ICES WKREBMSE Report 2018

ICES Advisory Committee

ICES CM 2018 / ACOM:52

Report of the Workshop on the evaluation of harvest control rules for *Sebastes mentella* in ICES areas 1 and 2 (WKREBMSE)

June – August 2018

By correspondence

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

Recommended format for purposes of citation:

ICES. 2018. Report of the Workshop on the evaluation of harvest control rules for *Sebastes mentella* in ICES areas 1 and 2 (WKREBMSE), June–August 2018, by correspondence. ICES CM 2018/ACOM:52. 32 pp.

Contents

1 Terms of Reference

WKREBMSE - Workshop on the evaluation of harvest control rules for *Sebastes mentella* in ICES areas 1 and 2

2018/2/ACOM50 The Workshop on the evaluation of harvest control rules for *Sebastes mentella* in ICES areas 1 and 2 (WKREBMSE), chaired by Daniel Howell*, Norway, and attended by two invited external experts Colm Lordan, Ireland and Sarah Kraak, Germany, will be established and will meet by correspondence during June-August 2018 to:

a) Evaluate, according to ICES guidelines, whether the assessment for this stock should be treated as a Category 1 or Category 2 (relative) assessment. Determine appropriate reference points for the stock depending on the category chosen.

b) Address the request from Norway and Russia for an evaluation of a set of proposed harvest control rules for *Sebastes mentella* in ICES areas 1 and 2 (reb.27.1-2).

c) Draft advice for the reb.27.1-2 stock (including catch scenarios for those HCR considered precautionary).

WKREBMSE will report by 24 August 2018 for the attention of the ACOM.

2 Introduction

A pelagic fishery for *Sebastes mentella* has developed in the Norwegian Sea outside EEZs since 2004. This fishery is managed by the Northeast Atlantic Fisheries Commission (NEAFC). In the Barents Sea, where fisheries are managed by the Joint Norwegian-Russian Fisheries Commission (JNRFC), there had been no directed fishery for this stock from the 1990s until 2013. A new directed demersal and pelagic fishery was opened in the Norwegian Economic Zone in 2014. The 44th Session of JNRFC decided to split the total TAC among countries as follows: Norway: 72%, Russia: 18%, Third countries: 10% (as bycatch in the fishery protection zone at Svalbard (Spitsbergen): 4.1%, and international waters of the Norwegian Sea (NEAFC-area): 5.9%). This split was reconducted at the 47th session of JNRFC in 2017, but there is no agreement between JNRFC and NEAFC concerning this split.

No management plan exists for this stock. From 1995 to 2012, the ICES advice had been no directed catch/lowest possible level. From 2013 onwards, the basis of the advice has been somewhat ad hoc and has varied (F0.1 in 2013, status quo catch in 2014, precautionary approach in 2015–2018).

ICES was requested to test a wide range of harvest control rules for this stock. The request is given in Annex 1 and the list of participants in Annex 2.

This working group was conducted by correspondence, with reviewer comments available after the draft report was finalized. We have therefore inserted some sections to address the comments and concerns.

1.1 Category 1 vs category 2

Given that the *S. mentella* assessment model was recently benchmarked by ICES as a category 1 stock, and that the WKREBMSE is being conducted by correspondence and without active reviewer participation during the work of the group, WKREBMSE is reluctant to recommend reclassifying this stock as category 2. We would recommend that such a judgement is better made at a benchmark meeting with a broader participation and reviewer involvement in the discussions. However, the workshop notes that there is a significant uncertainty in the absolute biomass level associated with the somewhat arbitrary choice of the catchability for a specific survey (ICES, 2018a). Historically, a c. 10-year period (1974–1985) of catches around and above 100 000 tonnes led to a reduction in stock size and subsequent catches. Given the current assessed good stock status, applying an F from an HCR will result in initially large catches, and there is concern that these could be damaging to the stock if there is a bias in the assessment.

