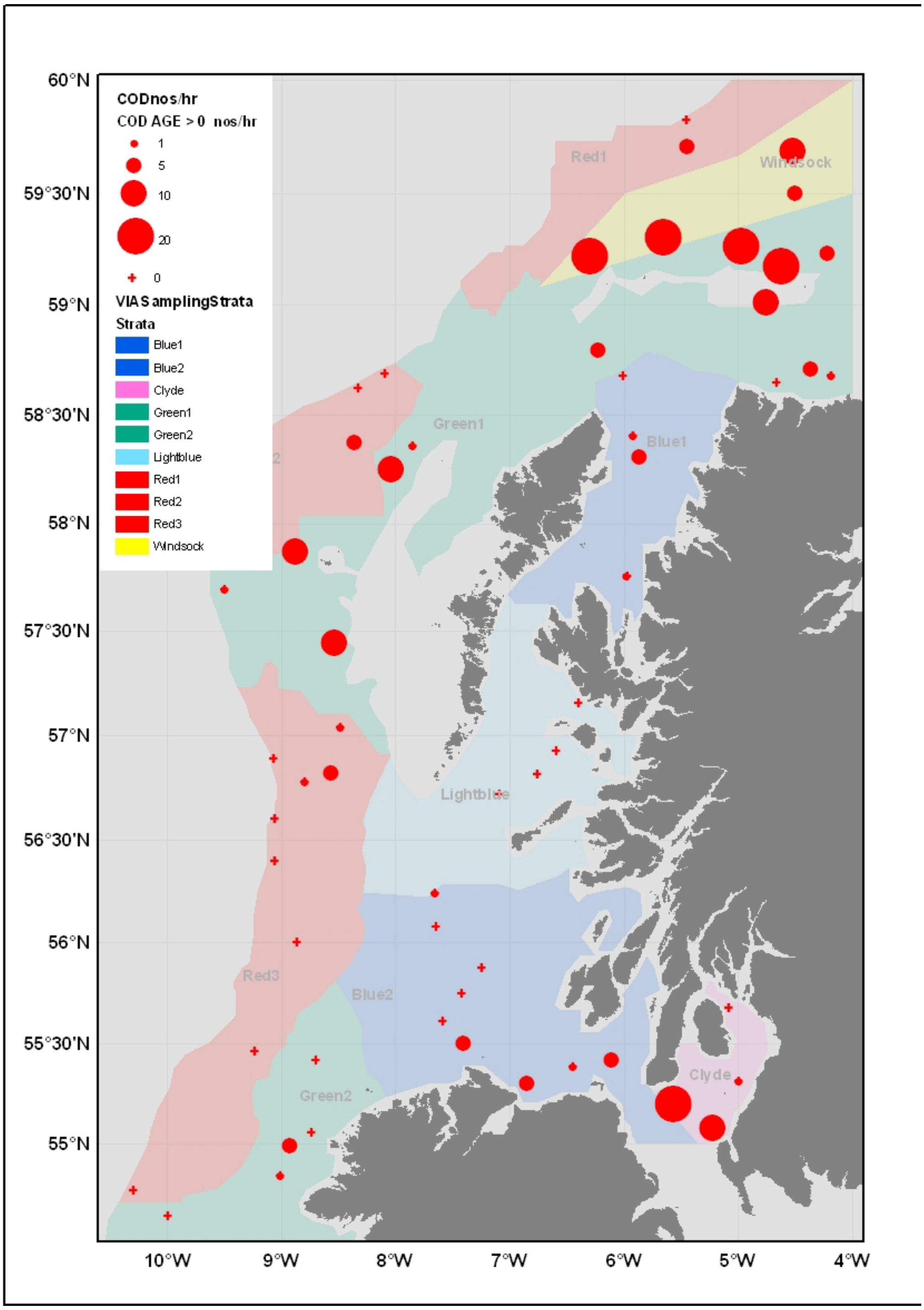


Figure 3.3.20. Cod in 27.6.a. CPUE of 1+ fish in numbers per 60 minute tow for Scottish surveys in 2010 quarter 1.

3.4.5 Reference points

In deriving F_{MSY}, a decision has to be taken about the definition of yield – ICES defines this as catch above MCRS which in the case of a stock with significant high grading (such as 6.a cod) is different to landings. We follow the approach taken at IBPCod6.a (ICES, 2019a), when the current reference points were derived, and define yield as the total catch minus discards where the discard proportion is taken to be the average of the historical discards calculated from well before the change in discarding practices (over 1981—2000).

3.4.6 Input data

The first step in defining reference points is to agree the data to be used in the calculations. The IBP agreed that there was no evidence of a regime change and therefore the full time-series of stock and recruitment data (excluding the final year which depends only on a single datapoint) was used in the estimation of biomass and F reference points (both MSY & PA). (See WD 4.9).

Eqsim provides MSY reference points based on the equilibrium distribution of stochastic projections. Stochasticity is included in biological and fishery parameters by resampling at random from the recent stock assessment. The default year range (most recent ten years) for these data was used (with a check on the sensitivity to mean weights), with the exception of i) using 1981–2000 discard proportions to approximate above MCRS yield and ii) the use of catch mean weights instead of landings mean weights for ages 2 and above to avoid the use of mean landings weight affected by high-grading. The default setting for inclusion of recruitment autocorrelation (TRUE) in the simulations was also used. Given that there is no evidence of significant first order autocorrelation in recruitment (WD 4.9), this setting has no impact on the estimates of reference points.

3.4.7 Defining PA reference points

Following the ICES guidance, the stock is Type 2: a stock with a wide dynamic range of SSB, and evidence that recruitment is or has been impaired. In such cases the ICES guidance suggests B_{lim} is set at the segmented regression change point (Figure 3.3.13). However, there is no clear asymptote in the stock–recruitment data and for this reason the estimated breakpoint occurs at a very high SSB with very wide 95% confidence intervals (19 679 t, CI: 10 857–30 921). As a consequence, this would likely result in a B_{pa} (derived from B_{lim}) at a level close to that at the start of the time-series as the stock was declining from the gadoid outburst. Such an approach may therefore be inappropriate for this stock.

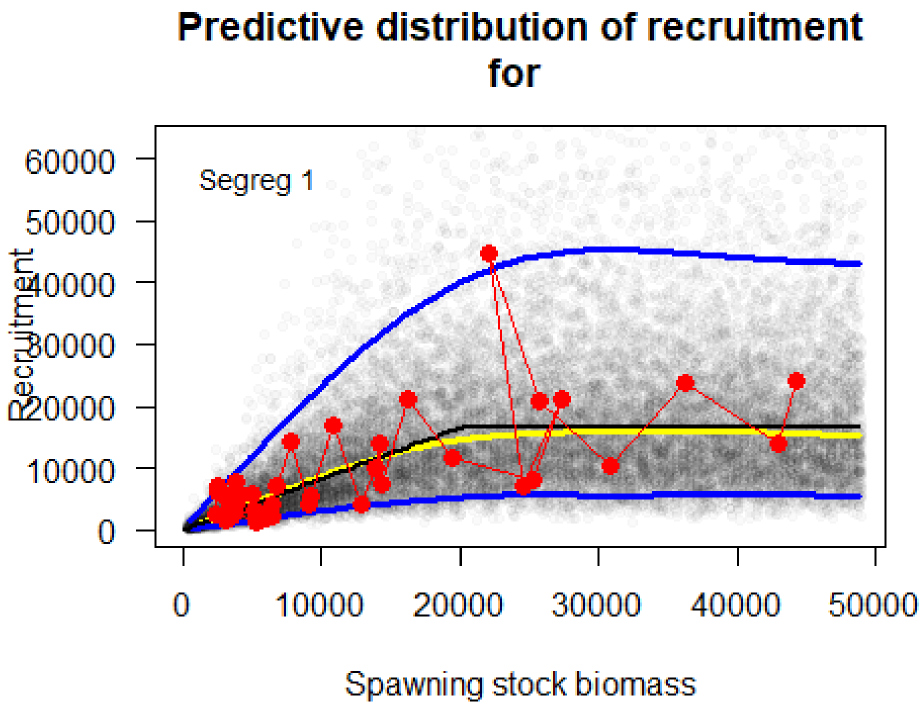


Figure 3.3.13. Cod.27.6a. Stock—recruitment relationship - fitted segmented regression.

As an alternative approach for estimating B_{lim}, we consider the SSB at which low recruitment can be avoided, by looking at the frequency of above/below average recruitments occurring across the range of SSB values. Figure 3.3.14 shows the proportion of SSB points which result in above average recruitment within a moving window (5000 t wide) of SSB intervals (in 500 t increments). Choosing B_{lim} as the value of the lowest SSB with a very high probability of above average recruitment (proportion=1.0), then B_{lim} is equal to 14 500 t. This value would be consistent with either i) retaining the previous B_{lim} or ii) adjusting it to be consistent with the new estimate of the 1992 SSB = 14 376 t. We choose the latter (a point which lies within the confidence bounds of the estimated segmented regression breakpoint) and also update the estimate of B_{pa}, which is now calculated as 20 126 t (based on the typical ICES definition of 1.4 x B_{lim}).

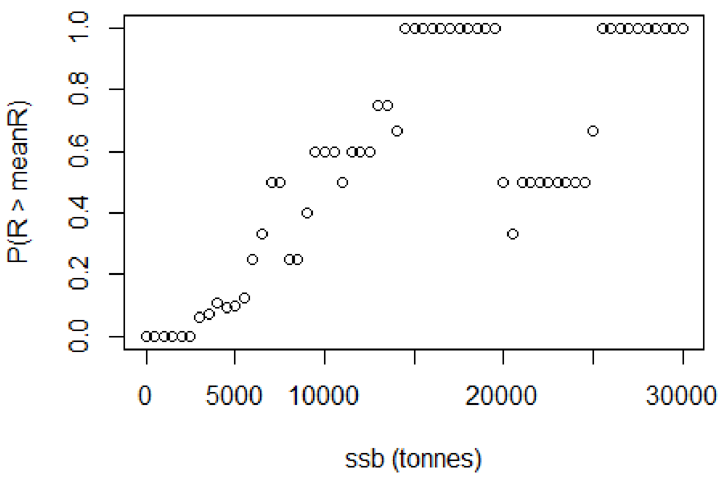


Figure 3.3.14. Cod.27.6a. Proportion of SSB/R pairs with above average recruitment within a moving SSB window 5000 t wide (Point is marked at SSB interval lower bound).

F_{lim} estimation was performed using Eqsim (without assessment/advice error) to derive the F that has 50% probability of SSB falling below B_{lim} using a segmented regression stock–recruitment relationship with the breakpoint fixed at B_{lim}. F_{lim} was estimated as 0.73. F_{pa} was calculated using the default value of σF (0.2) resulting in F_{pa}=F_{lim}/1.4=0.52.

3.4.8 Calculating F_{MSY}

F_{MSY} calculations require the use of a stock–recruitment relationship. The model averaged fit of the Beverton–Holt, Ricker and segmented regression (fixed breakpoint to avoid very high breakpoint) relationships is shown in Figure 3.3.15. In situations where the stock–recruitment relationship is uncertain, the ICES guidance suggests using the model averaging approach in the estimation of F_{MSY}. However, further exploration of the model fits showed that the Beverton–Holt plateau and peak in the Ricker curve occur well outside the range of historical data and therefore these stock–recruit relationships are excluded from the calculation of F_{MSY}.

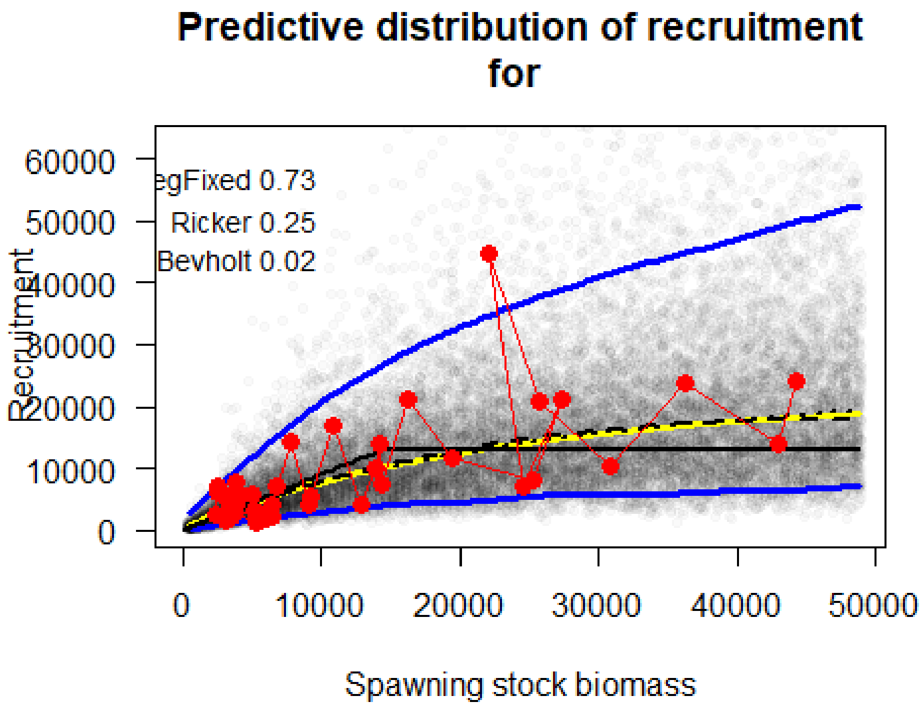


Figure 3.3.15. Cod.27.6a. Stock–recruitment data (red points) with fitted relationship using Ricker (dashed black line), Beverton–Holt (dotted line) and segmented regression (solid black line). Blue lines are 5th and 95th percentiles. Yellow line: 50th percentile.

F_{MSY} is initially calculated by running EqSim with assessment/advice error, but without application of the ICES advice rule (MSY $B_{trigger}$). To include assessment and advice error, the values Fcv=0.212 and Fphi=0.423 (default values suggested by WKMSYREF4 (ICES, 2017)) were used. The median F_{MSY} estimated by Eqsim applying a fixed F harvest strategy was 0.30. The upper bound of the F_{MSY} range giving at least 95% of the maximum yield was 0.49 and the lower bound 0.18.

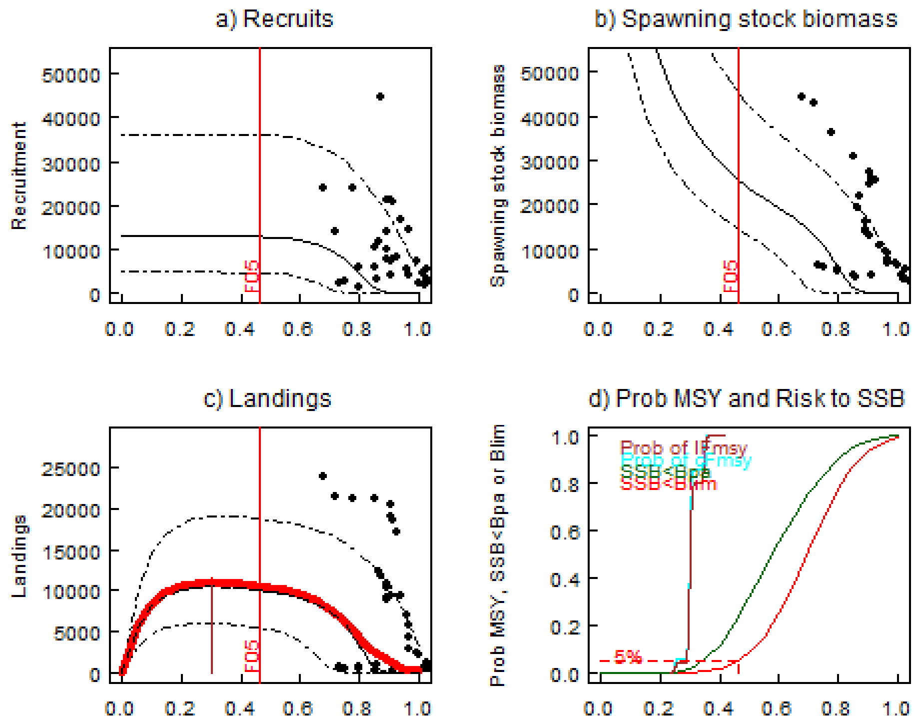


Figure 3.3.16. Cod.27.6a. Eqsim summary plot. Panels a-c: historic values(dots), median (solid black) and 90% intervals (dotted) for recruitment, SSB and yield for exploitation at fixed values of F. Panel c also shows mean yield (red). Panel d shows the probability of SSB less than B_{lim} (red), less than B_{pa}(green) and the cumulative distribution of F_{MSY} based on > MCRS yield (brown) and catch (cyan).

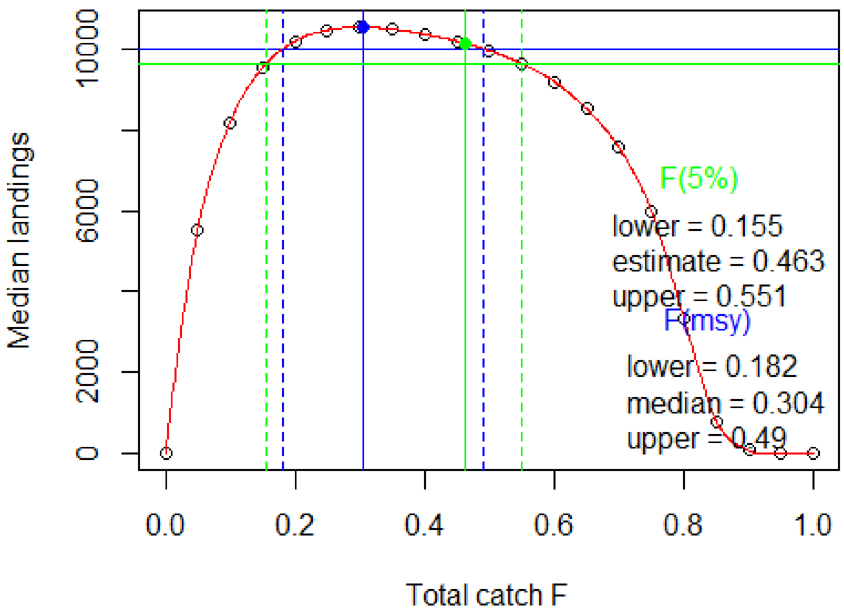


Figure 3.3.17. Cod.27.6a. Median yield curve with estimated reference points for fixed F. Blue lines: F_{MSY} estimate (solid) and range at 95% of maximum yield (dotted). Green lines: $F_{p.05}$ estimate (solid line) and range at 95% of yield at $F_{p.05}$ (dotted line).

The next step is to set MSY $B_{trigger}$. According to ICES guidelines, MSY $B_{trigger}$ is set equal to B_{pa} since the stock has been fished well above F_{MSY} for the last five years. The ICES MSY advice rule is then evaluated to check that the F_{MSY} and MSY $B_{trigger}$ combination fulfils the precautionary criterion of having a less than 5% annual probability of SSB < B_{lim} in the long term. (The evaluation includes assessment/advice error). The $F_{p.05}$ is calculated as 0.57 (see Figure 3.3.18 below) which is greater than the F_{MSY} (and F_{MSY} upper) without the advice rule and therefore the F_{MSY} reference points are not limited by $F_{p.05}$.

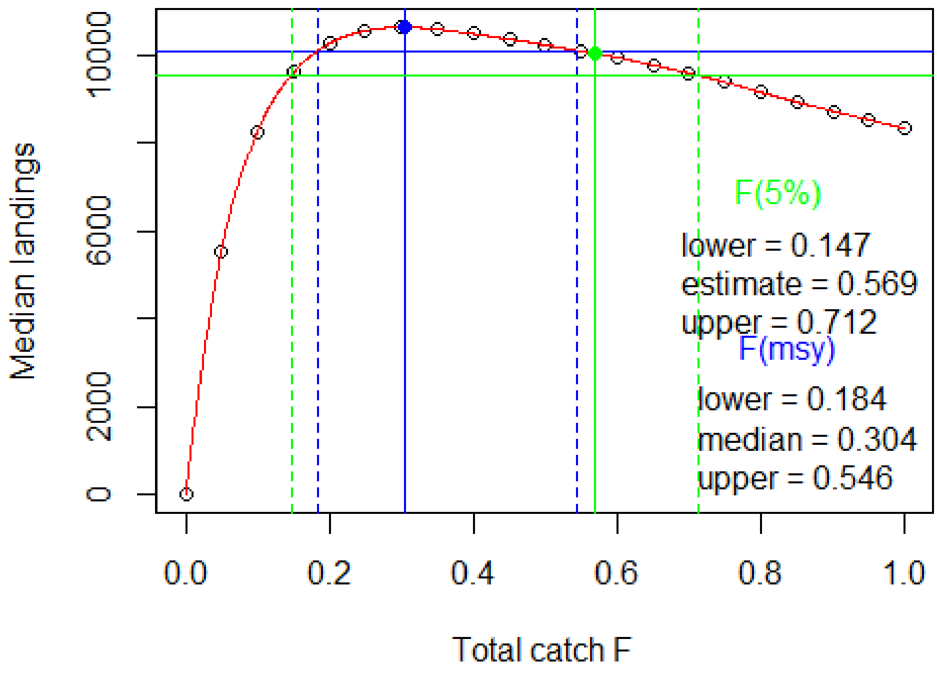


Figure 3.3.18. Cod.27.6a. Median yield curve with estimated reference points when applying the ICES advice rule with $B_{trigger}$=20 000 tonnes. Blue lines: F_{MSY} estimate (solid) and range at 95% of maximum yield (dotted). Green lines: $F_{p.05}$ estimate (solid line) and range at 95% of yield at $F_{p.05}$ (dotted line).

Further outputs from the final EqSim runs and sensitivity testing which shows the estimate of F_{MSY} to be relatively insensitive to the year range of the input biological data can be found in WD 4.9.

3.4.9 General discussion

One of the issues to be addressed at this benchmark meeting was consideration of alternative assessment methods, and given the difficulties with the current implementation of TSA (in that it is slow to converge and difficult to optimise) and likely limited future development support, there was a real motivation for moving to a more practical option. The approaches proposed ahead of the meeting included SAM and a4a, both publicly available assessment approaches which are used widely within ICES. The SAM model was used in the final assessment as opposed to the a4a, largely due to the inability of a4a to appropriately model the years with missing catch data (In a4a, there is currently no option to estimate a catch scaling factor i.e. the model cannot include age compositions without including the absolute values). The TVM model was presented to the meeting as an exploratory assessment and was used to highlight areas of uncertainty with the current assessment, which still remain following this benchmark.

The final SAM model configuration was chosen by comparing model residuals, AIC and retrospective patterns. There remain some patterns in the residuals particularly in the later surveys which are very noisy and the various sensitivity analyses had little impact on these. Similar patterns were evident in previous TSA assessments and the comparison of plots of observed and modelled values from the TVM assessment also suggest a similar lack of fit in some years of the surveys, with this latter model also showing clear bias/residual patterns in both discards and landings. The retrospective analysis in the SAM shows a small amount of overestimation of fishing mortality during the initial years of decline in mean F (although not persistent across all years of the retrospective analysis), which may suggest the model reacts slowly to changes in fishing mortality.

The final SAM assessment estimates that SSB in the stock declined steeply to around 2006 with a slow increase afterwards, followed by another decline in more recent years. The estimated mean F in this assessment remains high until around 2009 and then declines by around 30% over the following years. The results are robust to the assumptions about fishery selectivity, survey catchability and survey data inclusion. They are also consistent with the exploratory a4a assessment runs and the results from the TVM model run in which the fishery selectivity is forced to be flat-topped. In contrast, under other configurations, the TVM model estimates the mean F to be much lower across the whole time-series (indeed much lower than other neighbouring cod stocks such as N Sea cod) and the decline in mean F to begin much earlier and to occur to a much greater degree than in any of the configurations of the other assessment models which have been presented. It also estimates a somewhat greater increase in SSB (although it appears to remain below the putative B_{lim} in all scenarios).

Cook (2019a) previously argued that one would expect the timing and magnitude of the decline in fishing mortality to be more closely related to the decline in fishing effort than seen in the previous TSA assessment and the same could be said for the final SAM assessment presented here. Major decommissioning of fishing vessels active in the Scottish demersal fishery took place in 2001 and 2003, and the reported effort of the main fleets operating in the fishery (TR1 +TR2 from both regulated and unregulated gears, STECF database) declined by around 50% between 2003 and 2015. Yet, despite the reduction in reported effort, the fishing mortality in the SAM assessment does not decline until after 2009. However, given that the fishing mortality is estimated from an assessment which accounts for underreported and area misreported landings it would not be wholly unexpected if the reported effort and estimated fishing mortality were not closely related. Given the known issues with officially reported landings, it is clear that the associated reported effort may not be completely representative of the actual effort in the fisheries in Division 6a. In contrast to the SAM model the decline in estimated mean F in the TVM model is proportionately more in line with the decline in effort although there is a substantial decline before the major decommissioning schemes occur so it is also difficult to conclude that this decline in fishing mortality is due to the reduction in effort.

The area of the main demersal trawl fishery (catching cod) has contracted and operates largely in the northern part of Division 6.a. Research vessel survey data indicate that cod in Division 6.a is now largely confined to the same area and that very large hauls of cod still do occur in this area, suggesting that there are still small areas with a high density of cod. This could potentially explain why the estimated mean F does not decline at the same rate (or to the same extent) as the fishing effort.

Some concern was expressed at the benchmark regarding the credibility of the estimates of survey catchability from the final SAM assessment which show an increase with increasing age for most of the surveys which is consistent with previous TSA assessments. The usual expectation is for survey catchability to be either flat-topped at older ages or to potentially decline as large fish are able to out-swim a short survey tow. However, that is not the case here. Recent Scottish survey data, particularly at older ages, are characterised by a small number of very large hauls and many zeros which would be consistent with aggregating behaviour and hence increasing survey catchability with age. This feature is also evident in the exploratory a4a assessments presented here. It is also apparent in the survey catchability estimates in the SAM assessment of N sea cod (for the NS-IBTS Q1 and Q3 surveys, available on stockassessment.org), so it is not a feature which is unique to the West of Scotland cod assessment (although the North Sea surveys do use a different ground gear so might be expected to have a different catchability).

The differences in estimated fishery selectivity pattern between the exploratory TVM assessments and other assessments presented here appear to be the critical factor in the differences in stock trends. Both flat-topped and dome-shaped selectivity patterns are credible scenarios, and both occur across other stocks. There is some doming estimated in the SAM assessment for the neighbouring North Sea cod, however not to the degree estimated by the TVM in which there is a very steep decline in estimated fishing mortality at older ages with F at-ages 5 and 6 much lower than that at-age 2.

Sensitivity analysis in which the survey catchability was forced to something more 'expected' (i.e. flat-topped) resulted in only minor differences to the final agreed SAM assessment in terms of both estimated stock biomass and fishing mortality. Although the SAM model runs do suggest some change in fishery selectivity pattern later in the time-series, in none of the SAM sensitivity runs was there evidence of significant dome-shaped selectivity. Only in the a4a assessment with forced survey catchability, does the fishery selectivity become significantly dome-shaped and this doming occurs only in the assessment of the period from 2007 onwards (with estimated fishery selectivity in the early period remaining resolutely flat-topped). This is in contrast with the exploratory TVM assessments which show significant doming throughout the assessment period. Given the likely spatial structuring in the age composition of the stock and the relative spatial distribution of the fleets exploiting the stock such an extremely dome-shaped pattern may be plausible. However, without consideration of additional fleet-based catch data it is difficult to conclude which pattern is more likely.

Cod consumption by seals (derived from diet composition studies and seal abundance estimates) is estimated to be 7632 tonnes (95% CI: 3542–13 937) in 2010 (Hammond and Wilson, 2016) compared to a TSB estimate of just under 6000 tonnes from the SAM assessment. Most configurations of the TVM model result in much higher stock biomass more consistent with an estimate of a stock exploited by both seals and the fishery. However, there is uncertainty as to whether the seals are actually exploiting the same population as the fishery. Seal foraging mostly occurs on the continental shelf (Russell *et al.*, 2017) including rocky areas which are unsuitable for trawl fishing and are not surveyed on RV trips, while most of the cod landings are taken along the continental shelf-edge in the north of Division 6a (STECF, 2016) and thus the seals and fishery are largely operating in different areas. Given the complex stock structure and the presence of coastal cod populations (WD 4.1), it is clear there is potential for the seals and fishery to be exploiting different substocks.

The final SAM assessment assumes natural mortality to be a function of stock weight-at-age (Lorenzen, 1996) which are in turn derived from smoothed catch weights-at-age. The time-varying natural mortality estimates from TVM (a function of size with estimated parameters) are substantially higher than the values used in the SAM model and also show extremely high interannual variability across many of the ages. If the predation mortality is largely due to seals as has been previously hypothesized, it is unclear how such interannual variability would be achieved given the relative between year stability of the seal population. Natural mortality clearly remains a major source of uncertainty in this assessment and incorrect assumptions regarding its trend and magnitude can have a significant impact on estimates of stock status.

3.5 Future considerations

3.5.1 Multiple assessment models

There was much criticism of ICES at the benchmark meeting regarding its general approach to providing advice, and its reliance on a single assessment model. ICES stock assessment results generally only report model fit and uncertainty of estimated quantities for the 'agreed' model, ignoring other plausible models and hence ignoring (or not fully quantifying) uncertainty in the results. This seems to be a particular problem for this cod assessment with the exploratory TVM model estimating a greater decline in fishing mortality and a somewhat greater increase in SSB (although likely with the same stock status) than the other models which have been presented. Although sensitivity analysis can be presented in the assessment WG report to demonstrate the how advice might change under different model structures, there is currently no formal method for including this in final advice within the ICES system.

For stocks such as this with significant uncertainty across a range of assumptions including fishery selectivity and survey catchability, it may be appropriate in future to consider a model ensemble approach for the provision of advice. Such an approach accounts for process uncertainty in particular aspects of the model by fitting multiple models and integrating across model results (Millar *et al*., 2015; Scott *et al*., 2015). ICES is planning to explore how such an approach could be operationalised for the provision of advice through WKENSEMBLE which meets later this year.

3.5.2 Multifleet stock assessment

One of the critical issues identified at the IBP conducted in 2019 (and not resolved here) is the fishery selectivity pattern. The commonly used ICES stock assessment models (SAM, TSA, a4a) all estimate a flat-topped fishery selectivity for this stock through most of the assessment time period while the TVM model estimates it to be extremely dome-shaped. Both are considered to be plausible options. Fleet-disaggregated catch-at-age data were requested (and provided) as part of the data call ahead of this benchmark, but due to the challenges of using InterCatch with area misreported landings data, there was limited time available to process the fleet-disaggregated data. The application of a stock assessment in which the main fleets (demersal fish target and *Nephrops* target fisheries) are modelled separately may help to sort out these selectivity issues.

3.5.3 Stock structure

As described in WD 4.1, the stock structure to the West of Scotland is complex and the latest evidence suggests that cod to the west of Scotland are believed to comprise of at least three subpopulations that remain geographically separated throughout the year. A similarly complex picture is apparent in the North Sea and a process is underway at ICES, which it is expected will ultimately result in an assessment for cod in the North Sea which allows for multiple stocks. The current schedule includes a workshop on stock identity planned for summer 2020, with a data call to collate data at the appropriate spatial scale to follow in late 2020 and a benchmark workshop in early 2021. Given the likely biological linkage between the northern offshore component of cod in Division 6a and the northwestern component of North Sea cod, it would make sense to also consider cod in Division 6a cod as part of this process.

3.6 Reviewers' comments

The base run of the Cod 27.6a model was moved to the SAM model, but TSA [the model used in past benchmarks] was available for comparison purposes. The AIC, retrospective pattern, and residuals were improved with the 6a cod benchmark. We support the justification for moving from TSA to SAM and note the minor F retrospective patterns seemed to reduce in the terminal year. We believe the current model is a reasonable model to base advice. The move to SAM was a good decision moving forward due to software help and development and the speed of estimation. The proposed base model run had smoothed time- and age-varying natural mortality based on a Lorenzen curve and time varying weights-at-age. The model had a catch scaling factor in order to account for missing catch information. The stock–recruitment curve was not a functional form, but rather a random walk informed by catch and survey data.

Additional runs were requested to explore the differences between TSA and SAM configurations, as well as explore the influence of assumptions on outcomes of the model. Specifically, the catch scaling factor was explored given some difference between TSA and SAM. The lead analyst checked the years in which the scaling factor was implemented and checked on the ability to over-report catches within the model. Next, SAM was run with different selectivity options; specifically, forcing the survey to have flat-topped selectivity and to estimate dome-shaped selectivity for the fishery.

The review panel agreed with the decisions to use 1995–2006 for the catch scaling time period based on knowledge of the data available. We also agreed with the decision to include increased variance on age-1 and age-7 because the variance is often largest for the youngest and oldest age classes. In addition, there was increased variance in the catch time-series for 2006 to the present to account for a change in legislation in the fishery, which led to more discards being reported than catch. This makes sense considering the technical difficulties with estimating discards. In addition, the appearance of the bubble plots improved with this change with the scale being similar, but the pattern of larger bubbles in the latter part of the time period disappeared.

Finally, the panel compared runs with various selectivity and survey catchability configurations. Fixed flat-top selectivity was chosen because the model independently estimated a flat-top selectivity, which was the run with the lowest AIC value. Specifically, the selectivity was fixed to flat-topped after age-4, which makes sense because age-3 selectivity over time seems to be changing and less than full selection. In addition, estimating ages-5 and -6 catchability for the survey separately is in accordance with the data since there seems to be a signal in the data that the two ages are different.

One general comment for this stock is the apparent issues with defining stock structure. As demonstrated in the supporting documentation and the stock annex, there appears to be a large amount of cod consumed by grey seals every year. This portion, conservatively, is much larger than the estimated SSB of 27.6a cod. There is almost certainly overlap between seals and the 27.6a groundfish fishery, but the discrepancies are noteworthy. We note that for the SAM model, the F trajectory was robust to different assumptions of selectivity (flat-topped vs. domed). It is unclear at this point if stock structure has changed in recent years coinciding with increases in grey seal population/predation. If 6a cod are a single population, it is very likely that only a portion of the stock is being sampled. More likely, because they seem to have responded to changes in catch, biologically they appear to display stock dynamics consistent with independence from the stock that has apparently been feeding the sprawling grey seal population. This uncertainty created debate during the benchmark, particularly in the setting of biological reference points (addressed below).

For the choices regarding reference points, we believe the decisions are sound. The linking of the calculation of reference points to the life history in the most recent time period makes sense [including weights, M, etc.] given that the population is expected to be consistent with the most recent time period. The choice for B_{lim} was systematically done within the context of the ICES guidelines [although see comment above about ICES guidelines]. The determination of F_{MSY} was consistent with the decisions made for B_{lim}.

For the future, the Cod 27.6a assessment should consider the use of fleet-specific catch-at-age to account for spatial dynamics within the population. The lead analyst indicated that the data were recently available, and the exploration of the utility of those data would be worthwhile.

Lastly, a supporting model was presented at the workshop that had different assumptions than the base run and was vetted with simulated data. Differences between the base run and the supporting model were not resolved at the workshop. Several hypotheses were proposed as to why the models were not consistent including selectivity assumptions, natural mortality assumptions, the catch scaling factor, and specification of priors. In addition, we suggest that the estimation model should fit the simulated data better, especially at older ages, in order to make equitable comparisons. If supporting models are brought to the table in the future, then systematic work needs to be done to align the models and determine where/when differences arise. The onus is on the analyst of the supporting model to align their model with the base run. Then, full discussions of assumptions can be undertaken.

3.7 References

Cook, R. 2019a. Stock collapse or stock recovery? Contrasting perceptions of a depleted cod stock. ICES Journal of Marine Science, 76 (4): 787–793. https://doi.org/10.1093/icesjms/fsy190.

Cook, R. M. 2019b. Inclusion of discards in stock assessment models. Fish and Fisheries, 20:1232—1245.

Cook, R. M., Holmes, S. J. and Fryer, R. J. 2015. Grey seal predation impairs recovery of an over-exploited fish stock. Journal of Applied Ecology, 52(4), 969—979.

Doyle, A., Davie, A., Wright, P., Coull, K. and C Angus. 2016 Determination of the Distribution of the Resident Inshore and Offshore Migratory Cod Populations around Shetland (IVa) and Westwards into VIa. Fishing Industry Science Alliance (FISA) Project 01/13 Scottish Marine and Freshwater Science Vol 7 No 28.

Fryer, R. 2001. TSA: Is it the way? WD to the Working Group on Methods on Fish Stock Assessment, 2001. ICES CM 2002/D:01

Fryer, R. 2011. TSA: current implementation. WD to the Workshop on the Analysis of the Benchmark of Cod in Subarea IV (North Sea), Division VIId (Eastern Channel) and Division IIIa (Skagerrak) (WKCOD 2011), 7–9 February 2011, Copenhagen, Denmark. ICES CM 2011/ACOM:51.

Gudmundsson, G. 1994. Time-series Analysis of Catch-at-age Observations. Journal of Applied Statistics, 43:117–126. ICES. 2012.

Hammond, P.S. and Wilson, L.J. 2016. Grey Seal Diet Composition and Prey Consumption. Scottish Marine and Freshwater Science Report Vol 7 No 20. Doi: 10.7489/1799-1.

Holmes, S.J., Miller, C.P., Fryer, R.J. and Wright, P.J. 2014. Gadoid dynamics: differing perceptions when contrasting stock vs. population trends and its implications to management. ICES Journal of Marine Science (2014), 71(6), 1433–1442.

ICES. 1997. Report of the Working Group on the Assessment of Northern Shelf Demersal Stocks. ICES Headquarters, Copenhagen, Denmark. 26 June–5 July 1996.

ICES. 2008. Report of the Workshop on Maturity Ogive Estimation for Stock Assessment (WKMOG), 3–6 June 2008, Lisbon, Portugal. ICES CM2008/ACOM:33. 72 pp.

ICES. 2012. Report of the Benchmark Workshop on Western Waters Roundfish (WKROUND), 22–29 February 2012, Aberdeen, UK. ICES CM 2012/ACOM:49. 283 pp.

ICES. 2015. Report of the Inter-Benchmark Protocol of West of Scotland Roundfish (IBPWSRound), February–April 2015. By correspondence. ICES CM 2015/ACOM:37. 99 pp.

ICES. 2017. Report of the Workshop to consider F_{MSY} ranges for stocks in ICES categories 1 and 2 in Western Waters (WKMSYREF4), 13–16 October 2015, Brest, France. ICES CM 2015/ACOM:58. 187 pp.

ICES. 2019a. Inter-benchmark Workshop on West of Scotland Cod (6.a) (IBPCod6.a). ICES Scientific Reports. 1:13. 171 pp. http://doi.org/10.17895/ices.pub.4976.

ICES. 2019b. Working Group for the Celtic Seas Ecoregion (WGCSE). ICES Scientific Reports. 1:29. 1604 pp. http://doi.org/10.17895/ices.pub.4982.

Jardim, E., Millar, C. P., Mosqueira, I., Scott, F., Osio, G. C., Ferretti, M., Alzorriz, N., and Orio, A. 2015. What if stock assessment is as simple as a linear model? The a4a initiative. ICES Journal of Marine Science, 72: 232–236.

Jones, E.L., McConnell, B.J., Smout, S., Hammond, P.S. Duck, CD, Morris, C, Thompson, D, Russell, DJF, Vincent, C, Cronin, M, Sharples, RJ and Matthiopoulos, J. 2015. Patterns of space use in sympatric marine colonial predators reveal scales of spatial partitioning. Marine Ecology Progress Series 534:235–249. https://doi.org/10.3354/meps11370.

Lorenzen K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49, 627–647.

Millar, C. P., Jardim, E., Scott, F., Osio, G. C., Mosqueira, I., and Alzorriz, N. 2015. Model averaging to streamline the stock assessment process. ICES Journal of Marine Science, 72; 93–98.

Nielsen, A. and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models, Fisheries Research, 158, 96–101.

Russell, D. J. F., Jones, E. L. and Morris, C. D. 2017. Updated Seal Usage Maps: The Estimated at-sea Distribution of Grey and Harbour Seals. Scottish Marine and Freshwater Science Report Vol 8 No 25.

Scott, F., Jardim, E., Millar, C. P., and Cervino, S. 2016. An Applied Framework for Incorporating Multiple Sources of Uncertainty in Fisheries Stock Assessments. PLoS ONE 11(5):e0154922. Doi:19.1371/journal.pone.0154922.

STECF. 2016. Scientific, Technical and Economic Committee for Fisheries (STECF) – Fisheries Dependent Information (STECF-16-20); Publications Office of the European Union, Luxembourg; EUR 27758 EN; doi:10.2788/502445.

4 Whiting (*Merlangius merlangus*) in Division 6.a (West of Scotland) (Whg.27.6a)

4.1 Why a benchmark?

The last benchmark for this stock was carried out in February 2012 (ICES, 2012) with a subsequent inter-benchmark (IBP) in February 2015 (ICES, 2015). With new data being gathered in the following years, it was deemed proper to conduct a new benchmark assessment.

The main problem with assessment of this stock is related to the commercial data (ICES, 2019a). Incorrect reporting of landings (species and quantity) is known to have occurred in the past and directly affecting the perception of the stock. The assessments conducted in 2012–2019 used the Time-Series Analysis (TSA, Gudmundson, 1994; Fryer, 2002; Needle and Fryer, 2002). The model allowed omission of unreliable catch data for 1995–2005 and use of age structure data only from the catch. In the past, the survey data and commercial catch data contained contradicting signals concerning the stock. This was particularly the case with the discontinued Scottish surveys before 2011. The TSA assessments allowed the model to interpret the mismatch in the signals from the two data sources through a persistent trend (increase) in survey catchability. One of the goals of this benchmark was to conduct a sensitivity analysis with and without a survey catchability trend.

The assessments conducted in 2015–2019 used five tuning time-series. At the inter-benchmark in 2015, the option was considered of using one combined index for the two Q4 surveys, the Irish and Scottish ones (ICES, 2015). One rationale for combining the two indices was the fact that the Irish survey is mainly limited to the southern part of Division 6a and it is unclear to what extent the index represents the population size. A single (combined) index would thus be more representative of the whole stock. The first attempt to combine the two surveys was done at WGISDAA 2018 (ICES, 2018) and an update with recent data was done for this benchmark.

The benchmark for whiting in 6.a, uses catch data recently uploaded to InterCatch which is hoped to improve the accuracy of the whiting assessment. In addition, a new maturity ogive and updated estimates of natural mortality varying in time are proposed to be used in subsequent assessments of the stock.

4.2 Summary of final model

Previous assessments of whiting in Division 6.a were conducted with TSA. At that time, the stock was classified as category 1. The updated information on the biological parameters (maturity and natural mortality) was meant to be used as input to the TSA model.

During the benchmark process, it became clear that running TSA with the new data and changed survey configuration posed a serious challenge. Poorly converged optimisation runs (with some parameters being found on the boundary of the assumed parameter space) in conjunction with excessive running times, were a major obstacle to complete the assessment successfully. In these circumstances, it was decided *ad hoc* to run the benchmark assessment using an alternative method, the age-aggregated stochastic Surplus Production in Continuous Time (SPiCT) model (Pedersen and Berg, 2017). At the same time, the stock was downgraded to category 3 according to the ICES guidelines for data-limited stocks (ICES, 2019b).

The historical and newly available catch data for whiting in 6.a, as well as survey indices (including the combined Irish and Scottish Q4 survey) were made available for the assessment. The summary of these data is given below:

Data type	Year range	Note
Catch numbers-at-age	1978–2018	As reported catch data
Catch weights-at-age	1978–2018	As reported catch data
ScoGFS-WIBTS-Q1 (survey)	1985–2010	Scottish survey Q1 (old)
UK-SCOWCGFS-Q1 (survey)	2011–2018	Scottish survey Q1 (new)
ScoGFS-WIBTS-Q4 (survey)	1996–2009	Scottish survey Q4 (old)
IGFS-UK-SCOWCGFS-Q1 (survey)	2011–2018	Irish and Scottish survey Q4 (combined index)
IGFS-WIBTS-Q4 (survey)	2003–2010	Irish survey Q4 (truncated time-series)

To be used as input to the SPiCT model, catch numbers-at-age were multiplied by catch weights-at-age giving the total catch biomass (only for age groups 1+ that are vulnerable to commercial fleets).

For each survey time-series, the indices for numbers-at-age per 10 hours (for age groups 1+) were converted to survey biomass-at-age per 10 hours using catch weights-at-age. The latter were assumed to represent stock weights-at-age. Finally, stock biomass-at-age per 10 hours were summed up giving the total survey biomass per 10 hours. The complete input data to be run with the SPiCT model are shown in Table 4.1.