This uncertainty is, to some extent addressed in Section 1.9 where we investigate the sensitivity of HCR performance to bias in the overall estimated biomass. We find the adding of a cap on annual TAC of 50 000 tonnes to each of the HCRs analysed proves precautionary to relatively large biases in the assessment (50%). This catch is below the short-term peak in the various HCRs, but above the long-term stable catches. Given that this is a long-lived stock, such a cap would not impact severely on long-term catches, and would provide more stable year-to-year catches, and smooth changes in biomass levels. The WK therefore recommends that the stock can be managed as a category 1 stock, with two provisos to be precautionary against a risk of stock collapse caused by unidentified biases in the assessment. These provisos are:

1. Whatever HCR is adopted should include, alongside a formula for target F at different stock sizes, an absolute cap on TAC. This cap to be set to be in line with sustainable historical catches, ie. around 50 000 tonnes.
2. The trend in assessed biomass be monitored, and the HCR be revised if the stock is seen to be declining significantly more rapidly than projected.

The projections of catches under the different HCRs analysed here include a short-term peak in catches, before the long-term catches stabilize. It is this short-term peak which poses the greatest risk if there is a bias in the stock assessment. Given the long-lived nature of this species, a cap on catches would act to "smooth out" this peak, providing increased precautionarity by reducing the risk of rapid stock collapse without negatively impacting on the long-term yield-per-recruit from the stock. We stress that because of the suspected bias in the assessment model, a HCR without a cap on annual TAC cannot be said to have been assessed as sustainable in this workshop.

1.2 Reference points and changes in time series since WKREDFISH 2018

At the recent Benchmark (WKREDFISH, ICES, 2018a), the range of SSB in time series was (324 kt,1088 kt) and the reference points were then set to B_{lim} = 324 kt (lowest observed) and B_{pa} = 450 kt (using the magic formula with σ = 0.2, i.e. B_{pa} = exp(1.645 σ). At AFWG 2018 (ICES, 2018b) weight at age of 19+ was revised and set to constant over time, and also a new year of data (2017) was added. There was a considerable effect on the SSB due to the revision of 19+ weight at age and the range is now (227 kt, 940 kt). The two SSB time series are shown in Figure 1. Using the same approach as at WKREDFISH 2018 would give B_{lim} = 227 kt and B_{pa}= 315 kt, and we suggest these values as reference points for this stock.