Different model scenarios were tested to obtain robust results in terms of relative fishing mortality and relative biomass (Table 4.2). The uncertainty of these estimates was generally very high, but the model performed considerably better with the assumption of a Schaefer model, i.e. with the shape parameter n fixed to be equal to 2 (details of the different tested scenarios and tested model settings can be found in WD 5.4). Among the tested survey configurations, the one with these four surveys:

- old Scottish Q1 survey,
- new Scottish Q1 survey,
- combined Irish and Scottish Q4 survey, and
- truncated Irish Q4 survey.

performed best, particularly in the retrospective analysis. Inclusion of the old Scottish Q4 survey seemed to cause some noise, while the truncated Irish Q4 survey time-series, being relatively short, still seemed to provide useful information on the dynamics of the stock.

The final run (under Scenario 4b) used the full catch time-series from 1978–2018 and the above-mentioned four survey time-series (Figure 4.1). A Schaefer model was assumed. For the period 1995–2005 (with unreliable catch data), the uncertainty of catch biomass was scaled by a factor of 4. The results of the final run are summarised in Table 4.3 and shown in Figure 4.2. The relative biomass (B_t/B_{MSY}) was less than 1. The relative fishing mortality (F_t/F_{MSY}) was very low with very wide confidence intervals. No strong residual patterns, no issues with autocorrelation and no serious violation of the assumption of normality were found for the catch or indices (Figure 4.3). In the final run, the model converged successfully in all the years of the retrospective analysis also showing generally little retrospective pattern (Figure 4.4).

4.3 Investigations undertaken (summary)

4.3.1 Catch data

The data imported from InterCatch included national landings and discards, and their age composition for the period 2003–2018. This was preceded by a data call for such information from the countries involved in the whiting fishery in Division 6.a.

The approach adopted and approved at this benchmark for the catch estimation was to allocate sampled catch data to unsampled catches and to aggregate all catch data, separately for landings and discards with a distinction being made between TR1 and TR2 fleets. Details of this aggregation are presented in WD 5.1.

The estimated catch and catch weights (by catch category) did not differ substantially from those reported earlier (ICES, 2019a). They are shown in Figures 4.5 and 4.6 and in Tables 4.4–4.10.

4.3.2 Surveys

Up to now, five tuning time-series have been used in the assessment of whiting in 6.a. At this benchmark, it was decided to replace the index from the new Scottish Q4 survey (run since 2011) with a combined index for the two Q4 surveys, the Irish and Scottish ones. For this purpose, an extensive analysis was carried out with survey data for 2011–2018. Details of this analysis and its results are presented in WD 5.2.

For the GAM analysis, hauls were selected that were taken in the southern part of Division 6.a, considered as the "common area" for the two surveys. A statistical model (a negative binomial GAM for counts with a log link function) was used to estimate catch rates. This was run separately for age groups (0, …, 6 and 7+), as well as for an additional aggregate age group (1+). The raised numbers were the response variable in the model. The explanatory variables included vessel, year, time of day, depth, longitude and latitude. The log-transformed tow duration term was added to the model as an offset (Zuur *et al.*, 2009). Subsequently, a chi-square test was conducted with the full model and the reduced model (without the 'vessel' variable) to establish the significance of the difference between the two surveys. The effect of vessel was quite variable across the different age groups. While the catch rates tended to be higher in the Irish survey for nearly all age groups (notably for fish at-age 3–5), for fish at age 6, they were significantly higher in the Scottish survey. In relative terms, the ratio of catch rates in the Scottish survey to those in the Irish survey was 0.6–1.8, depending on the age group. For the aggregate age group 1+, it was 0.7.

To combine the two surveys, all hauls taken in area 6.a were initially considered to form one dataset. The age frequencies were calculated in the same way for each haul, irrespective of the survey. As the next step, the frequencies in the Irish hauls were modified by using the ratios of catch rates established earlier through the GAM analysis. As a result, the final modified index could be derived. Figure 4.7 compares the three tuning series in operation in 2011–2019 (Irish Q4 surveys and new Scottish surveys for Q1 and Q4) with the combined Irish and Scottish Q4 survey.

The combined index appeared to produce less noise and be more informative of the population densities (see WD 5.2). It was used (with other survey time-series) in the final assessment with SPiCT but can be used in assessments with other methods as well.

4.3.3 Assessment model

As mentioned earlier, different scenarios were run with the SPiCT model and data from Table 4.1 as the input. All these runs converged (disregarding retrospective runs). Table 4.2 summarises the tested scenarios.

Scenarios 1a–c were run with the data from 1985 to 2018, i.e. for the years with a survey. Scenarios 1a and 1b used default priors: distributions for the shape parameter n and the two uncertainty ratios α and β. The uncertainty of the assessment was found to be very high, also causing problems in estimating some of the derived quantities. In Scenario 1c, the modification with a Schaefer model resulted in slightly smaller confidence bands. Also, the derived quantities previously inestimable could now be estimated. However, the retrospective analysis showed that this was not the case for all the preceding years (especially with respect to the relative biomass and relative fishing mortality).

Scenarios 2a–d were run with the full catch time-series (1978–2018) and priors as in Scenario 1c. There were small differences in the assessed quantities among Scenarios 2a–d. In Scenarios 2a (with all the five survey time-series) and 2c (with four surveys that included the old Q4 survey), the model converged but for one year in the retro analysis it failed to converge. In Scenarios 2b (with four surveys) and 2d (with three surveys), which both did not include the old Q4 survey, the model converged, and it did so in all the years of the retrospective analysis. The output of these two scenarios is shown in WD 5.4, Annex 1.

Scenarios 3a–d were run with the truncated catch time-series for the years 1978–1994 and 2006–2018, and with the survey configurations analogous to Scenarios 2a–d. There were small differences in the estimated quantities among Scenarios 3a–d, and generally between Scenarios 2 and 3, but the retrospective analysis was successful, for all the retrospective years, only for Scenario 3b (shown in WD 5.4, Annex 1).

In Scenarios 4a–d (also with the survey configurations as in Scenarios 2a–d or 3a–d), account was taken of the period 1995–2005 with unreliable catches. The uncertainty for this period was incorporated in the model; in this case, it was scaled by a factor of 2–5. Among Scenarios 4a–d, only Scenario 4b provided a model that converged successfully in all the years of the retrospective analysis. Based on the model's performance in the retrospective analysis, showing a low between-assessment variation in estimates of the relative biomass and relative fishing mortality, the scaling factor of 4 was selected. The model in Scenario 4b with this scaling factor was the final model.

4.3.4 Maturity ogive

Maturity and natural mortality were modified in the benchmark process providing more accurate estimates. The maturity ogive had previously been assumed to be knife-edge, with the value 0 at-age 1 and full at-age 2+. An analysis of Scottish survey data conducted at this benchmark, following the guidelines from the ICES WKMOG report (ICES, 2008) showed no clear temporal trends in maturity (more details are presented in WD 5.3). The analysis provided coefficients of the logistic model: -6.165 (intercept) and 5.103 (slope) model (Figure 4.8). The midpoint of the modelled maturity ogive, A50, was estimated to be 1.208 (±0.033) years. The logistic model for the ogive was considered to sufficiently accurate for future assessments of whiting in 6.a, especially when contrasted with the earlier knife-edge assumption (see the table below).

Age	1	2	3	4	5	6	7+
WGCSE 2019	0	1	1	1	1	1	1
WKDEM	0.257	0.983	1	1	1	1	1

4.3.5 Natural mortality

In previous assessments of whiting in Division 6.a, natural mortality was assumed to vary and be dependent on fish weight (Lorenzen, 1996). Mean catch weights-at-age (assumed to be equivalent to stock weights-at-age) over the whole time-series were used in these calculations. As observed in annual assessments of the stock, catch weights-at-age of whiting in 6.a show some temporal trends. Therefore, at this benchmark, it was agreed to use smoothed catch weights-at-age which follows the method applied to cod in Division 6.a (see the section for cod in 6.a of this benchmark report). The resulting catch weights-at-age (and stock weights-at-age for that matter) were subsequently used to estimate natural mortality rates using the Lorenzen (1996) function. These estimated are shown in Table 4.10 and in Figure 4.9.

4.4 Future considerations

Given the unsuccessful assessment of whiting in 6.a with TSA, and highly uncertain assessment with the SPiCT model, there is a compelling need to consider alternative assessment methods for this stock. Using an age-based model such as SAM provides such an opportunity. This can be achieved through extensive exploration and testing of the model before this method is accepted. Work on this can be progressed immediately as the data requirements are the same as for the previous TSA assessment. The aim should be to develop a satisfactory preliminary SAM assessment and schedule for re-benchmarking (in order to return to Category 1) as soon as practically possible.

Given that for the time being this stock is now considered data-limited, it would be worth processing the catch length composition data which are available in InterCatch for 2003 onwards. This would enable the use (or at least) consideration of length-based methods in the assessment/advisory process. In addition, given that there are extensive survey data for this stock, it may also be useful to consider survey-based assessment such as SURBA.

New methods should be considered for calculation of survey indices. Those currently in use do not take into account important explanatory variables. Properly standardised indices (e.g. with GLMs or GAMs) could improve accuracy and precision of the data inputs to whiting assessments.

4.5 Reviewers' comments

This stock was benchmarked in 2012, and the last inter-benchmark was in 2015. Going into the 2020 benchmark, we evaluated this stock with an assessment using the TSA platform tuned with five survey indices.

The Whiting 27.6.a base run that was proposed at the beginning of the workshop exhibited some concerning behaviours. First, the stock–recruitment curve did not appear to fit the data well, and other options for configuration needed to be considered. Second, the catchability for the two longest survey time-series was increasing substantially over time, but no explanation was provided as to why that might be the case. Based on the stock and fishery dynamics, we do not find this reasonable. These two topics consumed most of the week for this assessment.

The assessment review panel agreed to have the following additional runs completed at the workshop:

1. Truncated time-series (2000–present);
2. Dropping 2 surveys [2 from earliest part of the time-series];
3. Constant catchabilities on all surveys;
4. Fixed age at maximum fishery selection [move from age-4 to age-6];
5. 4 and 3 combined;
6. S–R curve choice-based on the runs with constant catchability, the options provided in TSA did not allow for a stock–recruitment function that wasn't dependent upon the entire time-series.

In the end, the currently proposed TSA model was not performing well. Specifically, the model was taking longer to run than normal and convergence criteria were not always met. Several of the runs requested during the meeting did not reach convergence. The time it took for TSA to search for convergence suggests difficulty in searching the parameter space, meaning that the search algorithm is not efficient. Based on the long run-time, over the week a retrospective analysis was not produced. The lead analyst indicated that two parameters were reaching bounds, which was likely causing the convergence issues. However, by that time in the meeting and given the time it takes to run TSA for this species, the review workshop needed to develop a new plan of action for the assessment.

The first model (TSA) was rejected. The benchmark reviewers tried to salvage a SPiCT model for the 27.6.a stock.

At the time of the benchmark, we were unable to come up with reference points. However, it should in theory be possible to provide length-based reference points at a later date. We should emphasize that the litany of issues with this stock were disappointing, and probably to some extent could have been avoided with more work prior to the benchmark meeting. We expect that as the newer surveys accumulate a longer time-series, there will likely be better model specification. The final SPiCT model exhibited a retrospective pattern, though the trends of the peels were similar. We should note that the retros were also present and accepted as a category 1 stock after the last inter-benchmark. These are most likely due to the recent introduction of multiple surveys and the model not being able to track a period of less than a cohort. Again, as survey duration increases, theoretically, this will become less of an issue. Regardless, consistency and use of surveys should be thoroughly evaluated for this stock going forward.

This moved the stock into Category 3 within ICES. The surplus production model wasn't able to provide estimates with confidence intervals that informed the stock status but was able to provide overall trend information.

While the assessment had complexities that could not be overcome during the review workshop meeting, the data for the stock are informative and this should be pursued as a Category 1 stock in the future. This assessment would benefit from a platform that was more flexible, was supported with development, and with a faster solving/computing time. It is highly recommended that the data be moved into another assessment platform. The signals within the survey data seem to be congruent and should provide information on the trajectory of the stock. We agree with the combined survey index for heavily overlapping areas; survey indices should be combined when measuring the same segment of the population.

Based on the downgrading of this stock, we recommend the following guidelines for the next 27.6.a whiting assessment:

1. Assign more than one person to assess Whiting 27.6a.
2. Move model into SAM or an alternative structure that is supported and meets the needs of the data for the species.
3. Finalize survey data to be used, if necessary; five surveys indices were used in the base run presented here; we agree with the combined survey index that covers the same area and quarter.
4. Explore the fishery selectivity age at full selection [ages-4 and -6 were explored, but also explore other ages and choose based on biology and data fits].
5. Explore survey catchability options within the model [including constant and time varying as a random walk].
6. Explore stock–recruitment curve options in the model [include a random walk option that isn't reliant upon historical mean values].
7. Run sensitivity analyses to explore options for natural morality [be sure to include a constant value, size based (e.g. Lorenzen), age and time varying].
8. Run leave one out index runs as a diagnostic.
9. Run retrospective analysis.
10. Move forward by comparing changes in the fits to the data [residuals], AIC, considering biology and ecosystem dynamics, etc.

4.6 References

Fryer, R. J. 2002. TSA: is it the way? Appendix D in report of Working Group on Methods on Fish Stock Assessment. ICES CM 2002/D:01.

Gudmundsson, G. 1994. Time-series analysis of catch-at-age observations. Applied Statistics 43: 117–126.

ICES. 2008. Report of the Workshop on Maturity Ogive Estimation for Stock Assessment (WKMOG), 3–6 June 2008, Lisbon, Portugal. ICES CM2008/ACOM:33. 72 pp.

ICES. 2012. Report of the Benchmark Workshop on Western Waters Roundfish (WKROUND), 22–29 February 2012, Aberdeen, UK. ICES CM 2012/ACOM:49, pp. 283.

ICES. 2015. Report of the Inter-Benchmark Protocol of West of Scotland Roundfish (IBPWS-Round), February–April 2015, By correspondence. ICES CM 2015/ACOM:37. 72 pp.

ICES. 2018. Report of the Working group on improving use of Survey Data for Assessment and Advice (WGISDAA). 3–5 July 2018. Copenhagen, Denmark. ICES CM 2018/EOSG: 15. 29 pp.

ICES. 2019a. Working Group for the Celtic Seas Ecoregion (WGCSE). ICES Scientific Reports. 1:29. 1604 pp. http://doi.org/10.17895/ices.pub.4982.

ICES. 2019b. Advice basis. In Report of the ICES Advisory Committee, 2019. ICES Advice 2019, Book 1, Section 1.2. https://doi.org/10.17895/ices.advice.5757.

Needle, C. L. and Fryer, R. J. 2002. A modified TSA for cod in Division VIa: separate landings and discards. Working document to the ICES Advisory Committee on Fisheries Management, October 2002.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., and Smith, G. M. 2009. Mixed Effects Models and Extensions in Ecology with R. Springer, New York, 574 pp.

Table 4.1. Whiting in Division 6.a. Input data for the SPiCT model.

Year	Land-ings(tonnes)	Discards (tonnes)	Total catch (tonnes)	Scottish Q1 old (kg/10 h)	Scottish Q1 new (kg/10 h)	Scottish Q4 old (kg/10 h)	Scottish and Irish Q4 combined (kg/10 h)	Irish Q4 truncated (kg/10 h)
1978	14677	4669	19346					
1979	17081	3019	20100					
1980	12816	1782	14598					
1981	12203	2132	14335					
1982	13871	5485	19356					
1983	15970	6294	22264					
1984	16458	4017	20475					
1985	12893	4840	17733	955				
1986	8454	2669	11123	610				
1987	11544	11918	23462	1144				
1988	11352	8132	19484	895				
1989	7531	5876	13407	453				
1990	5643	4530	10173	517				
1991	6660	4883	11543	472				
1992	6004	9249	15253	1164				
1993	6872	4759	11631	2341				
1994	5901	3455	9356	1477				
1995	6076	5771	11847	1358				
1996	7156	7940	15096	1731		694		
1997	6285	5251	11536	1767		709		
1998	4631	9216	13847	1438		581		
1999	4613	3975	8588	1651		380		
2000	3010	13285	16295	1134		516		
2001	2438	4263	6701	909		647		
2002	1709	2851	4560	906		383		
2003	1331	1697	3029	981		551		1083
2004	799	2679	3477	932		389		354

Year	Land-ings(tonnes)	Discards (tonnes)	Total catch (tonnes)	Scottish Q1 old (kg/10 h)	Scottish Q1 new (kg/10 h)	Scottish Q4 old (kg/10 h)	Scottish and Irish Q4 combined (kg/10 h)	Irish Q4 truncated (kg/10 h)
2005	335	753	1087	275		96		246
2006	378	531	908	275		89		127
2007	481	314	795	225		179		548
2008	441	140	582	115		94		526
2009	480	412	892	267		122		557
2010	338	988	1326	272				381
2011	229	246	474		573		509	
2012	304	711	1015		534		429	
2013	216	1139	1355		548		389	
2014	181	498	678		468		467	
2015	223	836	1059		572		612	
2016	226	851	1078		1319		987	
2017	178	1044	1222		1484		1100	
2018	190	552	742		1277		447	

Table 4.2. Whiting in Division 6.a. Different scenarios tested with the SPiCT model.

Scenario	Catch time-series	Survey	Survey time-series	Setting	Comment
1a*	1985–2018	Scottish Q1 (old) Scottish Q1 (new) Scottish Q4 (old) Irish & Scottish Q4 (combined) Irish Q4 (truncated)	1985–2010 2011–2018 1996–2009 2011–2018 2003–2010	Default priors	High uncertainty; some derived quantities inestimable; not converged for some years in retro
1b*	1985–2018	Scottish Q1 (old) Scottish Q1 (new) Irish & Scottish Q4 (combined)	1985–2010 2011–2018 2011–2018	Default priors	High uncertainty; some derived quantities inestimable; not converged for some years in retro
1c*	1985–2018	Scottish Q1 (old) Scottish Q1 (new) Irish & Scottish Q4 (combined)	1985–2010 2011–2018 2011–2018	Fixed n=2	High uncertainty; not converged for some years in retro
2a	1978–2018	Scottish Q1 (old) Scottish Q1 (new) Scottish Q4 (old) Irish & Scottish Q4 (combined) Irish Q4 (truncated)	1985–2010 2011–2018 1996–2009 2011–2018 2003–2010	Fixed n=2	High uncertainty; not converged for some years in retro
2b*	1978–2018	Scottish Q1 (old) Scottish Q1 (new) Irish & Scottish Q4 (combined) Irish Q4 (truncated)	1985–2010 2011–2018 2011–2018 2003–2010	Fixed n=2	High uncertainty
2c	1978–2018	Scottish Q1 (old) Scottish Q1 (new) Scottish Q4 (old) Irish & Scottish Q4 (combined)	1985–2010 2011–2018 1996–2009 2011–2018	Fixed n=2	High uncertainty; not converged for some years in retro
2d*	1978–2018	Scottish Q1 (old) Scottish Q1 (new) Irish & Scottish Q4 (combined)	1985–2010 2011–2018 2011–2018	Fixed n=2	High uncertainty

Scenario	Catch time-series	Survey	Survey time-series	Setting	Comment
3a	1978–1994, 2006–2018	Scottish Q1 (old) Scottish Q1 (new) Scottish Q4 (old) Irish & Scottish Q4 (combined) Irish Q4 (truncated)	1985–2010 2011–2018 1996–2009 2011–2018 2003–2010	Fixed n=2	High uncertainty; not converged for some years in retro
3b*	1978–1994, 2006–2018	Scottish Q1 (old) Scottish Q1 (new) Irish & Scottish Q4 (combined) Irish Q4 (truncated)	1985–2010 2011–2018 2011–2018 2003–2010	Fixed n=2	High uncertainty
3c	1978–1994, 2006–2018	Scottish Q1 (old) Scottish Q1 (new) Scottish Q4 (old) Irish & Scottish Q4 (combined)	1985–2010 2011–2018 1996–2009 2011–2018	Fixed n=2	High uncertainty; not converged for some years in retro
3d	1978–1994, 2006–2018	Scottish Q1 (old) Scottish Q1 (new) Irish & Scottish Q4 (combined)	1985–2010 2011–2018 2011–2018	Fixed n=2	High uncertainty; not converged for some years in retro
4a	1978–1994 (1), 1995–2005 (4), 2006–2018 (1)	Scottish Q1 (old) Scottish Q1 (new) Scottish Q4 (old) Irish & Scottish Q4 (combined) Irish Q4 (truncated)	1985–2010 2011–2018 1996–2009 2011–2018 2003–2010	Fixed n=2	High uncertainty; not converged for some years in retro
4b*	1978–1994 (1), 1995–2005 (4), 2006–2018 (1)	Scottish Q1 (old) Scottish Q1 (new) Irish & Scottish Q4 (combined) Irish Q4 (truncated)	1985–2010 2011–2018 2011–2018 2003–2010	Fixed n=2	High uncertainty, but slightly less compared to other scenarios;
4c	1978–1994 (1), 1995–2005 (4), 2006–2018 (1)	Scottish Q1 (old) Scottish Q1 (new) Scottish Q4 (old) Irish & Scottish Q4 (combined)	1985–2010 2011–2018 1996–2009 2011–2018	Fixed n=2	High uncertainty; not converged for some years in retro

Scenario	Catch time-series	Survey	Survey time-series	Setting	Comment
4d	1978–1994 (1),	Scottish Q1 (old)	1985–2010	Fixed n=2	High uncertainty; not converged for some years in retro
	1995–2005 (4),	Scottish Q1 (new)	2011–2018		
	2006–2018 (1)	Irish & Scottish Q4 (combined)	2011–2018		

*** The output of these scenarios is shown in WD 5.4, Annex 1.**

In Scenarios 4a–d, different uncertainties (indicated as 1 or 4) were applied for different time periods.

Table 4.3. Whiting in Division 6.a. Output from the final assessment with the SPiCT model.

```
## Convergence: 0  MSG: relative convergence (4)
## Objective function at optimum: 56.7383462
## Euler time step (years):  1/16 or 0.0625
## Nobs C: 41,  Nobs I1: 26,  Nobs I2: 8,  Nobs I3: 8,  Nobs I4: 8
##
## Priors
##      logn  ~  dnorm[log(2), 0^2] (fixed)
##  logalpha  ~  dnorm[log(1), 2^2]
##   logbeta  ~  dnorm[log(1), 2^2]
##
## Model parameter estimates w 95% CI
##            estimate       cilow        ciupp    log.est
##  alpha1 5.182532e-01   0.2031801 1.321913e+00 -0.6572914
##  alpha2 2.771281e-01   0.0484502 1.585132e+00 -1.2832754
##  alpha3 4.602956e-01   0.1459654 1.451523e+00 -0.7758865
##  alpha4 1.439539e+00   0.7568139 2.738155e+00  0.3643232
##  beta   2.729842e-01   0.0733924 1.015369e+00 -1.2983414
##  r      1.878875e-01   0.0467444 7.552077e-01 -1.6719121
##  rc     1.878875e-01   0.0467444 7.552077e-01 -1.6719121
##  rold   1.878875e-01   0.0467444 7.552077e-01 -1.6719121
##  m      9.419787e+04  40.4417294 2.194080e+08 11.4531529
##  K      2.005411e+06 818.9114715 4.910997e+09 14.5113594
##  q1     5.159000e-04   0.0000002 1.557562e+00 -7.5695846
##  q2     1.455400e-03   0.0000005 4.466461e+00 -6.5324844
##  q3     1.086500e-03   0.0000004 3.337118e+00 -6.8248202
##  q4     7.189000e-04   0.0000002 2.194346e+00 -7.2377523
##  n      2.000000e+00   1.9996080 2.000392e+00  0.6931472
##  sdb    3.763513e-01   0.2729221 5.189772e-01 -0.9772322
##  sdf    3.357096e-01   0.1977264 5.699842e-01 -1.0915088
##  sdi1   1.950453e-01   0.0916112 4.152620e-01 -1.6345236
##  sdi2   1.042975e-01   0.0189128 5.751639e-01 -2.2605076
##  sdi3   1.732329e-01   0.0569552 5.268985e-01 -1.7531186
##  sdi4   5.417726e-01   0.3080859 9.527135e-01 -0.6129089
##  sdc    9.164340e-02   0.0359192 2.338169e-01 -2.3898502
##
## Deterministic reference points (Drp)
##           estimate       cilow        ciupp   log.est
##  Bmsyd 1.002705e+06 409.4557364 2.455498e+09 13.818212
##  Fmsyd 9.394370e-02   0.0233722 3.776038e-01 -2.365059
##  MSYd  9.419787e+04  40.4417294 2.194080e+08 11.453153
## Stochastic reference points (Srp)
##           estimate       cilow        ciupp   log.est rel.diff.Drp
##  Bmsys 5.865830e+05 255.0725976 1.348948e+09 13.282069   -0.7094006
##  Fmsys 5.861970e-02   0.0072558 4.735862e-01 -2.836685   -0.6025965
##  MSYs  1.968617e+04   2.3021332 1.683417e+08  9.887671   -3.7849783
##
## States w 95% CI (inp$msytype: s)
##                    estimate       cilow        ciupp    log.est
##  B_2018.81      4.873125e+05 157.3326474 1.509372e+09 13.0966610
##  F_2018.81      1.117300e-03   0.0000004 3.468024e+00 -6.7968214
##  B_2018.81/Bmsy 8.307649e-01   0.2351850 2.934584e+00 -0.1854084
##  F_2018.81/Fmsy 1.906050e-02   0.0000061 5.941482e+01 -3.9601363
##
## Predictions w 95% CI (inp$msytype: s)
##                  prediction       cilow        ciupp    log.est
##  B_2019.00      4.930150e+05 158.9845964 1.528851e+09 13.1082949
##  F_2019.00      1.116000e-03   0.0000004 3.479626e+00 -6.7980286
##  B_2019.00/Bmsy 8.404864e-01   0.2347211 3.009603e+00 -0.1737745
##  F_2019.00/Fmsy 1.903750e-02   0.0000061 5.961635e+01 -3.9613435
##  Catch_2019.00  5.683770e+02 247.4550810 1.305499e+03  6.3427849
##  E(B_inf)       0.000000e+00          NA           NA         NA
```

Table 4.4 Whiting in Division 6.a. Annual reported landings (international), estimated discards and catch (all in tonnes) as estimated by WKDEM. No revisions were made to pre-2003 data.

Year	Landings	Discards	Total catch
2003	1331.1	1697.4	3028.5
2004	798.5	2678.9	3477.4
2005	334.7	752.7	1087.3
2006	377.8	530.5	908.4
2007	480.7	314.0	794.7
2008	441.4	140.1	581.5
2009	480.2	411.6	891.8
2010	337.6	988.2	1325.9
2011	228.7	245.5	474.2
2012	304.3	711.1	1015.3
2013	215.7	1138.8	1354.5
2014	180.5	497.6	678.1
2015	223.2	835.6	1058.8
2016	226.5	851.1	1077.5
2017	177.9	1044.5	1222.4
2018	190.3	551.8	742.1

Table 4.5. Whiting in Division 6.a. Landings numbers-at-age (thousands). No revisions were made to pre-2003 data.

	Age						
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7+**
2003	98	652	1309	1481	414	93	2
2004	49	699	544	517	620	74	33
2005	26	273	460	145	107	49	5
2006	83	135	386	276	67	86	25
2007	193	190	294	361	152	31	53
2008	3	277	387	335	150	54	25
2009	108	255	258	417	107	49	14
2010	50	81	150	148	141	43	52
2011	0	256	144	94	27	26	8
2012	13	39	374	203	53	16	9
2013	4	41	76	269	74	19	6
2014	13	26	130	101	101	23	11
2015	7	74	56	157	71	73	30
2016	19	93	147	77	86	19	28
2017	17	37	167	69	52	39	10
2018	0	73	89	199	60	8	8

Table 4.6. Whiting in Division 6.a. Discards numbers-at-age (thousands). No revisions were made to pre-2003 data.

	Age						
Year	1	2	3	4	5	6	7+
2003	9448	2489	1775	375	25	7	1
2004	14941	5095	1011	660	125	4	2
2005	3246	2298	769	60	22	8	4
2006	4691	528	637	169	29	6	2
2007	1016	966	283	88	38	3	0
2008	630	144	114	31	37	4	0
2009	6880	114	66	44	15	4	0
2010	17678	1581	264	37	54	6	16
2011	2047	998	122	7	2	0	0
2012	7810	429	547	94	19	1	0
2013	16415	1578	172	255	8	2	2
2014	9831	51	55	27	30	8	3
2015	7930	909	287	112	18	17	0
2016	5506	1910	268	16	12	4	2
2017	7563	788	889	65	160	2	0
2018	2371	962	469	276	21	5	0

Table 4.7. Whiting in Division 6.a. Total catch numbers-at-age (thousands). No revisions were made to pre-2003 data.

	Age						
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7+**
2003	9546	3141	3083	1856	439	100	3
2004	14990	5794	1556	1176	745	78	35
2005	3272	2571	1229	205	129	57	10
2006	4773	663	1023	445	96	93	27
2007	1209	1156	578	449	190	33	53
2008	632	421	500	366	187	58	25
2009	6988	370	324	462	123	53	14
2010	17729	1662	414	185	196	49	68
2011	2048	1254	267	101	29	26	8
2012	7823	469	920	298	72	17	9
2013	16419	1619	247	523	82	21	7
2014	9844	77	185	127	130	31	14
2015	7937	983	343	269	90	90	30
2016	5525	2003	415	92	98	23	30
2017	7580	825	1056	134	212	41	10
2018	2371	1035	557	475	81	13	8

Table 4.8. Whiting in Division 6.a. Landings mean weights-at-age (kg). No revisions were made to pre-2003 data.

	Age						
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7+**
2003	0.236	0.272	0.301	0.373	0.349	0.409	0.659
2004	0.189	0.257	0.296	0.342	0.376	0.378	0.305
2005	0.215	0.253	0.297	0.366	0.426	0.455	0.383
2006	0.221	0.290	0.321	0.395	0.452	0.496	0.574
2007	0.215	0.289	0.356	0.416	0.497	0.598	0.667
2008	0.285	0.245	0.319	0.379	0.516	0.534	0.652
2009	0.288	0.317	0.406	0.446	0.439	0.444	0.603
2010	0.286	0.353	0.436	0.540	0.647	0.654	0.575
2011	0.201	0.356	0.396	0.502	0.571	0.578	0.370
2012	0.320	0.300	0.374	0.504	0.594	0.665	0.482
2013	0.225	0.325	0.355	0.441	0.546	0.597	0.770
2014	0.248	0.295	0.375	0.457	0.528	0.641	0.678
2015	0.261	0.347	0.447	0.468	0.508	0.596	0.600
2016	0.137	0.325	0.483	0.509	0.606	0.676	0.664
2017	0.340	0.352	0.413	0.546	0.497	0.510	0.684
2018	0.173	0.407	0.396	0.435	0.520	0.472	0.564

Table 4.9. Whiting in Division 6.a. Discards mean weights-at-age (kg). No revisions were made to pre-2003 data.

	Age						
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7+**
2003	0.091	0.161	0.193	0.243	0.209	0.291	0.278
2004	0.091	0.178	0.223	0.233	0.302	0.343	0.282
2005	0.074	0.145	0.207	0.188	0.302	0.289	0.368
2006	0.047	0.195	0.233	0.285	0.311	0.494	0.361
2007	0.064	0.157	0.232	0.223	0.231	0.787	0.266
2008	0.076	0.211	0.305	0.350	0.423	0.233	0.289
2009	0.051	0.283	0.227	0.262	0.250	0.248	NA
2010	0.040	0.119	0.239	0.360	0.360	0.382	0.224
2011	0.034	0.136	0.307	0.256	0.228	NA	NA
2012	0.057	0.152	0.292	0.362	0.356	0.386	NA
2013	0.041	0.209	0.229	0.358	0.385	0.299	0.371
2014	0.045	0.182	0.289	0.362	0.427	0.422	0.757
2015	0.072	0.171	0.212	0.336	0.316	0.427	NA
2016	0.068	0.206	0.276	0.292	0.304	0.261	0.367
2017	0.066	0.197	0.351	0.409	0.331	0.881	NA
2018	0.067	0.184	0.250	0.307	0.414	1.107	NA

Table 4.10. Whiting in Division 6.a. Total catch mean weights-at-age (kg). No revisions were made to pre-2003 data.

	Age						
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7+**
2003	0.092	0.184	0.239	0.347	0.341	0.401	0.516
2004	0.091	0.188	0.249	0.281	0.364	0.377	0.304
2005	0.075	0.156	0.241	0.313	0.405	0.432	0.376
2006	0.050	0.214	0.266	0.353	0.410	0.495	0.557
2007	0.088	0.179	0.295	0.378	0.444	0.613	0.666
2008	0.077	0.233	0.316	0.376	0.498	0.514	0.648
2009	0.054	0.307	0.369	0.429	0.415	0.430	0.603
2010	0.040	0.130	0.311	0.504	0.567	0.622	0.492
2011	0.034	0.181	0.355	0.485	0.546	0.578	0.370
2012	0.057	0.164	0.325	0.459	0.531	0.643	0.482
2013	0.041	0.212	0.268	0.401	0.530	0.571	0.679
2014	0.045	0.220	0.349	0.437	0.505	0.581	0.694
2015	0.072	0.185	0.250	0.413	0.469	0.565	0.600
2016	0.068	0.211	0.349	0.472	0.568	0.601	0.649
2017	0.066	0.204	0.361	0.480	0.372	0.524	0.684
2018	0.067	0.199	0.273	0.361	0.492	0.731	0.564

Table 4.11. Whiting in Division 6.a. Natural mortality-at-age.

	Age						
Year	1	2	3	4	5	6	7+
1965	0.700	0.634	0.572	0.525	0.468	0.457	0.436
1966	0.703	0.634	0.572	0.526	0.471	0.459	0.438
1967	0.705	0.634	0.573	0.527	0.474	0.461	0.441
1968	0.706	0.635	0.574	0.527	0.478	0.463	0.443
1969	0.707	0.635	0.574	0.528	0.481	0.465	0.445
1970	0.708	0.635	0.575	0.529	0.484	0.467	0.447
1971	0.707	0.636	0.576	0.530	0.487	0.469	0.449
1972	0.705	0.636	0.576	0.531	0.490	0.471	0.451
1973	0.703	0.637	0.577	0.531	0.492	0.473	0.453
1974	0.701	0.637	0.578	0.532	0.495	0.475	0.455
1975	0.699	0.637	0.579	0.533	0.497	0.477	0.457
1976	0.698	0.638	0.579	0.534	0.499	0.478	0.459
1977	0.699	0.638	0.580	0.535	0.501	0.480	0.461
1978	0.701	0.638	0.581	0.536	0.503	0.482	0.463
1979	0.705	0.639	0.582	0.538	0.505	0.484	0.465
1980	0.711	0.639	0.583	0.539	0.506	0.486	0.467
1981	0.718	0.640	0.584	0.540	0.508	0.489	0.468
1982	0.727	0.640	0.585	0.541	0.510	0.491	0.470
1983	0.736	0.640	0.586	0.543	0.511	0.492	0.472
1984	0.746	0.641	0.587	0.544	0.513	0.494	0.474
1985	0.755	0.641	0.587	0.545	0.514	0.496	0.475
1986	0.763	0.642	0.588	0.546	0.515	0.497	0.477
1987	0.770	0.642	0.589	0.547	0.516	0.499	0.478
1988	0.775	0.642	0.590	0.548	0.517	0.500	0.479
1989	0.779	0.643	0.591	0.549	0.518	0.501	0.480
1990	0.782	0.643	0.591	0.550	0.519	0.502	0.481
1991	0.784	0.643	0.592	0.551	0.519	0.503	0.481
1992	0.785	0.644	0.592	0.551	0.520	0.503	0.482

	Age						
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7+**
1993	0.787	0.644	0.592	0.552	0.520	0.504	0.483
1994	0.788	0.645	0.593	0.552	0.520	0.504	0.483
1995	0.791	0.645	0.592	0.553	0.520	0.504	0.483
1996	0.794	0.645	0.592	0.553	0.520	0.504	0.484
1997	0.799	0.646	0.592	0.553	0.520	0.504	0.484
1998	0.804	0.646	0.591	0.552	0.520	0.504	0.484
1999	0.811	0.646	0.590	0.551	0.519	0.503	0.484
2000	0.818	0.647	0.589	0.550	0.519	0.503	0.484
2001	0.827	0.647	0.588	0.549	0.518	0.502	0.484
2002	0.838	0.648	0.586	0.547	0.517	0.501	0.484
2003	0.850	0.648	0.584	0.545	0.516	0.500	0.483
2004	0.864	0.648	0.583	0.543	0.515	0.498	0.483
2005	0.880	0.649	0.581	0.540	0.513	0.497	0.482
2006	0.896	0.649	0.579	0.538	0.512	0.495	0.482
2007	0.913	0.649	0.577	0.535	0.510	0.493	0.481
2008	0.929	0.650	0.574	0.531	0.509	0.491	0.480
2009	0.943	0.650	0.572	0.528	0.507	0.488	0.479
2010	0.954	0.651	0.570	0.525	0.505	0.486	0.478
2011	0.961	0.651	0.568	0.522	0.504	0.483	0.476
2012	0.963	0.651	0.566	0.519	0.502	0.480	0.475
2013	0.959	0.652	0.565	0.517	0.501	0.478	0.473
2014	0.950	0.652	0.563	0.514	0.499	0.475	0.472
2015	0.937	0.653	0.561	0.512	0.498	0.472	0.471
2016	0.920	0.653	0.560	0.510	0.497	0.469	0.469
2017	0.901	0.653	0.558	0.508	0.496	0.467	0.468
2018	0.881	0.654	0.556	0.506	0.495	0.464	0.466

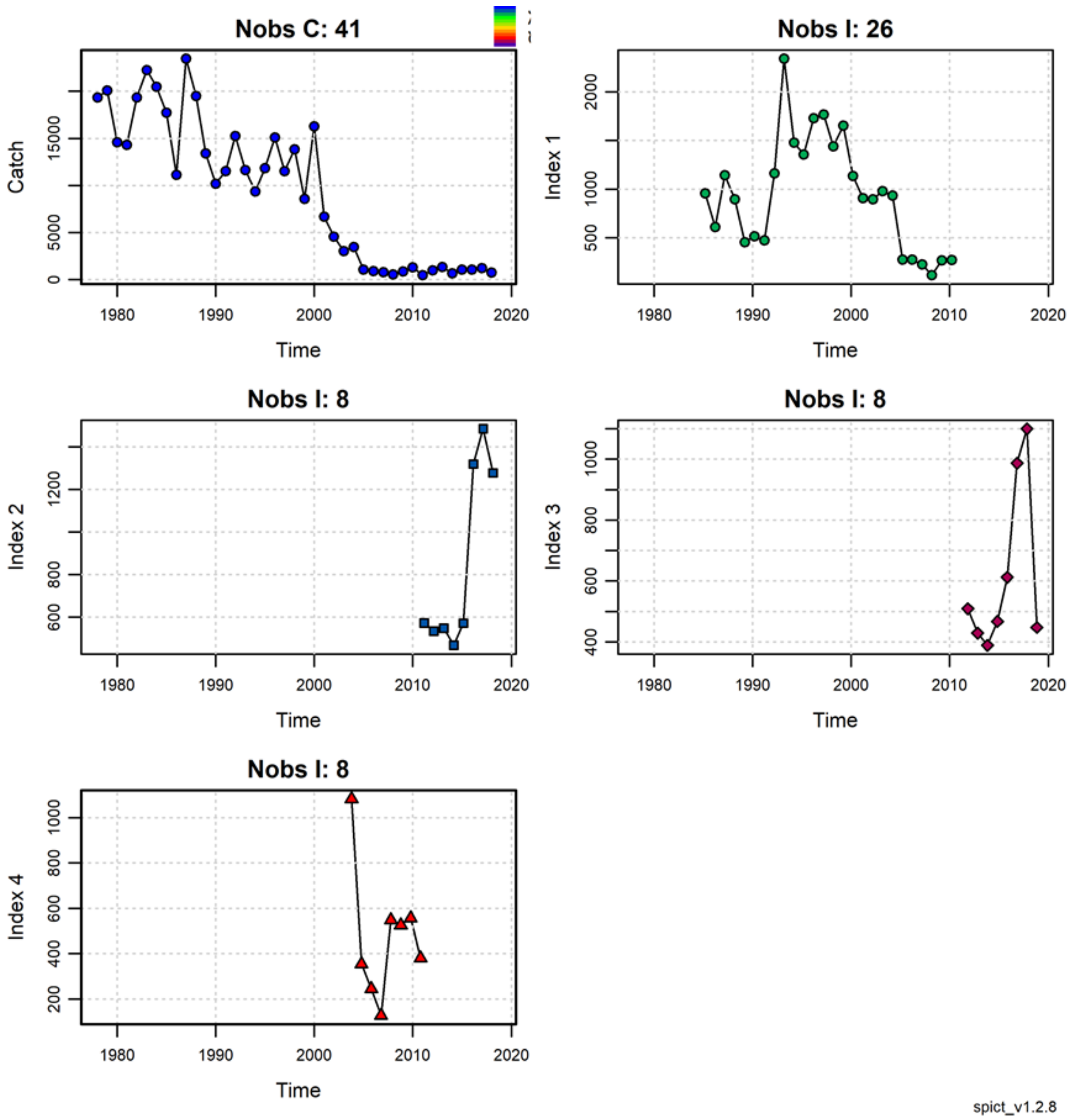


Figure 4.1. Whiting in Division 6.a. The catch and index observations in the final assessment with the SPiCT model.

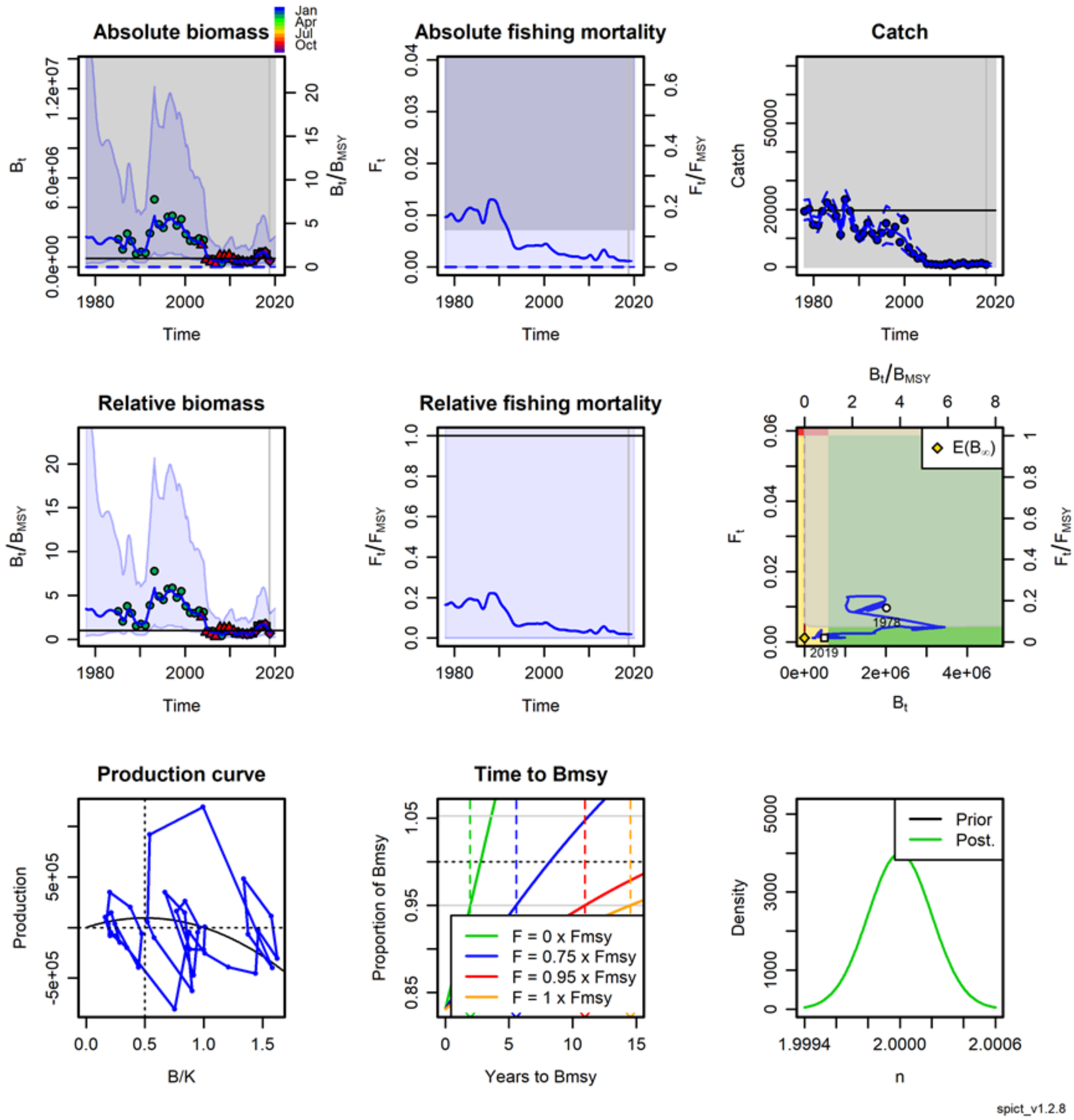


Figure 4.2. Whiting in Division 6.a. The multipanel plot with the most important outputs from the final assessment with the SPiCT model.