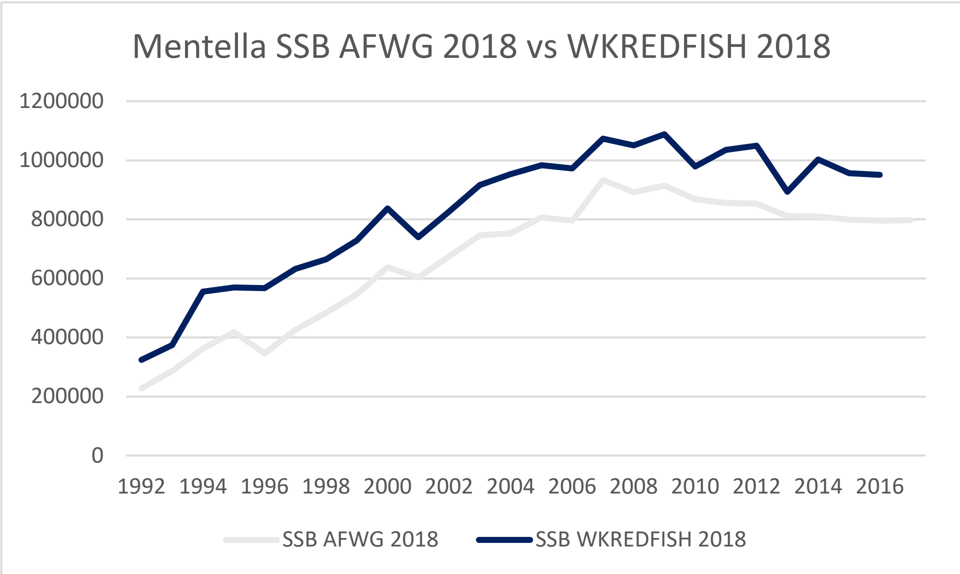


Figure 1. Comparison of SSB time series from WKREDFISH 2018 and AFWG 2018

We note that there is no legal requirement for the fishery in the Barents Sea to be conducted on a F_{MSY} basis, and that the HCRs evaluated here are based on those stated in the request rather than on an F_{MSY} basis. However, for completeness, we present such calculations here. In previous workshops (ICES, 2014; ICES, 2018a) $F_{0.1}$, estimated using yield per recruit (YPR) analysis, was proposed as a candidate F_{MSY}. Whilst the calculation was done for ages 12–18y earlier, WKREDFISH (ICES, 2018a) suggested to do the calculation for the 19+ group as it represents the bulk of the targeted biomass. Here $F_{0.1}$ was estimated for both age groups. For the 19+ group it amounts to 0.084 and to 0.074 for ages 12–18y. The estimated values for F_{max} are 0.236 (19+ group) and 0.207 (12–18y), respectively. The yield-per-recruit function is depicted in Figure 2. We stress that these calculations were conducted on a yield-per-recruit basis, and that these values have not been tested for precautionarity.

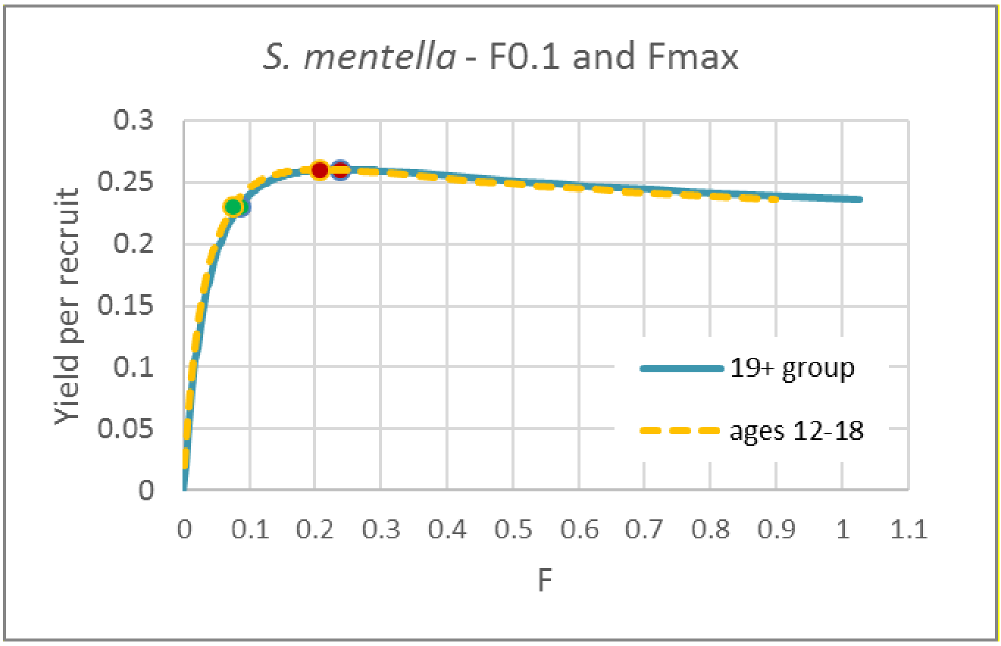


Figure 2. Yield per recruit (YPR) for age groups 12–18y and 19+. Green dots indicate $F_{0.1}$ and red dots F_{max} for the respective age groups.

1.3 Amendments and additions to the request

Following the change in reference points, we added simulations for B_{pa} = 315 kt as trigger point, for all three F values, in total this gives 3x4 = 12 runs. Following those runs we found it adequate also to investigate how given fixed TAC levels (cap on TAC) would perform in the medium term if the current stock size is overestimated. The reason for this will be explained below.

The option with reducing F at low recruitment was not run due to limitations in software. Instead, we explored how HCRs with (0.06, 450 and 0.08, 450) would have worked if implemented from 1992 onwards, i.e. how those rules would have coped with the actual recruitment failure observed.

1.4 Evaluation procedure

The PROST software was used for making long-term stochastic simulations. PROST is a tool for making single-fleet, single-area long-term stochastic projections and was used for the simulations. It is available on the ICES web page (http://www.ices.dk/marine-data/Documents/Software/WD2.%20prostguide.pdf).

PROST has previously been used in the evaluation of harvest control rules for northeast Arctic cod, haddock and saithe as well as during evaluation of HCRs for this stock during WKREDMP in 2014. In total, 10 000 simulations were run for each HCR. No assessment was conducted within the MSE simulations, rather the assessment was assumed to be "true population plus noise". Variability in the population in the operating model was achieved by having variation in the initial population and stochastic variability in the recruitment. In addition, separate simulations were run to investigate the impacts of potential bias in the assessment. These are detailed in the following sections.