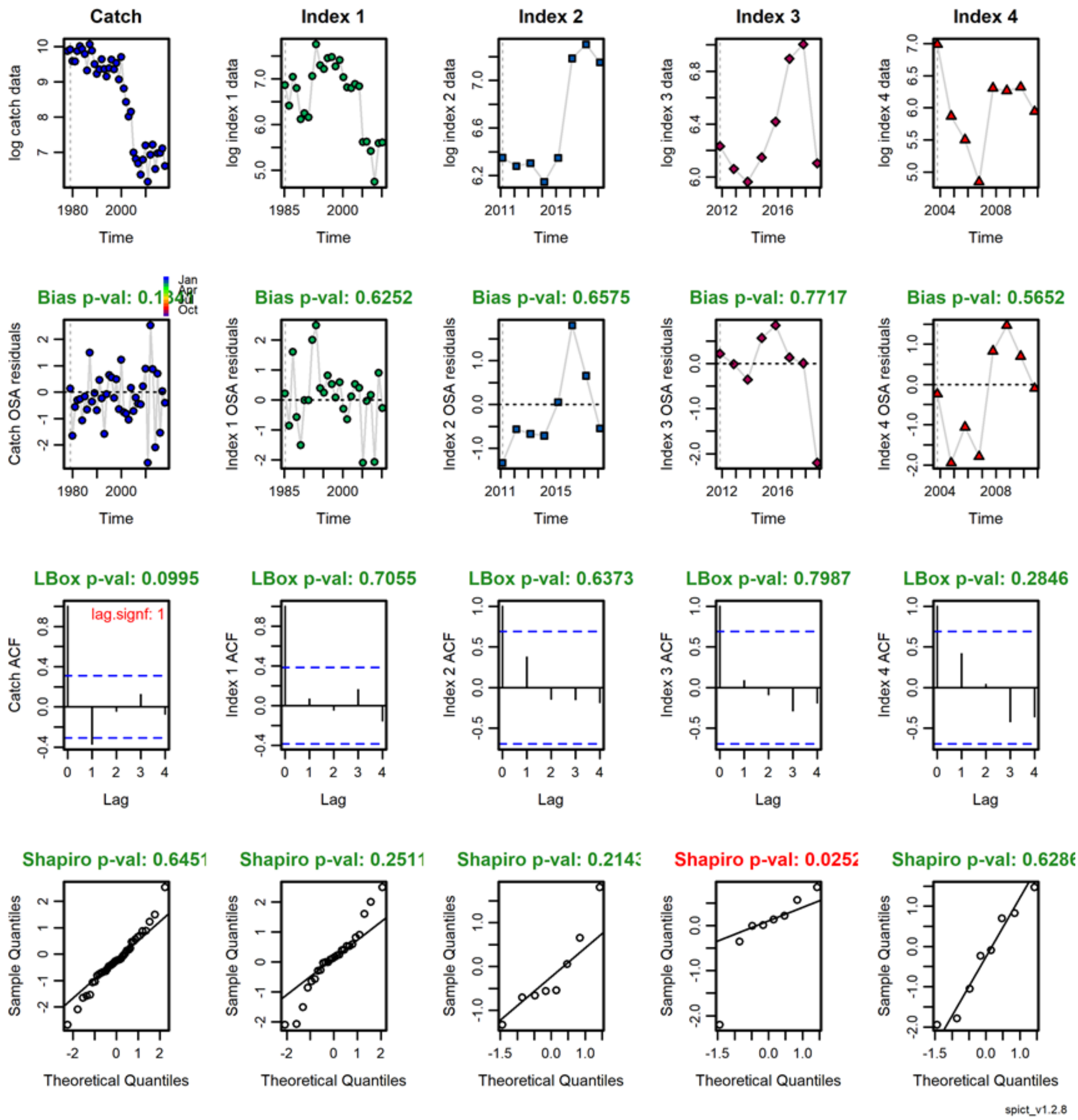
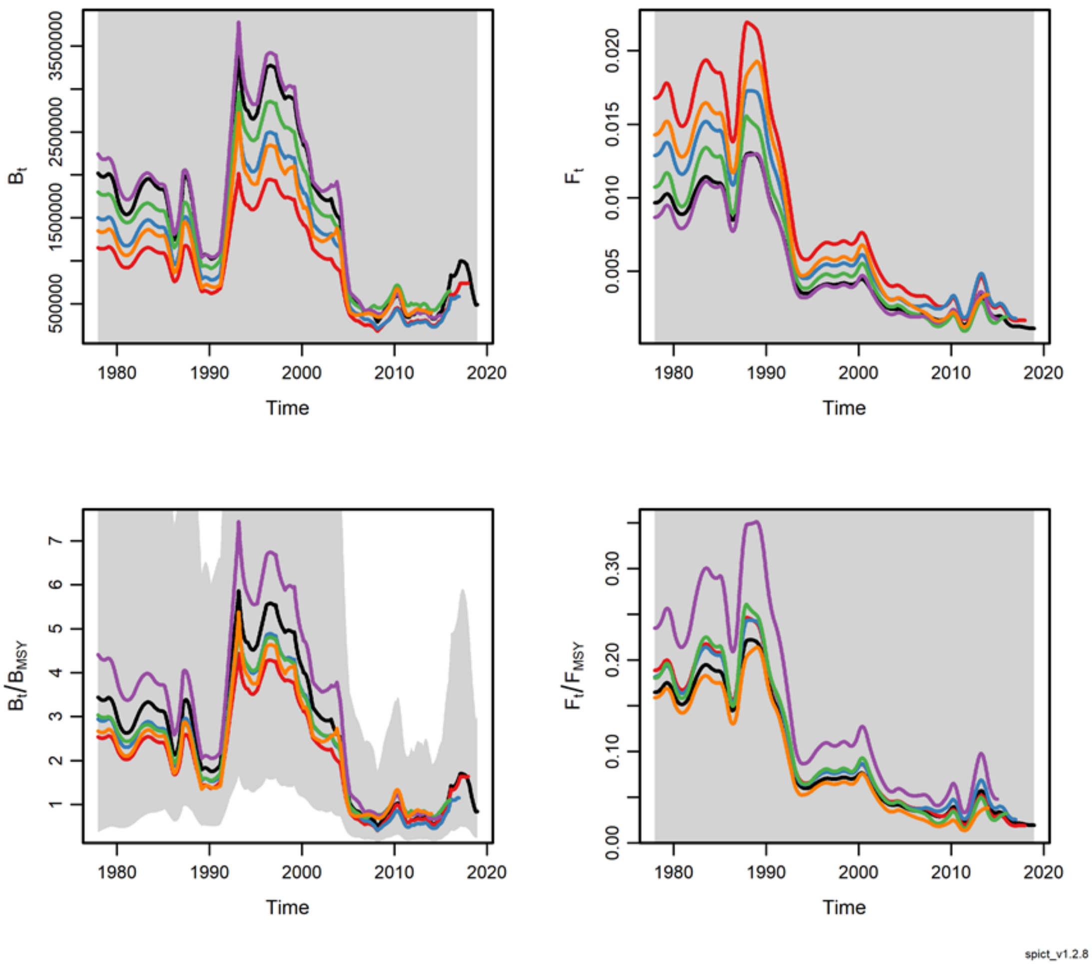


Figure 4.3. Whiting in Division 6.a. Plot of residuals and diagnostics from the final assessment with the SPiCT model.



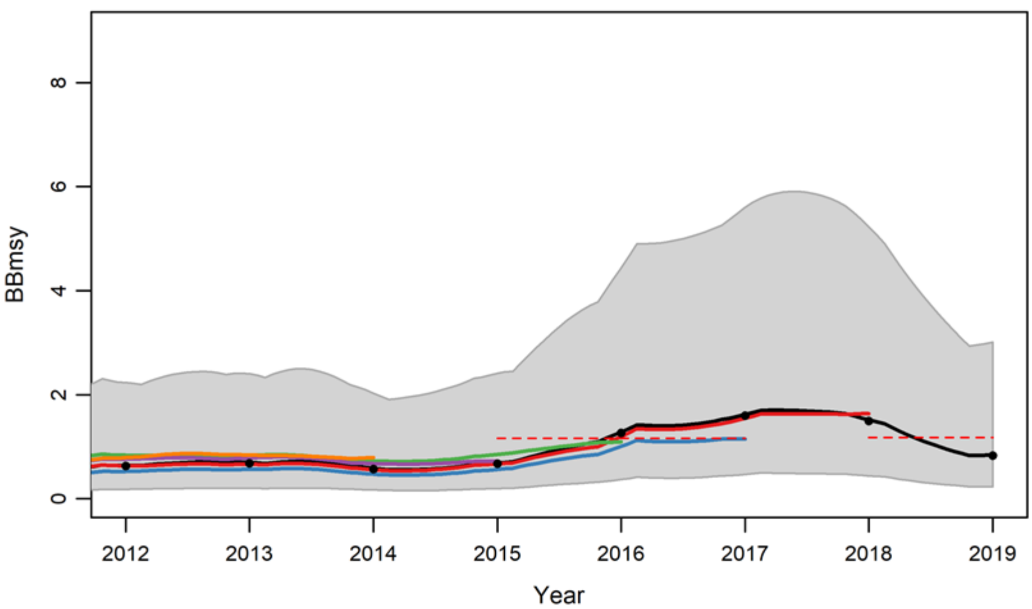


Figure 4.4. Whiting in Division 6.a. Retrospective analysis from the final assessment with the SPiCT model. The bottom plot shows the relative biomass for the most recent years.

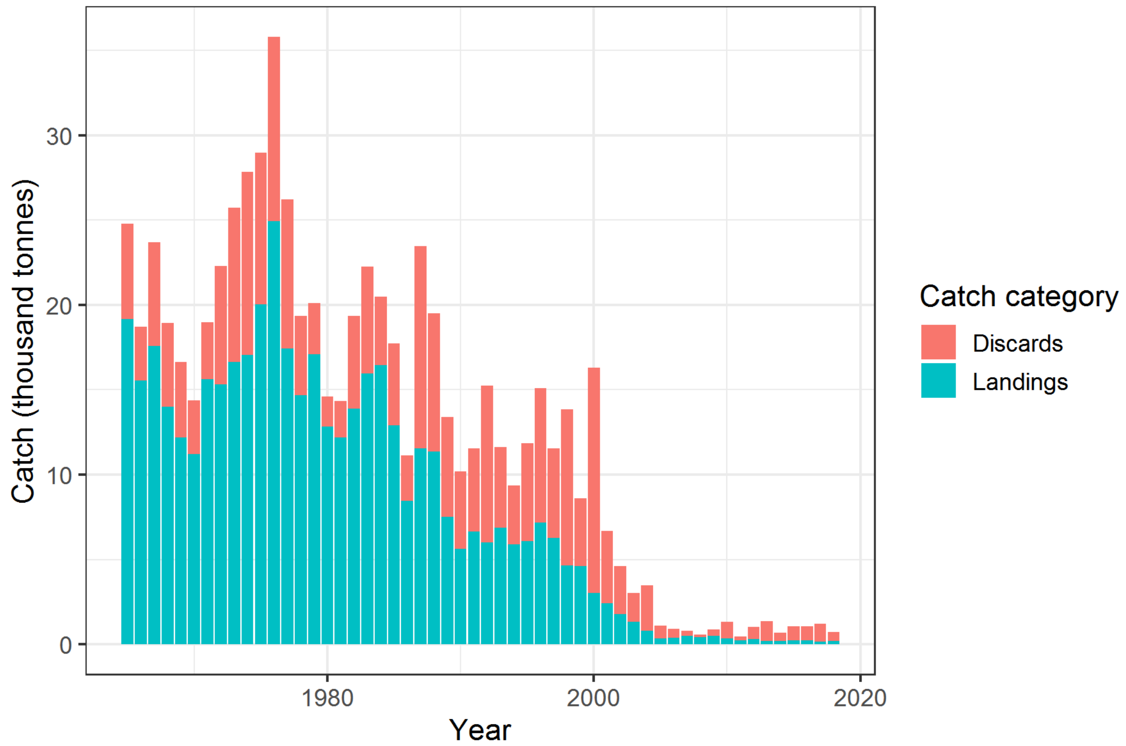


Figure 4.5. Whiting in Division 6.a. Landings, discards and catch (in tonnes, whiting at-age 1 and older) as officially reported to ICES.

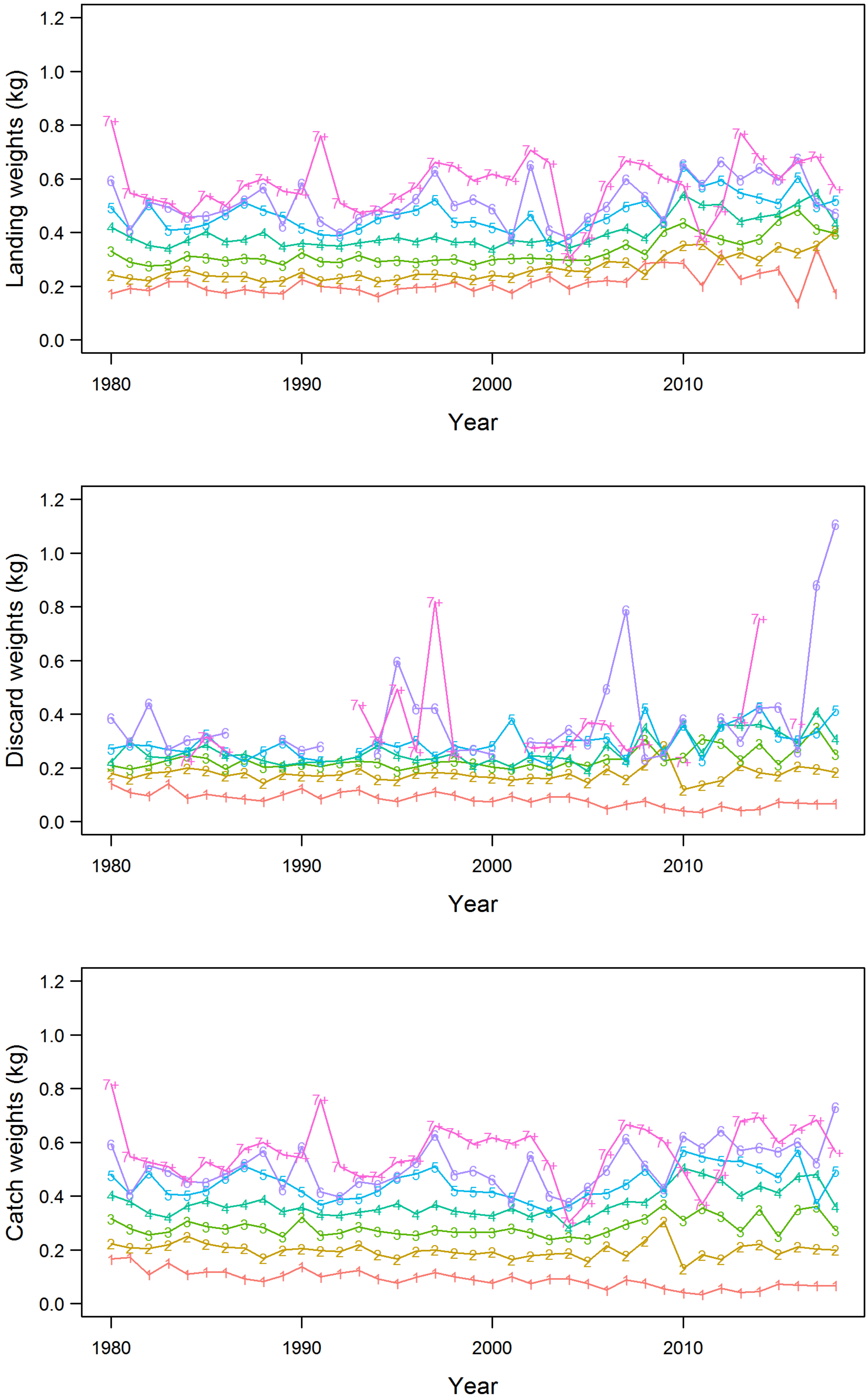


Figure 4.6. Whiting in Division 6.a. Mean weight-at-age in the landings (upper panel), discards (middle panel) and catch (lower panel).

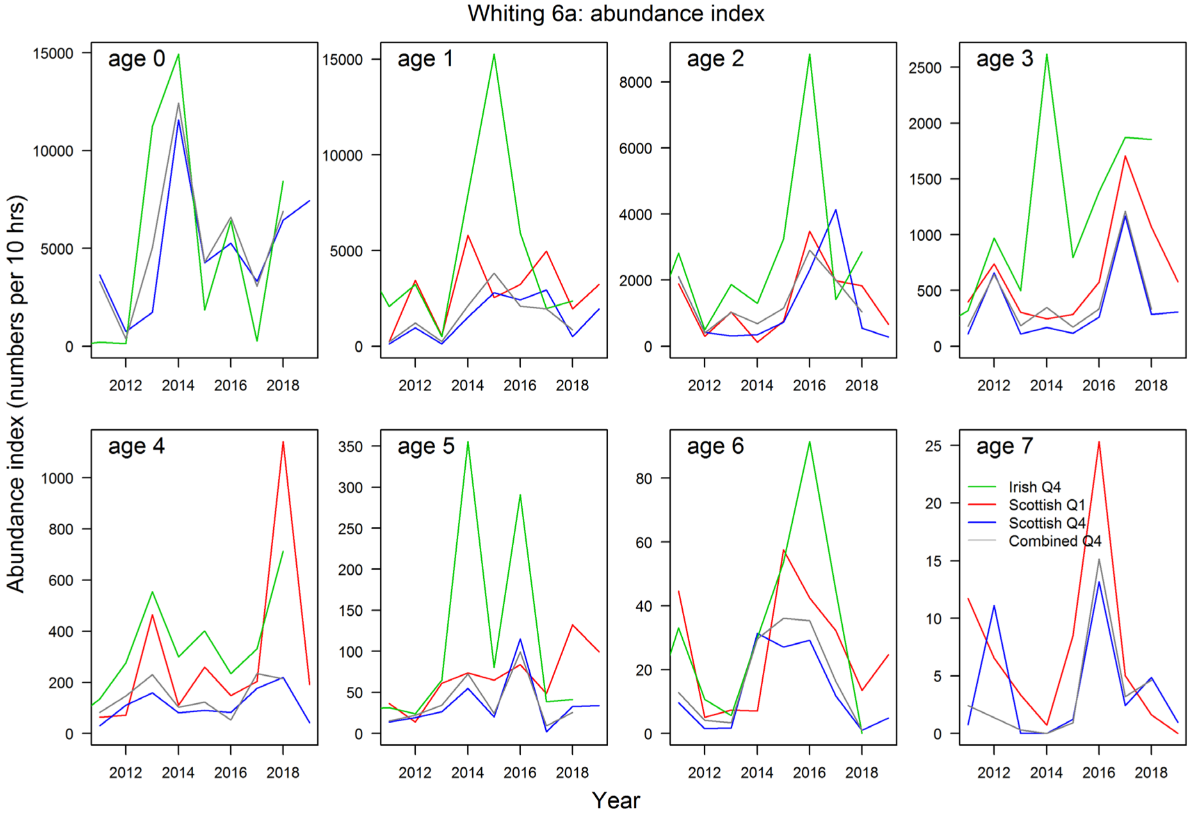


Figure 4.7. Whiting in Division 6.a. Abundance index in the four tuning series: Irish Q4 survey, new Scottish surveys for Q1 and Q4 and combined Irish and Scottish Q4 survey, in 2011–2019.

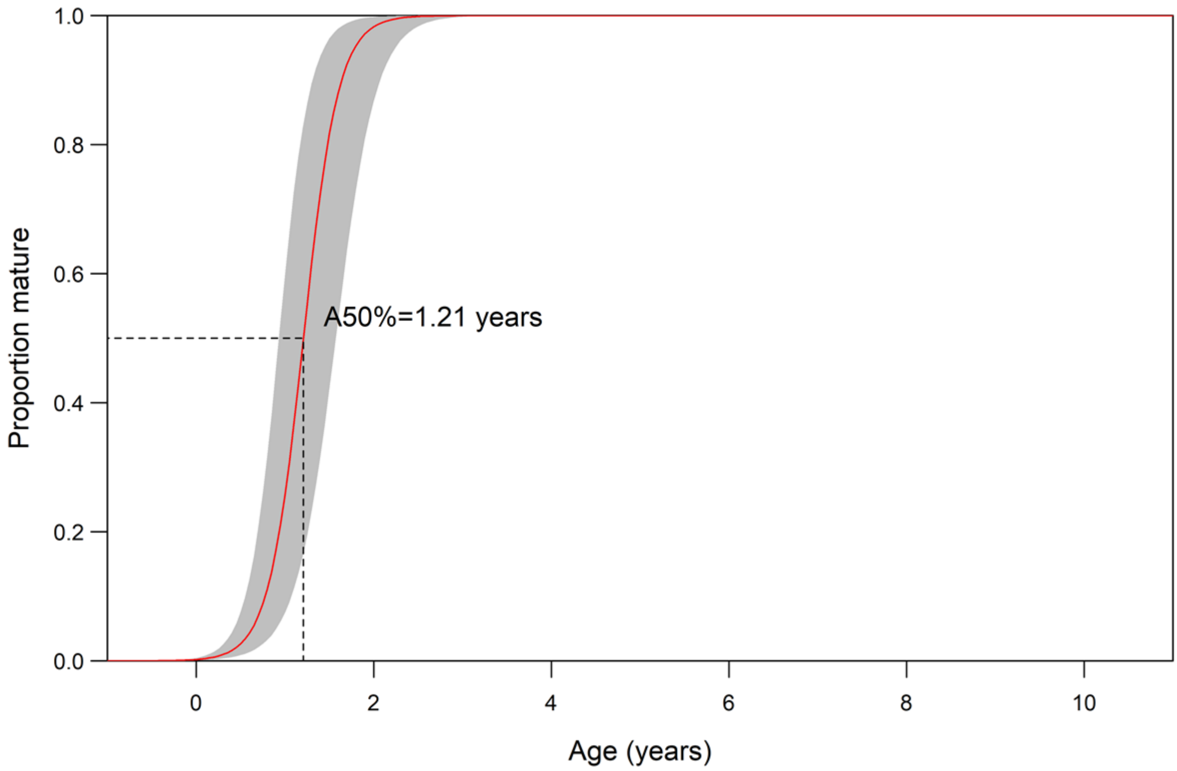


Figure 4.8. Whiting in Division 6.a. The estimated maturity ogive.

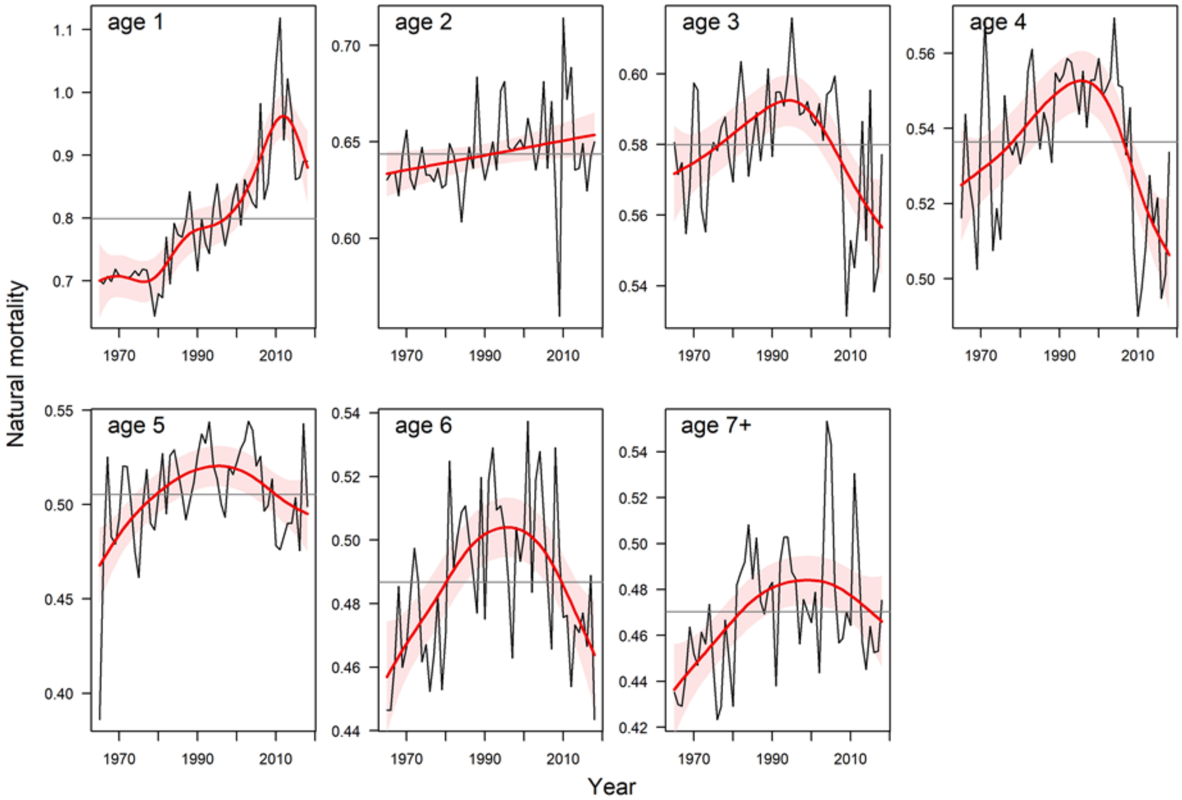


Figure 4.9. Whiting in Division 6.a. The estimated natural mortality.

5 Whiting (*Merlangius merlangus*) in Division 3.a (Skagerrak and Kattegat) (Whg.27.3a)

5.1 Why a benchmark?

Whiting in Division 3.a is a category 5 stock (ICES, 2019), i.e. a stock with only a short series of catches. There is currently no assessment for the stock and the status of the stock is unknown. ICES provides biennial advice for the stock based on the precautionary approach and information on previous catches. In 2013, the advice was based on the average catch in the years 2010—2012, i.e. 500 tonnes; since then the advice has remained the same with a reduction of 20% (precautionary buffer) every four years. In 2019, ICES advised that when the precautionary approach is applied the catches for each of the years 2020 and 2021 should be no more than 400 tonnes.

Nevertheless, more information is available for the stock that has not been used until now, but could be utilised to infer stock status and provide better catch advice. The aim of this benchmark is to present available time-series of catches (landings and discards), a new fisheries-independent biomass index that combines all relevant scientific surveys in the area, and provide an improved method for giving advice for whiting in Division 3.a.

5.1.1 Presentations and working documents

Three working documents were presented during the data evaluation meeting (9—13 December 2019, ICES HQ, Copenhagen, Denmark) and the benchmark workshop (10—14 February 2020, ICES HQ, Copenhagen, Denmark) are briefly summarised in Annex 2. Full copies of the working documents can be obtained from the authors, copies are held on the WKDEM SharePoint.

5.2 Summary of decision

The decision of the group was to raise the stock to category 3 and provide advice using the "2-over-3" trend-based approach with the new biomass index that was presented in this benchmark meeting.

There was an attempt to provide an analytical assessment for the stock using a surplus production model (SPiCT) without satisfactory results, mostly to the very high estimated uncertainty of the F/F_{MSY} time-series and the MSY reference point.

5.3 Investigations undertaken (summary)

5.3.1 Compilation of available data

5.3.1.1 Commercial catch

In connection with the WKDEM 2020 benchmark, there was an attempt to reconstruct the commercial catch time-series. New data became available from national data submitters of countries that are responsible for most of the catches of whiting in the area. A summary of the newly available information is given below; a complete presentation of the commercial catches is given in the corresponding working document (WD 6.2).

InterCatch 2002—2018: Landings and discards are now available for Denmark, Sweden, the Netherlands by subdivision, quarter of the year, and fleet. Most of the catches comes from Landings for some years are available for Norway. The coverage of discards is high for the stock (45—95%, Figure 5.1). Raising of discards for the fleets that lack discard information was done according to the following scheme:

- Industrial bycatch: no discards;
- Norwegian fleets: no discards;
- All other fleets: discard rate of non-sampled métiers was assumed to be equal to the weighted mean of all available discard rates per subdivision (Figure 5.2). The weights were equal to landings in kg.

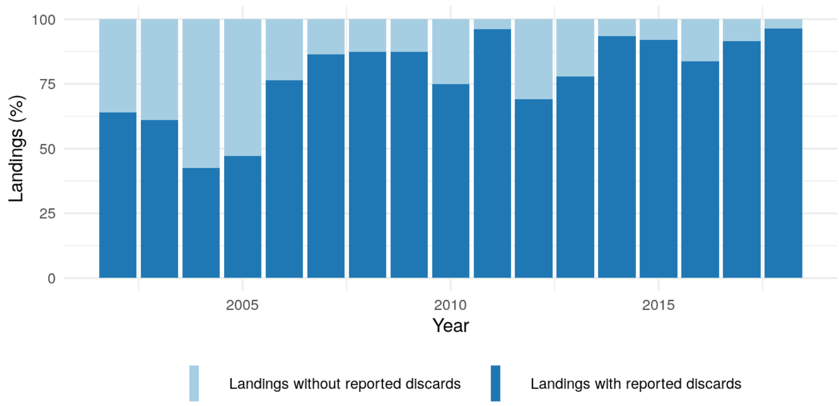


Figure 5.1. Percent of landings that have discard information (dark blue) and lack discard information (light blue).

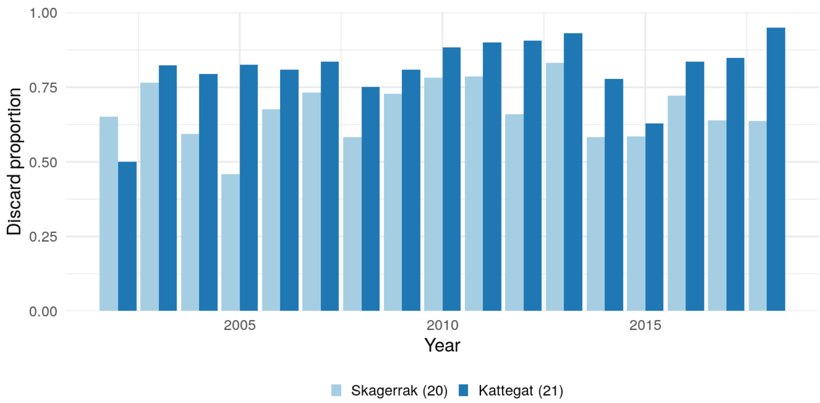


Figure 5.2. Proportion of discards in Skagerrak (light blue) and Kattegat (dark blue).

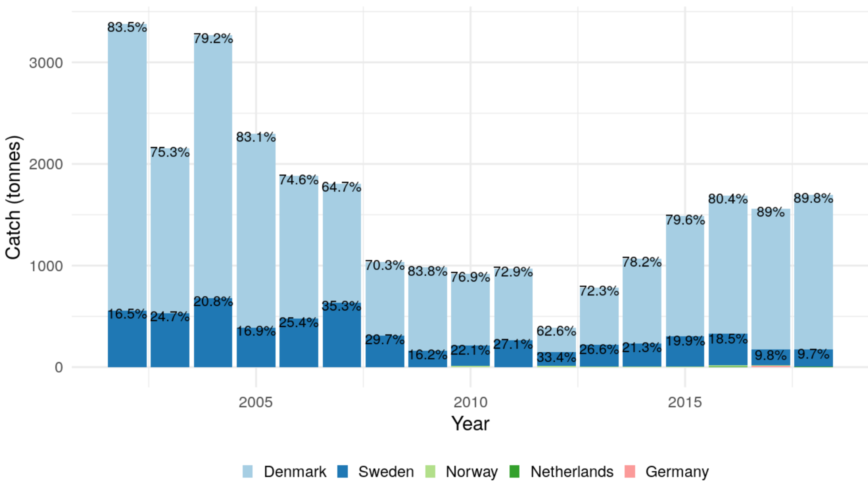


Figure 5.3. Catches (in tonnes) by country from 2002—2018. These correspond to imported landings and imported and raised discards.

The time-series of catch with the new discard raising is shown in Figure 5.3.

Length distributions: Length distributions from samples commercial catch are available, mostly from sampling of the Danish fleets. The industrial fleet that is responsible for a considerable part of the commercial catch is not being sampled adequately; therefore, length distributions are not representative of the catch and cannot be used for length-based analyses. Nevertheless, length distributions based on sampling of few hundred individuals per year were used to make decisions on the calculation of the new exploitable biomass index (Figure 5.4). It is apparent that individuals caught in the industrial fleet are at least as small as the smallest caught in the survey (Figure 5.5).

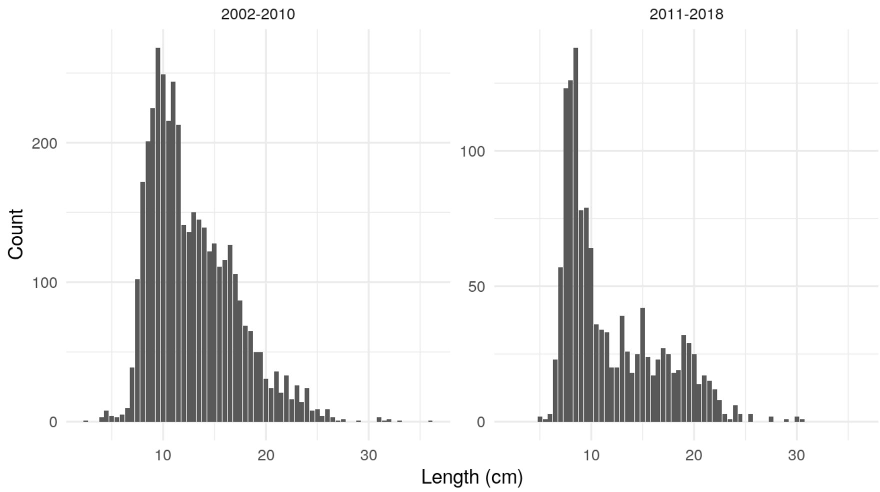


Figure 5.4. Length distribution from the Danish industrial fleet for the years 2002—2010 (left) and 2011—2018 (right).

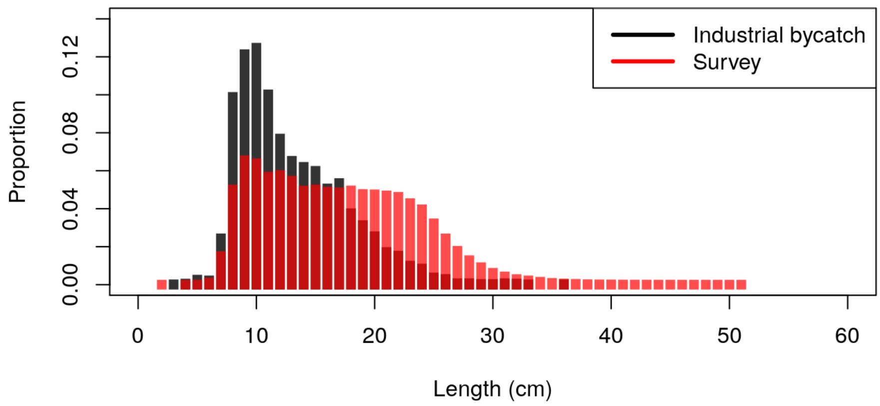


Figure 5.5. Industrial fleet length distribution (black) compared to survey length distribution (red) from 2002—2018.

5.3.1.2 Survey Data – fishery-independent biomass index

Several surveys operate annually or biannually in the area. The DATRAS database (https://datras.ices.dk) contains data from the two international bottom trawl surveys, the North Sea international bottom trawl survey (NS-IBTS), the Baltic bottom trawl survey (BITS). The Danish National Institute of Aquatic Resources (DTU Aqua) is conducting two national surveys in the area targeting cod and sole. The extent of the four surveys with some of subset of the stations is shown in Figure 5.6. The haul level information of these four surveys were combined in one generalised additive model (GAM) to produce standardised biomass indices. A brief summary of the new biomass index is given below. A complete description of the input data, modelling approach and results are in the survey index working document (WD 6.1).

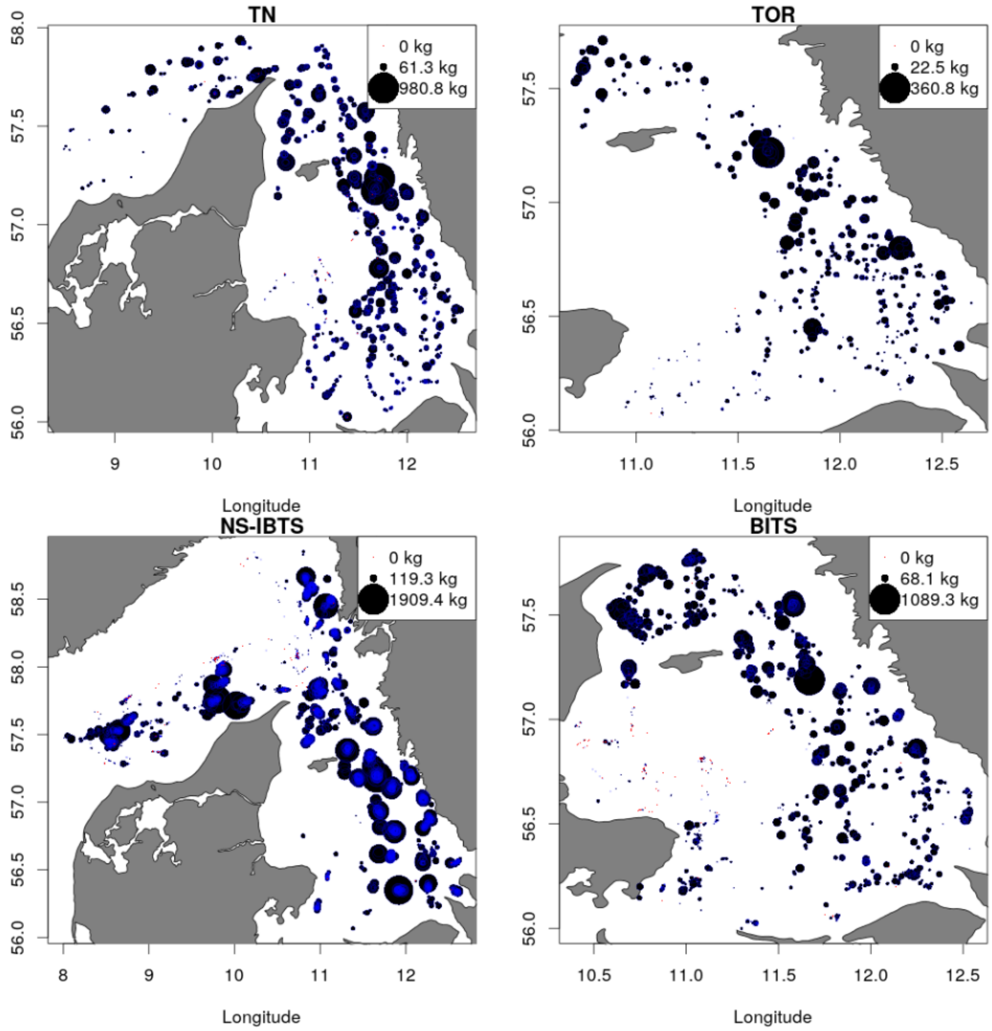


Figure 5.6. Catch per haul (kg) of whiting in Division 3a in the four surveys in the area: the two Danish national surveys targeting sole (TN) and cod (TOR), and the two international bottom trawl surveys in the North Sea (NS-IBTS) and Baltic Sea (BITS). Red dots show hauls that did not catch whiting.

The standardisation of the biomass index follows the method described in Berg *et al*. (2014) but models the catch of whiting in weight (in kg) in each haul as response variable, instead of numbers-at-age. The modelling is done using the R package *'surveyIndex'* (Berg, 2014). The space and time smoothers are decomposed into time-invariant spatial effect, a seasonal repeating pattern, and a space—time interaction effect that can capture smooth changes in the spatial distribution over longer time-scales. The Tweedie distribution (compound Poisson-Gamma) is used, because it is simpler and easier to work with and has a more consistent interpretation when sampling effort is not constant. Further, the model includes a smooth function for depth, fixed effects for gear, random effects for the interaction between ship and gear and an offset of natural logarithm of haul duration.

The fitted model from above allows prediction of the CPUE in any position in space and time when all nuisance parameters (gear, ship, quarter, haul duration) are fixed at constant values. The Q1 biomass index with corresponding uncertainty results from taking the sum of all predictions over a fine grid for each given year (Figure 5.7).

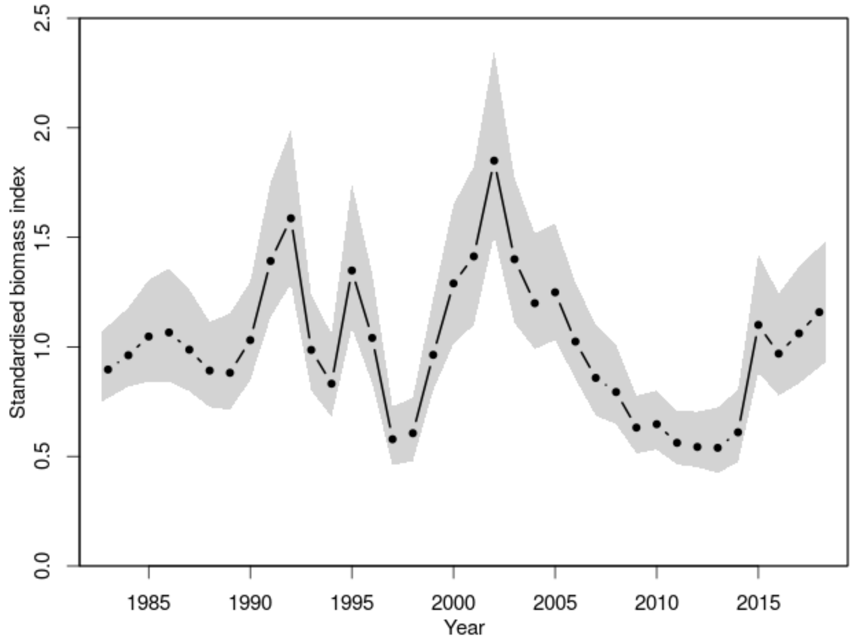


Figure 5.7. Standardised biomass index for whiting in Division 3a in quarter 1 for the years 1983—2018 with 95% confidence intervals (shaded area).

The goodness of fit was judged by visually inspecting residuals over space, time, and the combination of ship-gear-quarter. The robustness of the index is tested by retrospective analysis, where the time-series is shortened by excluding the last one, two and three years of available data (Figure 5.8). The effect of leaving each of the surveys out of the index calculation is shown in Figure 5.9. Leaving out NS-IBTS leads to a considerably different signal, indicating that the population in Skagerrak is different than in Kattegat, because the latter is mostly covered by the other three surveys.

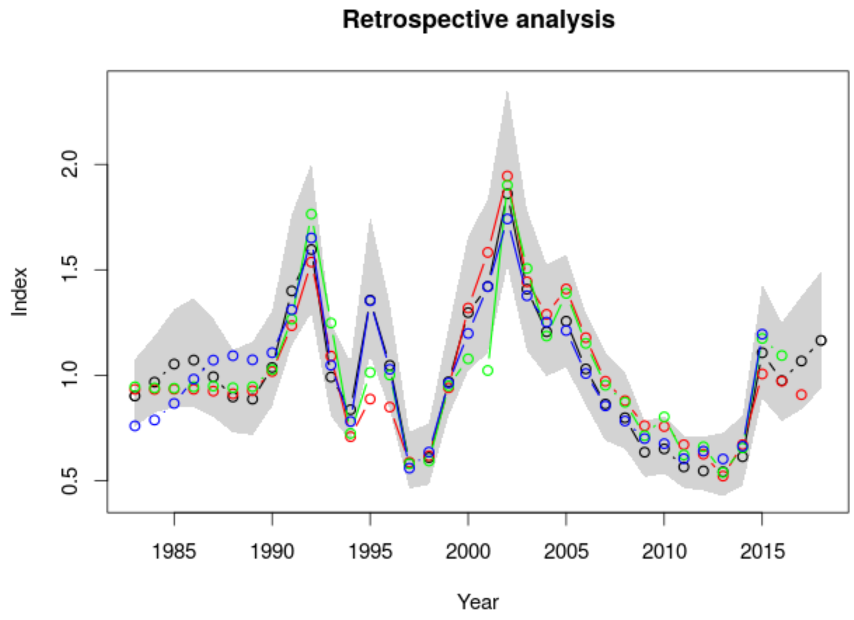


Figure 5.8. Retrospective analysis for the survey index. The 95% confidence bounds for the base run (all years included) is shown as shaded area.