1.5 Operating models

1.5.1 Biological model

1.5.1.1 Initial stock numbers

The stock size at the beginning of 2018 was taken from the last assessment (ICES, 2018a). The stock was projected through 2018 (intermediate year) assuming fishing mortality to be the same in 2018 as in 2017 (it is expected that catch level is approximately the same in 2018 as in 2017). The recruitment at age 2 for the 2016 and 2017 year classes was calculated based on two linear regressions between survey indices (i.e. the 0-group survey and the 5–9 cm fish in the winter (February) survey) and number at age 2 for the period 1992–2017 taken from the most recent stock assessment (ICES, 2018a). The coefficients of determination were $r^2 = 0.63$ and $r^2 = 0.62$ for the two regressions, respectively. The average value of the results of regressions was used. The values obtained for the 2016 and 2017 year classes at age 2 were 266 and 216 million, respectively. The 2018 year class at age 2 was set to the recent 10-year average (2006–2015 cohorts): 300 million.

Stochasticity was added to the projections by including uncertainty in the values for number-at-age for 2018. Uncertainty was higher for the most recent year classes (CV = 0.2 on log scale for the year classes prior to 2016 and 0.3 for the year classes 2016 and 2017).

1.5.1.2 Natural Mortality, weight-at-age and maturation

These are all assumed to be constant. M = 0.05, Weight-at-age – modelled – last line in Table 6.7 AFWG 2018, Maturity-at-age are modelled values (Table 6.19, modelled values used for 2014–2017). Weight-at-age in stock and catch are equal and the proportion of F and M before spawning is set to zero.

1.5.1.3 Recruitment

A spawning stock-recruitment plot for cohorts 1992–2015 is shown in Figure 3. Figure 4 shows the recruitment time series.

Figure 3. Recruitment age 2 vs SSB. Cohorts 1992–2015.

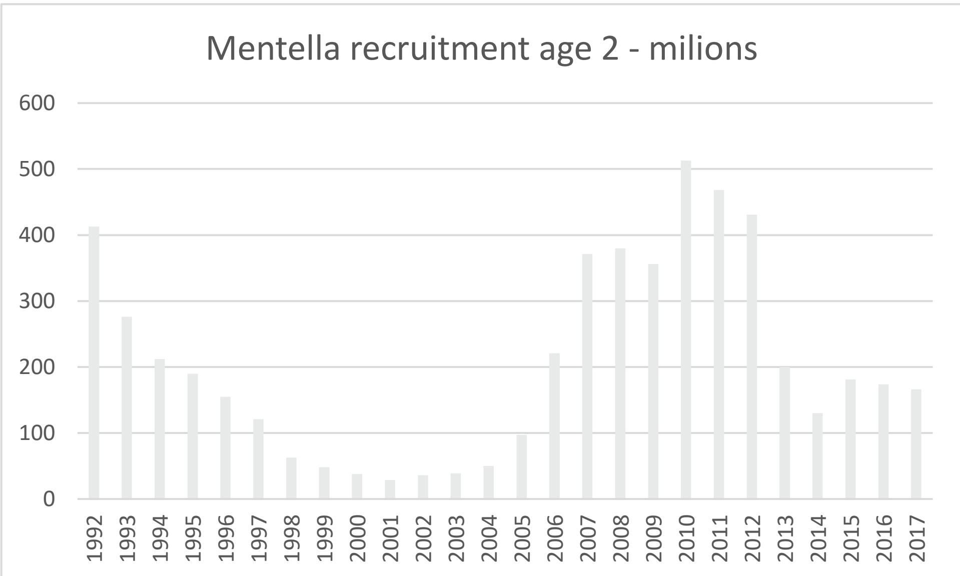


Figure 4. Recruitment time series (age 2 in 1992–2017 i.e. cohorts 1990–2015).

There is not much evidence of relationship between recruitment and spawners in either plot, and there are strong time trends in the series (1996–2003 year classes poor).