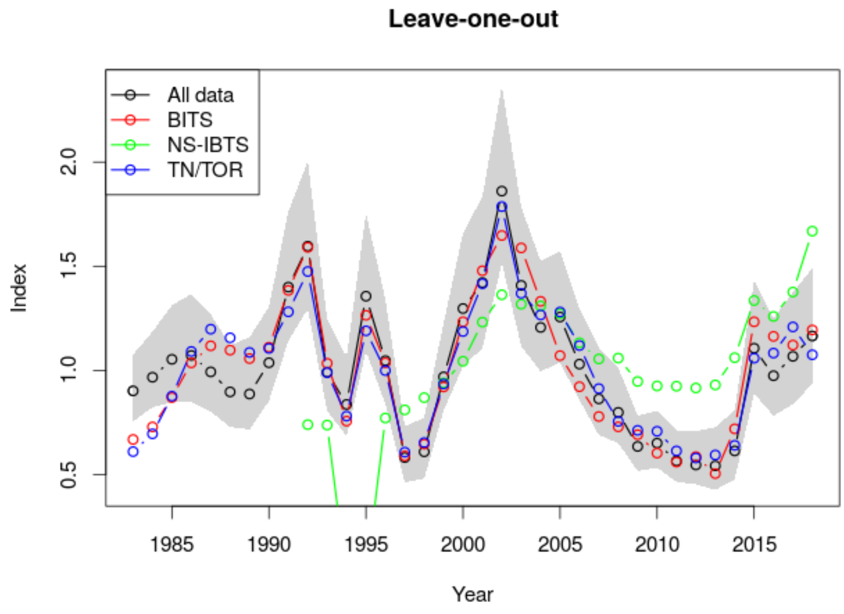


Figure 5.9. Survey index calculated by leaving out one of the two international bottom trawl surveys (green and red) or both Danish surveys (blue). The base run (black) includes all four surveys. The 95% confidence bounds for the base run is shown as shaded area.

5.3.1.3 Stock assessment

The Surplus Production model in Continuous Time (SPiCT, Pedersen and Berg, 2017) was considered for the assessment of the stock, as it is able to combine the catch and exploitable biomass index to provide estimates of important management quantities, i.e. B/B_{MSY} and F/F_{MSY}. The scenarios tested, with corresponding results, diagnostics are shown in the assessment working document (WD 6.3).

5.3.1.4 TAC advice

The advice for whiting in Division 3.a is still based on the ICES category "2-over-3" rule, i.e. the new advice will be equal to the last advice multiplied by the ratio of the mean of the biomass index of the last two years to the mean of the three years before that. The advice will be subject to an uncertainty cap, meaning that the new advice cannot be higher or lower than 20% compared to the previous advice. Furthermore, a precautionary buffer will be applied every four years, reducing the advice by 20%. The biomass index that is going to be used from now on will be the new standardised biomass index presented here that combines all four surveys in the area.

5.4 Future considerations

Whiting in Division 3.a lacks still after this benchmark an acceptable assessment that would be able to form the basis of providing advice for the stock. Few open research questions and topics for further investigations are given here.

SPiCT is not able at the moment to optimally deal with standardised survey indices that are smoothing observations over space and time. The main problem is that SPiCT assumes that the survey index observations are not correlated. Additional work is needed to implement correlation structures for the biomass index. That could allow including seasonal estimates of the biomass index in combination with seasonal catch data that could potentially improve the assessment of the stock.

More biological sampling of the stock in the area would substantially improve the advice of the stock, as it would allow for estimation of the stock status. Most important data on representative length or age distributions are lacking. At present there is very limited sampling done in the industrial fleet and no aging is done in the area.

A way to circumvent the lack of age—length keys from Division 3.a could be to borrow them from neighbouring areas, i.e. North Sea stock. Further investigation and thoughtful sensitivity analysis are necessary to ensure the robustness and adequacy of such an attempt.

5.5 Reviewers' comments

For this benchmark, historical catch was reconstructed based on newly available landing information: InterCatch (2002), official landings (1950—2018), and WGNSSK (1975—2018). From 2005, the Danish industrial fleet is included in the landings, and overall, there is good agreement between data streams. Whiting 27.3.a was previously a category 5.b stock. For this benchmark, the assessment was compiled using a SPiCT assessment model. During this benchmark, the panel was considering moving the stock to a Category 3 stock assessment in ICES, which includes information on catch and biomass indices.

This assessment exhibited a large degree of uncertainty and was not acceptable for the use of reference point calculation. Key parameters such as carrying capacity 95% credible intervals varied many orders of magnitude. Ultimately, we elected to use the biomass trends from the SPiCT model and upgrade this stock to Category 3. We commend the work put into reconstructing catches for this stock but highlight remaining uncertainties.

The majority of this fishery is bycatch from the Danish industrial fleet (~75%), and even given the data on hand, the stationarity of catch reporting bias through time is unlikely. Further, survey length frequencies did not seem to be totally representative of the Danish industrial fleet catch. Limited Danish industrial fleet data were acquired, but sample size is meagre, and unfortunately, lower in recent years. We strongly urge future assessors to request more length—frequency data from the Danish industrial fleet be gathered. With this information, the survey and fishery data may help reduce model uncertainty.

While catch data were available and compiled from 1950 to the present, the early years of catch data were considered very uncertain due to reporting issues. Thus, a SPiCT model starting in 1950 was not run; in addition, prior to the start of the survey time-series there was no effort data, which led to the concern that large catches could come from either large or small population sizes.

Some additional runs were requested by the review panel, but the main focus of the remaining discussion was compiling the most representative index for inclusion into the model. The proposed base run index used only the larger lengths for index creation, which missed those fish harvested by the industrial bycatch. Thus, the analyst modified the length information to keep smaller sizes to see if a signal is being masked by excluding data.

The proposed base run start year was 1983 in order to include high catches, the index time-series, and to address the most recent research regarding when a regime shift has occurred. The final proposed base run included the work to correct the length information to account for the industrial bycatch information. The index was meant to represent the catches or exploitable biomass of the population. However, in the end the surplus production model wasn't able to provide population estimates with confidence intervals that informed the stock status. Thus, the surplus production model could not be used to provide reference point advice. The reviewers and panel recommended using the survey-based method of Category 3, but not including length analyses due to missing information from the industrial fishery and spotty length information for the other fleet.

We applaud the work put into bringing this stock up from a category 5 and into category 3.

5.6 References

Berg, C. W. 2014. SurveyIndex: R package for calculating survey indices by age from DATRAS exchange data. https://github.com/casperwberg/surveyIndex.

Berg, C. W., Nielsen, A., and Kristensen, K. 2014. Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. Fisheries Research, 151:91–99.

Pedersen, M. W., and Berg, C. W. 2017. A stochastic surplus production model in continuous time. Fish and Fisheries, 18(2), 226–243. https://doi.org/10.1111/faf.12174.

6 Expert Reviewer Comments: 2020 ICES WKDEM Benchmark Review

6.1 General comments across all assessments

Multiple assessments in our opinion were not in a sufficient state to be evaluated for a benchmark at the beginning of the week. We commend all assessors and their support for determination and hard earned progress over the week. However, in many cases reports were insufficient and there was pressure to make decisions with incomplete information. Assessment documentation for each species should include more thorough biological information about the species of consideration to allow each member of the panel to start from the same place with respect to basic knowledge of the species. Most of the following difficulties could have been avoided with more thorough preparation ahead of the meeting.

The estimation of uncertainty does not seem to be properly formalized under the ICES framework. Some of the assessments considered for this benchmark used retrospective analyses and some sensitivity runs to consider uncertainty. However, each assessment could have had an expanded set of sensitivity runs to fully explore all of the assumptions in the model including life-history information. In addition, the error envelopes shown with the model outputs were based on the estimated SE, which may be an underestimate of actualized uncertainty. While the uncertainty in SAM is based on SE from model outputs the lead developer did note that some dataset bootstrapping had shown that the uncertainty envelopes were similar to the ones output based upon SE. Bootstrap runs on the data for each assessment would have provided a better sense of the uncertainty for each species.

During this assessment review process, it became apparent that an age-structured assessment would not be available by the end of the week for one of the stocks. When a proposed base run model is rejected for use and the panel moves to another option, guidance on the process related to that transition would be useful. It appears that the default is to go to the last benchmark model for advice; however, that model may have the same problems as the current model. Thus, that option is a problem.

In most of the cases, model choice decisions were vague. There seems to be an unspoken drive to use SAM as the generic model for all stocks. No groups proposed multiple model choices, but instead spent the week adjusting various formulations of SAM. While SAM is a perfectly fine model, it was in some cases only being compared with itself. Heuristically this is a dangerous precedent. Much like the VPA was over-reliant on in the 1980s and 1990s, care should be taken to consider alternate model choices.

As a general comment, biological or fishery-process justification should precede changes in model structure. Tweaking models, then vaguely justifying choices is not part of assessment best practices. Both reviewers agree that some choices felt of the latter. In practice, adjusting models is simple, but it is useful to consistently remind that processes should drive models, not the other way around.

"An uncertain assessment is okay; the model should not be tweaked, and parameters should not be fixed unless there is sufficient information and evidence to make such assumptions. Rather than avoiding uncertainty, it should be accounted for by means of stochastic harvest control rules" ICES, 2019.

Across most of the assessments, visualizations were often hard to interpret. Most of this is likely due to pre-written scripts for assessment visualization. In most cases these were not sufficient, particularly when stocks had experienced large fluctuations in key parameters over time. A good example is retrospective plots. In our opinion, normal and standardized retrospective plots should be provided for all stocks. Additionally, Mohn's rho or similar analytical retrospective should be provided as it is required by ICES. The reviewers should not have to ask for this. There was no clear guidance was provided on what constitutes a major or minor retro and what to do about it. ICES should pursue general advice for retrospective interpretation.

Bycatch corrections and justifications were opaque at best across the WKDEM stocks. The assessors acknowledged the challenges associated with changes in country reporting over many of the stock time-series, but the justifications for specific actions were often unclear. Ways to improve this aspect would be organizing useful figures summarizing the state of stock reporting that could be referred to in reviews and subsequent assessments. Given the importance of missing catch (and particularly relative changes in missing catch over time), this should be a priority.

The choice of benchmark based on the stock—recruitment information and guidance from ICES was difficult given that some of these stocks did not conform to the options available. Since this is used to provide management advice, it is critical that the advice be tested to determine if the goals and objectives for the science and management of the stock are met. In addition, guidance from ICES on how time varying life-history information should be considered in benchmark calculations would be useful. Some of these stocks are using time-varying weights-at-age for the stock and catch. Using the most recent time period to reflect recent dynamics makes sense, but it assumes that long-term dynamics will not occur in the near term, which may not be true depending upon the species of interest.

During the panel discussions, concern was expressed over the differences in F time-series across the species of interest given the similar historical patterns in the fishery. The question arose as to whether or not there should be consistent trends across different species occupying a similar area. Many reasons exist as to why trends could be inconsistent across species including spatial dynamics, migration, bycatch issues, and fishery dynamics. While this was an interesting discussion point, the panel was not prepared to have this discussion formally because the data had not been prepared ahead of time. In addition, this is a multispecies consideration of the data and is likely much more complex than meets the eye. If this discussion is to arise in the future, the basic work to compare across species must be done well ahead of the review workshop.

7 WKDEM: Chairs' report and recommendations

The Chairs were reasonably satisfied with the progression of the Benchmark. The inclusion of two WebEx meetings (one prior to the Data compilation workshop and one prior to the final Benchmark meeting) aided in ensuring progress had been made, and it was clear what was to be presented and discussed at each of the two meetings. As a minimum, two WebEx-type meetings should be included in the Benchmark process itinerary.

Both chairs would like to commend the group that were undertaking the Northeast Haddock Benchmark. The way that this group approached and executed the Benchmark should act as a template for future Benchmarks. This was a diverse group covering science and industry, and it was a truly international effort. They initiated and undertook a number of meetings (including electronic ones) prior to the Benchmark organised WebExes, and data compilation and final Benchmark meetings. They compiled a comprehensive set of Working Documents which covered all the salient points relevant for the Benchmark and had them uploaded to the SharePoint well in advance of the meetings. They also ensured there were sufficient participants (in this case 11) who covered all the necessary expertise to undertake the work and had the experience with the ecoregion and the specific stock. All personnel were engaged throughout the Benchmark and contributed to the very timely completion of the work and their section of the report.

Benchmarks are made more difficult and the outcomes are not always realised when there is insufficient personnel and resources allocated. There is a need for more than one expert for each stock being benchmarked. We would strongly recommend that at least two competent assessors attend the final Benchmark meeting for each stock. In addition, there is a need for experts on any assessment model being used at a Benchmark. In the case of the SAM model, this was the case, and this proved extremely useful with respect to understanding the assessment model behaviour and settings.

For all Benchmarks there should be a commitment of personnel and resources to undertake the background and preparation work at the beginning of the Benchmark. Also, there is a need for a commitment to complete the required work ahead of time or at least on time. The ability to meet these commitments should be considered during the Benchmark prioritization process.

There is a necessity for data explorations to be complete and data to be in the required format before the deadlines in order for full explorations to be completed by the end of the Data Exploration workshop.

It is essential that, prior to the last Benchmark Workshop, the suite of potential assessment models should be agreed. As part of this process, preliminary runs should be made and presented, and documentation of each model be made available well in advance of the last Benchmark Workshop. This should be part of the standard protocols for Benchmark Workshops. It is not possible to fully evaluate models which are presented at the final Benchmark due to time constraints. Therefore, preliminary final runs of expert agreed/suggested runs of models need to be completed and documented prior to the start of the final Benchmark. This should not, of course, be taken to preclude further refining of the model solutions during the benchmark.

Annex 1: List of participants

Name	Address	Country	E-mail
Elena Balestri	Scottish Fishermen's Federation	UK (Scotland)	e.balestri@sff.co.uk
Robert Boenish Invited Expert	Environmental Defense Fund	US	robert.boenish@gmail.com
Olav Nikolai Breivik	Norwegian Computing Center	Norway	Olav.Nikolai.Breivik@nr.no
Bjarte Bogstad	Institute of Marine Research	Norway	bjarte.bogstad@hi.no
Anatoly Chetyrkin	Knipovich Polar Research Institute of Marine Fisheries and Oceanography(PINRO)	Russia	chaa@pinro.ru
Robin Cook	Department of Mathematics and Statistics University of Strathclyde	UK	robin.cook@strath.ac.uk
Kenny Coull	Scottish White Fish Producers' Association	UK (Scotland)	kenny@swfpa.com
Gjert Endre Dingsør	Norwegian Fishermen's Association	Norway	gjert@fiskebat.no
Helen Dobby	Marine Laboratory Marine Scotland Science	UK (Scotland)	h.dobby@marlab.ac.uk
Johanna Fall	Institute of Marine Research	Norway	Johanna.fall@hi.no
Jenni Grossmann	ClientEarth	UK	jgrossmann@clientearth.org
Daniel Howell Chair	Institute of Marine Research	Norway	daniel.howell@hi.no
Andrzej Jaworski	Marine Laboratory Marine Scotland Science	UK (Scotland)	jaworskia@marlab.ac.uk
Edda Johannesen	Institute of Marine Research	Norway	edda.johannesen@imr.no
Alexandros Kokkalis	DTU Aqua -National Institute of Aquatic Resources	Denmark	alko@aqua.dtu.dk
Yuri Kovalev	Knipovich Polar Research Institute of Marine Fisheries and Oceanography(PINRO)	Russia	kovalev@pinro.ru
Paul MacDonald	Scottish Fishermen's Organisation	UK (Scotland)	paul.macdonald@scottishfishermen.co.uk
Tanja Miethe	Marine Laboratory Marine Scotland Science	UK (Scotland)	t.miethe@marlab.ac.uk

Name	Address	Country	E-mail
By WebEx			
Richard D.M. Nash External Chair	Cefas Lowestoft Laboratory	UK	richard.nash@cefas.co.uk
Anders Nielsen	DTU Aqua -National Institute of Aquatic Resources	Denmark	an@aqua.dtu.dk
Alfonso Pérez-Rodriguez	Institute of Marine Research	Norway	alfonso.perez.rodriguez@hi.no
Alexey Russkikh	Knipovich Polar Research Institute of Marine Fisheries and Oceanography(PINRO)	Russia	russkikh@pinro.ru
Amy Schueller Invited Expert	NOAA	US	amy.schueller@noaa.gov

Annex 2: Titles and summaries of Working Documents presented to WKDEM

NEA haddock (Section 2)

The final WDs are accessible from benchmark participants on the WKDEM SharePoint. A brief summary is given below. The full Working Documents can be obtained by contacting the first author of each WD.

WD 3.1—3.7 were made in preparation to the data evaluation meeting held in December 2019.

WD 3.1—3.3: Catch data. No change in the input data proposed.

WD 3.4 and 3.7: Survey data. Decided to include new strata in indices from NORU winter survey.

WD 3.5 and 3.6: Biological parameters. Maturity-at-age and weight-at-age from NORU winter survey were updated and revised. A ratio correction accounting for the discontinuation of the Russian survey was proposed.

WD 3.1: Bogstad, B. and Russkikh, A. 2019. NEA Haddock – issues concerning catch reporting.

The status of knowledge on discards and IUU haddock is summarised. In 2007, international legislation on 'port-state control' was introduced. This has strongly reduced IUU catches that was due to trans-shipping and that has been a large problem in the early to mid-2000s. IUU estimates for 2002—2008 are included in the catch data used for the assessment. The estimates of discard and IUU for recent years are considered too small or too unreliable to be of use.

WD 3.2: Berg, H.S.F., Clegg, T. and Nedreaas, K. 2019. Unreported catches of haddock (*Melanogrammus aeglefinus*) in Barents Sea longline and Norwegian coastal gillnet fisheries.

The WD present preliminary estimates of unreported catches of haddock in the Norwegian longline fishery and discards of haddock in Norwegian coastal gillnet fishery.

WD 3.3: Blom, G. 2019. The Norwegian landing statistics of Northeast Arctic haddock (*Melanogrammus aeglefinus*) during 2000—2018 - effects of applying accurate conversion factors to convert product weights to round weight equivalents.

Conversion factors are used to calculate total catches from gutted and fish without head. Accurate estimates of conversion factors were compared to the official factors used as standard. The accurate conversion factor was on average (2000—2018) 3.24% higher than the standard one, but the difference varied by year.

WD 3.4: Johannesen, E., Russkikh, A.A., Prozorkevich, D.V., Kovalev, Y.A., Fall, JJ and Chetyrkin, A.A. Fishery-independent data.

The surveys are described and evaluated based on internal and external consistency, and their effect on the assessment results. Overall the consistency is good. The bottom trawl indices from the NORU survey was revised using 1 cm length groups instead 5 cm length groups for the length-dependent sweep width, and mean of bootstraps, but this had a very small impact on the indices. The NORU winter survey was extended in 2014 by including new strata to the north in response to reduced ice and expanding stocks. On average the including new strata increased the indices between 4–15% for ages used in tuning, leas for 3–4 year olds, and more for 5–9 year olds. The impact of the revised indices on the assessment was evaluated, and was found to be relatively minor. We decided to use the revised indices including new strata.

WD 3.5: Perez-Rodriguez, A, Korsbrekke, K. and Johannesen, E. 2019. Least Squares to Maximum Likelihood: haddock length, weight, proportion mature at-age from winter survey data.

The method of smoothing the maturity and weight-at-age, as first implemented at the benchmark in 2006, was described in detailed here. The weight-at-age and maturity data from the NORU winter survey was corrected for errors and based on the revised data, using ML and R, smoothed maturity- and weights-at-age were re-calculated for 1994–2019. In the revised the estimates, mature individuals skipping spawning was treated as immature, whereas previously the practice has been variable and not properly documented in the stock annex or in assessment reports.

WD 3.6: Russkikh, A., Johannesen, E., Kovalev, Y. and Chetyrkin, A. 2019. NEA haddock: Calculation of spawners proportion and stock weight-at-age when data from one of the surveys are absent.

The working document gives a brief summary on growth and maturity of NEA haddock, describes the practice of combing estimates from the Russian autumn survey and the NORU winter survey and the current practise for adjusting for holes in the survey series by using mean ratios between the two surveys. The Russian survey is discontinued. We evaluated the impact of using mean ratios on the assessment to account for the lack of the Russian survey also in the future. Given that the difference in estimates of weight and ogives between the two surveys has been fairly constant, we could keep the old data on weight and proportion spawners by age and use the ratio method in the next assessment. However, the approach can't be continued indefinitely, and a better solution needs to be found by the next benchmark, basing the weights-at-age and proportion spawners at-age on the survey data that are updated annually.

WD 3.7: Johannesen, E., Russkikh, A., Prozorkevich, D., Johansen, G.O. and Kovalev, Y. Ecosystem survey bottom trawl indices from StoX and BIOFOX.

The ecosystem survey is one of the two ongoing surveys that are used in tuning. The indices from this survey is calculated using as method developed by PINRO using the BIOFOX software developed and run at PINRO. BIOFOX cannot be used outside PINRO. IMR have proposed an alternative approach using swept area and the StoX software, a freely available software developed at IMR. In the WD we argue the strength and weaknesses of the two approaches. We agreed to use the BIOFOX method, but we should work on a common estimation method that could be run by both IMR and PINRO.

WD 3.8: Rodriguez-Perez, A., Dingsør, G., Breivik, O., Chetyrkin, A., Fall, J., Johannsen, E., Russkikh, A. and Kovalev, Y. NEA Haddock exploring SAM settings.

This WD was compiled in preparation for the main benchmark meeting. After the data evaluation meeting, the data were updated with also including plus group and three year olds from the winter survey in the data. From this dataset, we modified the model settings from last year's assessment trying different preliminary alternatives for catchabilities and observation variance.

WD 3.9: Russkikh, A., Bogstad, B., Johannesen, E., Kovalev, Y., Rodriguez-Perez, A. and Fall, J.

Short-term prediction input NEA haddock. This WD explore different settings for the short-term prediction, including recruitment, weight-at-age (stock and catch) and maturity using updated data. The assumptions on natural mortality and selection pattern was described.

WD 3.10: Kovalev, Y. A revision and update of Northeast Arctic haddock - population model.

This working document was done to check and update the population model, as well as to re-evaluate reference points and HCRs evaluation taking into account new data. It contains a review of work done for haddock during the ICES workshop on Management Plan Evaluation on North-east Arctic cod and haddock and Barents Sea capelin (WKNEAMP-2). The submodels for simulating haddock growth and maturation was tested using new and updated data. For some part of the model new methods/predictors were proposed. The run of the updated model confirms the reliability of currently used values of BRP, including B_{lim} and F_{MSY}.

Cod 6a (Section 3)

WD 4.1: Wright P.J., Régnier, T, Gibb, F.M. Implications of population structuring to Division 6a cod assessment.

1. Since the last benchmark there is new genetic evidence for the north of 6.a and a re-analyses of historic tag–recapture data that considers this evidence.
2. Three or four subpopulations of cod have been indicated from genetic, tagging and otolith chemistry studies that remain geographically separate throughout the year.
3. No significant genetic differentiation was found between cod sampled in the north of 6.a in depths >100 m and samples from Shetland in Division 4. Similarly historic tag–recapture results also indicate some west to east movement across 4°W. This subpopulation region accounts for most of the cod landed in 6.a since 2010.
4. Tagged cod from three of the populations appear to be inshore groups that tend to reside within 90 kms of their releases sites. However, there is some exchange between southern inshore groups and the Irish Sea.
5. Significant differences in maturity-at-age were found among the populations with those inshore maturing around a year younger.
6. The identified population structuring means that recovery and the effectiveness of measures is likely to differ regionally within the stock. Further trends in cod abundance in the north Division 6.a are unlikely to be independent of those in 4.a.

WD 4.2: Baudron, A.R., Régnier, T., Wright, P., Miethe, T. and Dobby, H. Review of maturity estimates for cod in Division 6.a.

Clear differences were observed between the maturity ogives of the four subpopulations of cod in ICES division 6a. Inshore subpopulations show faster maturation with almost all individuals reaching maturity by age 3, while the inshore subpopulation shows much slower maturation with all individuals reaching maturity by age 6. The overall maturity ogive for 6.a cod weighed by regional abundances is similar to the ogive observed for the North Offshore subpopulation owing to most individuals being caught in that area. These results were observed from both recent data (2015 to 2019), and data spanning the whole time-series (1995 to 2019).

This weighted ogive differs from the one used in the stock assessment and shows circa 20% (ogive obtained with data from 2015 to 2019) to 25% (ogive obtained with data from 1995 to 2019) of mature individuals at-age 1 contrasting with 0% in the assessment ogive (all subpopulations showed a proportion of mature individuals at-age 1 above zero, both with data from 2015 to 2019 and 1995 to 2019), and lower proportions of mature individuals from age 2 onwards.

The identification of a clear temporal pattern in the shape of the maturity ogive was hampered by the poor data quality in the earlier years of the time-series. Ogives obtained for the four sub-populations with a five-year time step showed no conclusive temporal trend. The annual weighted ogives estimated after the changes in survey protocol (2011 onwards) did appear show a more consistent pattern: an increase in the proportion of mature individuals at-ages 1 (and 2 to some extent) resulting in a decline in A50. However, in the absence of strong evidence for a long-term directional trend, this recent increase is probably not worth considering in the stock assessment.

The differences observed between the overall weighted ogive obtained with both recent (2015 to 2019) and long-term (1995 to 2019) data and the one currently used in the stock assessment suggest that it might be preferable to update the ogive. Indeed, the ogive used in the assessment appear to underestimate the proportion of mature individuals at-age 1, a pattern that was observed for all subpopulations. This could have a significant impact of spawning–stock biomass estimates. In the absence of strong support for temporal changes in maturity ogives, and given the similarities between ogives obtained with recent and long-term data, using a fixed maturity ogive based on all available data (i.e. overall weighted ogive for 6.a cod obtained with 1995 to 2019 data) might be worth exploring.

WD 4.3: Baudron, A.R., Miethe, T. and Dobby, H. Review of natural mortality estimates for cod in Division 6.a.

The results presented here show a clear decreasing trend in the weight of 6.a cod at all ages. This pattern towards smaller sizes is consistent with existing literature. As a result, there is a clear increasing trend in the natural mortality-at-age estimated with the Lorenzen equation. It would be worth exploring the impact of accounting for this directional trend in the stock assessment model, especially since the retrospective analysis presented here shows little impact overall when adding a few years of data. Indeed, it has been shown by Trijoulet *et al*. (2018) that allowing natural mortality to vary through time can result in a different trend in estimated fishing mortality compared to the current stock assessment model (which uses a constant mortality-at-age) which closer resembles that of models accounting for predation (different values but similar temporal variations). Accounting for the trend in natural mortality-at-age could be done by using smoothed weight-at-age as in out to both the Lorenzen equation and the stock assessment model for the sake of consistency.

The Lorenzen equation used in the assessment has been parameterised using data from both marine and freshwater systems, at all latitudes. Data from temperate latitude, which cod inhabits, were the vast majority of the observations used so this may not be an issue. However, data from freshwater systems (lake and rivers) amounted to almost two thirds of the observations (Lorenzen, 1996). Using empirical data from freshwater systems to parameterise an equation applied to marine species when marine data are available seems difficult to justify. It may be worth exploring the use of the equation parameterised with marine data only, also given in Lorenzen (1996), rather than the equation parameterised with all data which are currently used.

The increasing seal predation may be impacting the natural mortality of cod in Division 6.a, as shown by the existing literature. Indeed, models that account for seal predation estimate a natural mortality far higher than estimated with the Lorenzen equation, especially for ages 2 and 3, which seem to be the most targeted by seals. However, data on seal predation are scarce which would probably prevent the inclusion of seal predation mortality in a tactical assessment model. Most importantly, the impact of seal predation on the component of 6.a cod which is actually targeted by fisheries is being challenged (Hammond and Wilson, 2016).

WD 4.4: Dobby, H. Estimating area-misreported catch for cod in Division 6.a.

Based on the analysis conducted in this working document, the area-misreported component of the Scottish catch is estimated as follows:

- Total landings are estimated by distributing reported daily landings (within a trip) across fishing pings for that day (within a trip).
- Area-misreported landings are estimated as the difference between estimated and reported landings for Division 6.a.
- Area-misreported landings are assumed to have the same discard proportion as the Scottish demersal fleet operating in Division 6.a.
- Area-misreported landings and discards are assumed to have the same age compositions as the Scottish demersal fleet operating in Division 6.a.

WD 4.5: Dobby, H. Preparation of catch data in InterCatch for Division 6.a cod.

WD 4.6: Sánchez, B.R., Jaworski, A., Clarke, L. and Dobby, H. Cod 6.a: Biomass estimates from the West Coast Demersal Fish Project.

This manuscript reports on the estimates of absolute biomass (and length compositions) for Division 6.a. cod which use area swept by the fishing gear and estimates of the herding coefficient. Analysis investigating the impact of an area closed to demersal fishing (known as 'the Windsock') was also carried out as part of this project (and based on the offshore surveys), but is not reported on here. In general, in the inshore surveys, cod were caught in very low numbers, with the exception being more moderate catches of juvenile cod in the Clyde in two of the four seasonal surveys. Due to the wide range of fishing gear used among trips on these surveys, biomass estimates have not been presented and this document focuses on the results of the offshore surveys.

WD 4.7: Barreto, E. and Clarke, L. Estimates of Cod Biomass in ICES Division 6.a for 2014–2019.

In 2005, Fisheries Research Services (FRS, now Marine Scotland Science - MSS) started a new survey to estimate the abundance and distribution of anglerfish on the Northern Shelf. Initially, and again in recent years, the survey has included contribution from research vessels of the Marine Institute in Ireland and is called the Scottish Irish Anglerfish Megrim Industry Science Survey (SIAMISS). This survey covers much of the area of the known distribution of northern shelf anglerfish (ICES Divisions 4a, 6a and 6b at Rockall), with the exception of the central and southern parts of Subarea 4 and the Skagerrak and Kattegat (Division 3a). Because it covers such a large area, the current design incorporates multiples vessels to survey the whole area. Although the focus of the survey is anglerfish and megrim, data on the other main commercial species are also recorded. This document focuses on the cod occurring on these surveys, in ICES Division 6a (West of Scotland) and presents biomass estimates for the last six years (2014–2019).

WD 4.8: Cook, R. An Assessment of cod in ICES Division 6a using a Time Varying Natural Mortality model.

An assessment of cod in 6.a is described using a model that estimates time-varying natural mortality. The model appears to fit the observed data reasonably well with plausible estimates of natural mortality. Additional models are used to investigate sensitivity to alternative assumptions about catch errors, exploitation patterns, seal predation and survey data. Most models suggest fishing mortality has declined by around 50% since the 1980s, and that SSB reached a minimum around 2006. While all models suggest some recovery in SSB, the strength if the recovery is uncertain with a further recent decline to below B_{lim}. Retrospective runs of the time-varying M model show good consistency over a five year period. The inclusion of M as a variable enables more realistic estimates of uncertainty in the assessment.

WD 4.9: Dobby, H. Cod in 6a: Reference Points.

This documents the calculation of reference points based on the final SAM assessment agreed at WKDEM, including as discussion on the choice of B_{lim}, stock–recruitment function and sensitivity of F_{MSY} to the biological parameters.

Whiting 6a (Section 4)

WD 5.1: Jaworski, A. Catch data for whiting in Division 6.a (West of Scotland).

WD 5.2: Jaworski, A. Combined abundance index for whiting in Division 6.a from the Q4 Scottish and Irish surveys.

The combined index for the two Q4 surveys seems to provide a more complete representation of population levels. It also simplifies, to some extent, the modelling procedure in the annual assessments of the stock.

WD 5.3: Jaworski, A. Estimation of the maturity ogive for whiting in Division 6.a.

WD 5.4: Kokkalis, A. and Jaworski, A. Assessment model for whiting in Division 6.a (West of Scotland).

Previous assessments of whiting in Division 6.a were conducted with the Time-Series Analysis (TSA, Gudmundson, 1994; Fryer, 2002; Needle and Fryer, 2002). At that time, the stock was classified as category 1. During this benchmark process, it was found that running TSA with the new data and changed survey configuration posed a serious challenge. Poorly converged optimisation runs in conjunction with excessive running times were a major obstacle to completing the assessment successfully. In these circumstances, it was decided to run the benchmark assessment using the age-aggregated stochastic Surplus Production in Continuous Time (SPiCT) model (Pedersen and Berg, 2017). At the same time, the stock category was changed to category 3 according to the ICES guidelines for data-limited stocks (ICES, 2019a).

Whiting 3a (Section 5)

WD 6.1: Berg, C.W. “Survey index calculations for whiting in Division 3.a and adjacent Waters.”

The proposed biomass index is based on data collected by four scientific surveys that cover the stock area, namely the two international bottom trawl surveys (NS-IBTS, BITS) and two Danish national surveys that target cod and sole. The working document describes the method of calculation of the new index. A Tweedie-GAM is used to model total catch (in weight) per haul as a function of the following smooth functions: time invariant spatial effect, seasonal repeating pattern, space–time interaction effect, depth effect. The logarithm of haul duration is used as an offset, i.e. the underlying assumption is that the catch is proportional to trawling duration. Fixed effects are used to model differences between gears and random effects to are used for interaction of ship and gear.

WD 6.2: Kokkalis, A. “Commercial landings and discards.”

Whiting in Division 3a is mostly caught by Danish fleets (60–90% in the last decade) unwanted catch because of its low value and is discarded. A considerable amount of catches comes from the Danish industrial fleet (11–82% of the total landings). The information about commercial catches is collated from three different sources, InterCatch, the official nominal catches that each country is submitting and ICES is providing, and the previous reconstruction of the catch time-series in the North Sea assessment working group (WGNSSK).

WD 6.3: Kokkalis, A. “Assessment with the surplus production model SPiCT.”

The working document describes all attempted scenarios, differing by input data and model settings. The conclusion is that there is no scenario that can be directly used to provide TAC advice for whiting in Division 3.a.

Annex 3: Terms of reference and agenda for the Benchmark meeting

Terms of reference

a) Evaluate the appropriateness of data and methods to determine stock status and investigate methods for short-term outlook taking agreed or proposed management plans into account for the stocks listed in the text table below. The evaluation shall include consideration of:
   1. Stock identity and migration issues;
   2. Life-history data.
   3. Review current sampling levels and adjust stratification levels for landings and discards accordingly;
   4. Examine alternative assessment models to the current model;
   5. Explore impact of all tuning fleets on assessment estimates;
   6. Further inclusion of environmental drivers, multispecies information, and ecosystem impacts for stock dynamics in the assessments and outlook;
   7. Examine mixed fisheries interaction.

b) Agree and document the preferred method for evaluating stock status and (where applicable) short-term forecast and update the stock annex as appropriate. Knowledge about environmental drivers, including multispecies interactions, and ecosystem impacts should be integrated in the methodology. If no analytical assessment method can be agreed, then an alternative method (the former method, or following the ICES data-limited stock approach) should be put forward;

c) Re-examine and update (if necessary) MSY and PA reference points according to ICES guidelines (see Technical document on reference points);

d) Develop recommendations for future improving of the assessment methodology and data collection;

e) As part of the evaluation:
   1. Conduct a three-day data evaluation workshop. Stakeholders are invited to contribute data (including data from non-traditional sources) and to contribute to data preparation and evaluation of data quality. As part of the data compilation workshop consider the quality of data including discard and estimates of misreporting of landings;
   2. Following the Data evaluation, produce working documents to be reviewed during the Benchmark meeting at least seven days prior to the meeting.

Agenda

Monday 10th February

10.00	Welcome, roundtable, housekeeping
10.20	ICES code of conduct, conflict of interest
10.30	Any ICES presentations
11.00	Begin presentation of the models for each stock
12.00–13.00	Lunch
13.00	Continue stock by stock presentations
15.00	Break
18.00	Close

Tuesday 11th February

9.00	Work as required
10.30	Break
12.00–13.00	Lunch
15.00	Break
18.00	Close

Wednesday 12th February

9.00	Work as required
10.30	Break
12.00–13.00	Lunch
15.00	Break
18.00	Close

Thursday 13th February

9.00	Finalising model specifications by stock
10.30	Break
12.00–13.00	Lunch
13.00	Begin reference points
15.00	Break
15.30	Continue per stock model specifications and reference points
18.00	Close

Friday 14th February

9.00	Planning report writing schedule
9.30	Reference points and finishing off per stock model specifications
10.30	Break
12.00–13.00	Lunch
16.00	Close

Annex 4: Agenda for the ICES WKDEM benchmark (Data Compilation)

ICES Headquarters, Copenhagen, Denmark, 9–13 December 2019 (Monday to Friday)

Chair: Richard D.M. Nash (Cefas, UK)

Begin at 12:00 the first day, 09:00am thereafter. Aim to close each day by 18:00, and finish at 12:00 on the final day.

Stock order

Had.27.1–2 stock leader(s) Alexey Russkikh, Edda Johannesen

Cod.27.6a stock leader(s) Helen Dobby

Whg.27.6a stock leader(s) Andrej Jaworski

Whg.27.3a stock leader(s) Alexandros Kokkalis

Monday

12.00 Opening of Meeting

Introduction, housekeeping and round table

12.30 Overview of work

13.00 Presentation of stocks:

Had.27.1–2 stock leader(s) Alexey Russkikh, Edda Johannesen

15.00 Break

15.30 Haddock, continued

17.00 Close

Tuesday

9.00 Work on specific stocks (order to be determined)

Cod.27.6a stock leader(s) Helen Dobby

12.00 Lunch

Cod continued

Whg.27.6a stock leader(s) Andrej Jaworski

Return to Haddock

18:00 Close

Wednesday

9.00 Work on specific stocks (order to be determined)

Whg.27.3a stock leader(s) Alexandros Kokkalis

Return to Haddock

13.00 Lunch

Return to Cod 6a, Return to Haddock (David Miller attending), Return to Cod 6a

18:00 Close

Thursday

9.00 Work on specific stocks (order to be determined)

NEA haddock, 6a Cod, 6a whiting, 3a whiting, 6a cod

13.00 Lunch

6a cod, 6a whiting.

Friday

9.00 Wrapping up of outstanding issues

10.00 3a whiting update, round table of completed tasks and timetable for completion of remaining tasks.

12.00 Close of meeting

General Points

1. All WD to be completed and loaded in to 04. Working Documents (relevant folder) on the WKDEM 2020 SharePoint by 17th January 2020.
2. A document explaining the relevance of each WD to the Benchmark process to be included in each WD stock folder by 24th January 2020.
3. All relevant data for input to assessment to be lodged in the appropriate folder in 06. Data on the WKDEM 2020 SharePoint.
4. Final WebEx prior to the Benchmark (presentation of results) January/February 2020. Start time xx:xx CET.

Issues list from the Data Compilation Workshop

Had.27.1–2 stock leader(s) Alexey Russkikh, Edda Johannesen

<u>IUU and discards</u>
No change to the input data, however there will be documentation available.

<u>Fishery-independent surveys</u>
Winter survey: include the extended area from 2014 onward.

BioFox and STOX for estimating indices for the joint Russian/Norwegian Barents Sea Ecosystem Survey (BESS). Both indices to be brought to the Benchmark, decision as to whether to maintain *status quo* (BioFox) or switch to an alternate (STOX). Ongoing investigations prior to the Benchmark.

A document to Sven Kupschus (ICES surveys) with an outline of the problem and ask for an expert opinion on the methodologies in relation to submitting an index to an assessment.

WDs re BioFox and STOX, along with summaries and documentation of deliberations. Reviews of the two methodologies/platforms to be uploaded before the 1st January.

<u>Biological parameters e.g. weights-at-age</u>
Weights in the catch to continue with the historic times up until the Russian survey ceases then to continue with the weights from the Norwegian winter survey with an invariant correction factor. The continuation gives values of a similar magnitude to the current time-series.

Note this is a short/medium solution which needs to be given a more permanent solution by the next Benchmark.

Mortality rates and consumption by cod
To remain as currently implemented. Note that this is very reliant on the cod assessment.

Maturity ogives
Needs to be rechecked for all datasets: any skipped spawners are not included in the mature stock i.e. treated as immatures.

Discards
To be included where possible and document (at least for the WG).

Conversion from dressed weights to whole weights
To be noted but the new relationships cannot be implemented here as complex due to no age data. Some uncertainty in the landings data as well.

Catch data sampling
Catches sufficiently sampled but not at the recommended level. No change to current data practices or catch matrices.

Cod.27.6a stock leader(s) Helen Dobby

Stock structure and substocks
0^{o} meridian probably a better division between eastern and western stocks.
Coastal versus offshore (resident versus migratory)
Spatial considerations
Mosaics of (meta) populations

Weightings for population characteristics and catches etc.
Mosaics of (meta) populations
A number of population characteristics including a single age matrix.

Discards – spatial and characteristics of the discarded fish
See below

Variable maturity
Taken from the surveys with a smoother.
WD to be on the SharePoint by 20th December.

Natural mortality
A number of suggestions, problem with the reliance on grey seal predation due to a mismatch with the spatial distribution.
Use of Lorenzen equation from mean weight-at-age to generate M. Use peels and Mohn's rho as an objective way of viewing the best choice. There is concern that M could vary over the time-series quite considerably on an annual basis and thus annually change the perception of the stock.
WD: Plots with a number of spans plus one for a GAM. Table with Mohn's rho (relative and absolute) with GAM. Text to describe results and suggest a preferred option for implementing M.

Surveys
The combination of the West of Scotland with Irish survey to provide a single time-series was discussed. Decided to keep them separate at the moment. This may be revisited before the Benchmark but highly unlikely.

Area misreporting, Catch data and discards
Area misreporting using VMS and then the associated discarding rates. Misreporting rates and composition assumed to be similar to the Scottish fleet. WD to be uploaded.

Selectivity by fleets
This was considered but there was insufficient time and insufficient resources to investigate this further with the resources available.

Whg.27.6a stock leader(s) Andrej Jaworski

Catch data
To be finalized before the Benchmark.

Survey data
Combining the WoS and Irish Q4 surveys.

Maturity ogives
New average, invariant ogive for this year, however, to investigate using historical data produce a time-series.
WD to be on the SharePoint by 20th December.

Natural mortality
Should this use the Lorenzen equation with the mean weights-at-age?
To replicate what was done for cod.
WD: Plots with a number of spans plus one for a GAM. Table with Mohn's rho (relative and absolute) with GAM. Text to describe results and suggest a preferred option for implementing M.
Insights as to GADGET outputs re mortality rates – WD?

Weights in the projections
Change to using the previous year as an alternative to a three-year mean.

Whg.27.3a stock leader(s) Alexandros Kokkalis

Objective: move from a category 5.2 to a category 3 stock!

Landings data
Official versus InterCatch still needs to be resolved.
Some inconsistencies in the length data.
Revising landings data, a 'work in progress'.

Survey data
Worked up to a Biomass index for 3a using four surveys. Still to look into the sensitivity of the index to losing a survey for some reason.

Annex 5: Stock Annexes

The table below provides an overview of the stock annexes updated at WKDEM. Stock Annexes for other stocks are available on the ICES website Library under the Publication Type "Stock Annexes". Use the search facility to find a particular Stock Annex, refining your search in the left-hand column to include the *year*, *ecoregion*, *species,* and *acronym* of the relevant ICES expert group.

Stock ID	Stock name	Last updated	Link
cod.27.6a	Cod (*Gadus morhua*) in Division 6.a (West of Scotland)	April 2020	WoS cod
had.27.1–2	Haddock (*Melanogrammus aeglefinus*) in subareas 1 and 2 (Northeast Arctic)	February 2020	NEA haddock
whg.27.3a	Whiting (*Merlangius merlangus*) in Division 3.a (Skagerrak and Kattegat)	April 2020	Skagerrak and Kattegat whiting
whg.27.6a	Whiting (*Merlangius merlangus*) in Division 6.a (West of Scotland)	April 2020	WoS whiting