We fitted a log-normally distributed recruitment function, independent of SSB, given by: $R = \alpha e^{\varepsilon}$

where ε is normally distributed, ε =N(0,σ). The plateau α was estimated to 139 million and the error term σ was estimated to 0.90 using the method outlined by Skagen and

Aglen (2002). The error term is smaller than in 2014 evaluation as recruitment in recent years is closer to average. The fit to the data is not very good, as shown in Figures 5–7. Whatever happens to recruitment below B_{lim} does not matter for these simulations, as long as the risk level is calculated as the chance of dropping below B_{lim} at least once in a defined period, which is the approach we decided to take.

Figure 5. Recruitment function (with log-normal error) fit to data for spawning stock and recruitment at age 2 for the cohorts 1992–2015.

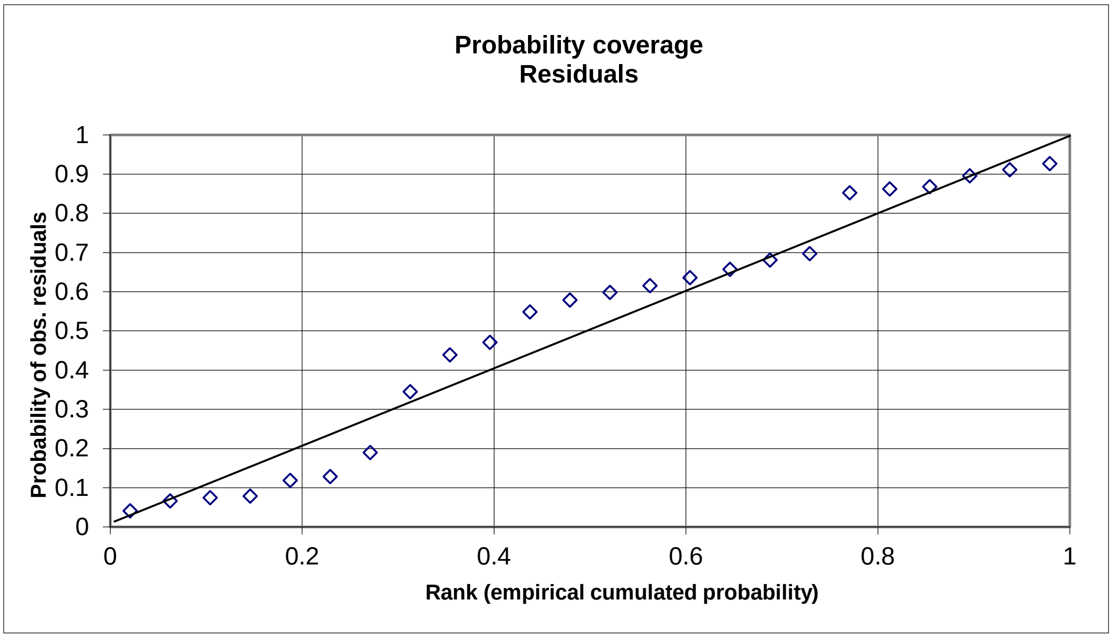


Figure 6. Probability coverage for stochastic stock–recruitment function.

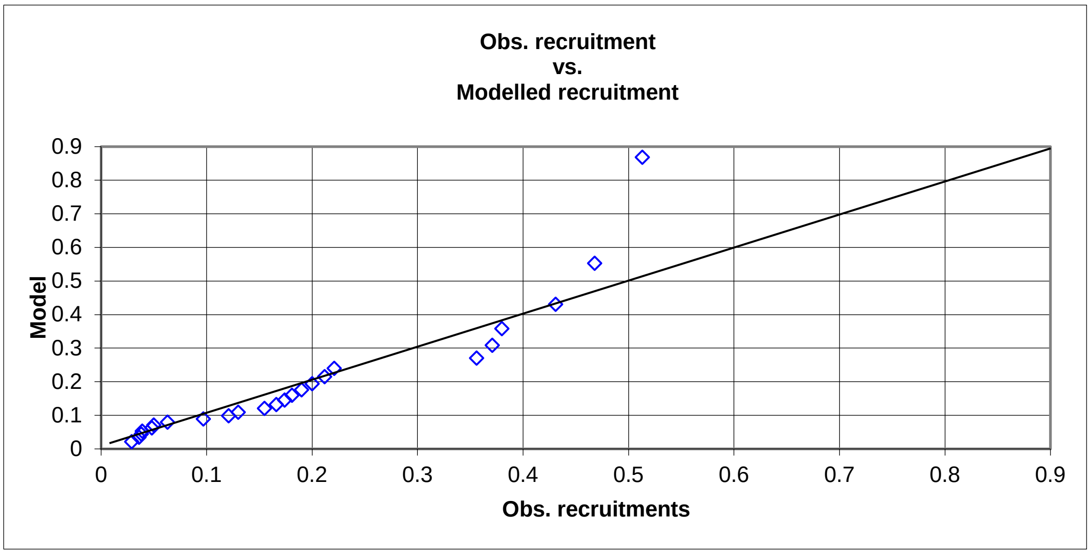


Figure 7. Observed vs. modelled recruitment for stochastic stock–recruitment function.

1.5.1.4 Selection at age

We calculated a recent average selectivity at age based on averaging the 2015–2017 total fishing mortality at age (thus assuming that the selectivity in each fishery and the pelagic/demersal ratio of catch in tonnes to continue unchanged in the future). This is shown in Figure 8. If there were to be major changes in the fisheries then a fresh analysis is recommended.

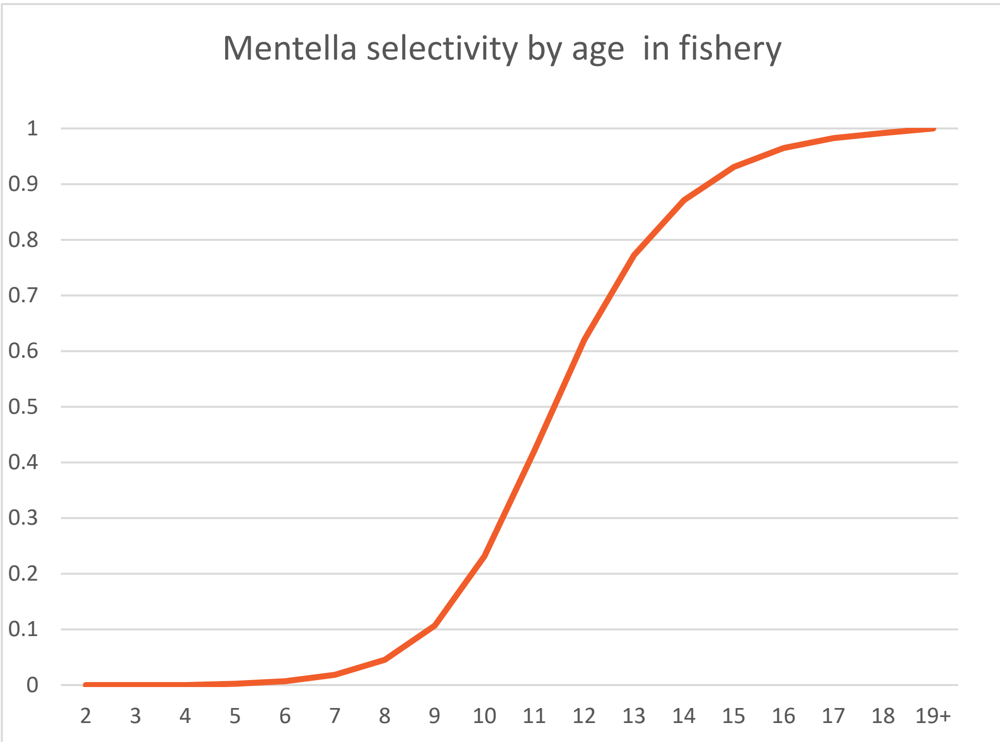


Figure 8. Selection pattern calculated as average of values for 2015–2017 from AFWG 2018.

1.5.2 Observation model

The simulations were not full feedback (i.e. an assessment was not run each year in the projection). Uncertainty in deriving the perceived view of the true stock (i.e. accounting for observation error in future catches and indices and model error) is included in a single 'assessment error' term. The assessment error was set to CV = 0.2 on log scale for all age groups in all years. The assessment error in a given year is uncorrelated between age groups. In the first instance the assessment was assumed to be unbiased, we investigate the possible effects of bias later.

1.5.3 Summary of stochasticity modelled

As mentioned before, stochastic long-term simulations are performed. For clarification, the following components are stochastic:

- Initial stock size
- Annual stock assessments
- Recruitment function (independent of SSB)

Making these components stochastic are considered to be sufficient to describe the overall uncertainty in the biological model, uncertainties in weight, maturation, mortality and fishing pattern are thought to be of minor importance in comparison.

1.6 Reality check

A run was made for 100 years with $B_{trigger}$ = 450 kt and F = 0.08. Catches and stock size seemed to level off at plausible levels (see Figure 9). SSB < B_{pa} in 10% of years. The average catch levels in this simulation was considered to be plausible in light of the historical catch level (Figure 10); the average since 1952 is 40 kt, ten year averages 1952–1961, 1962–1971 … 2002–2011, 2012–2017: 30, 17, 121, 50, 12, 13 and 22 kt respectively).

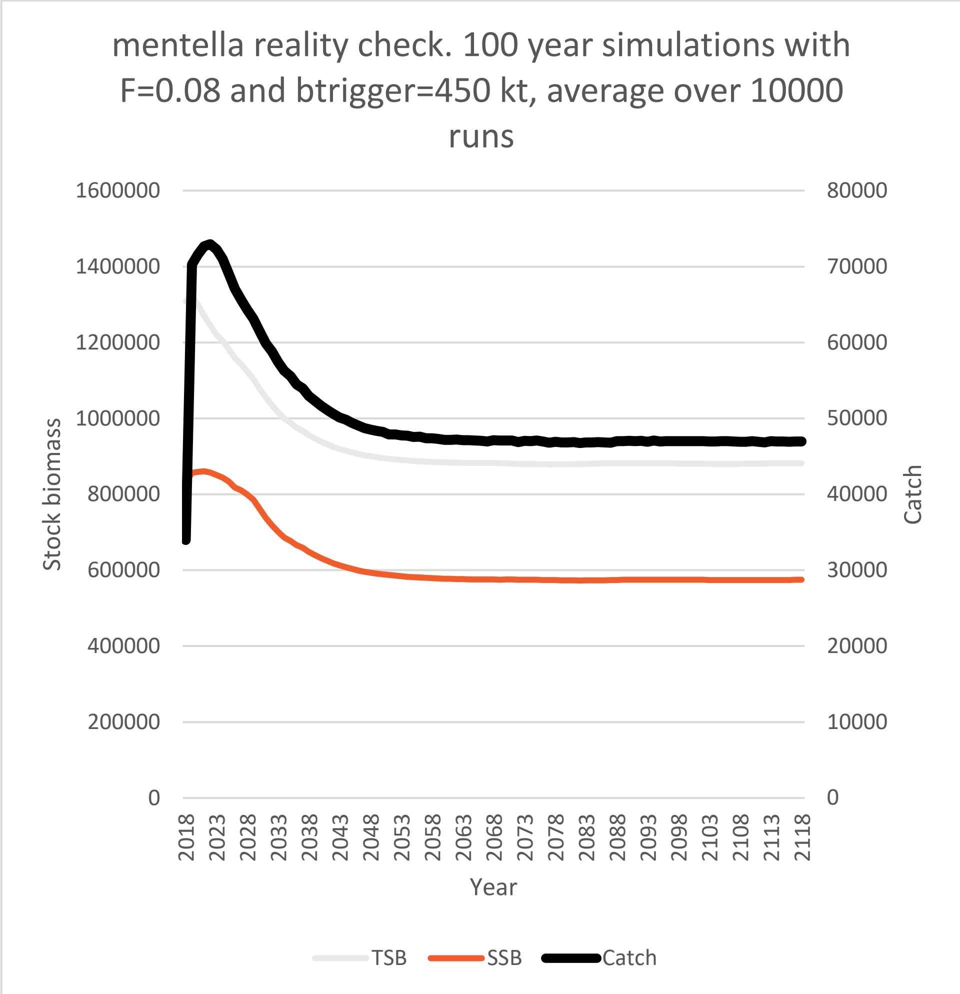


Figure 9. Results of reality check with F = 0.08 and $B_{trigger}$ = 450 kt.

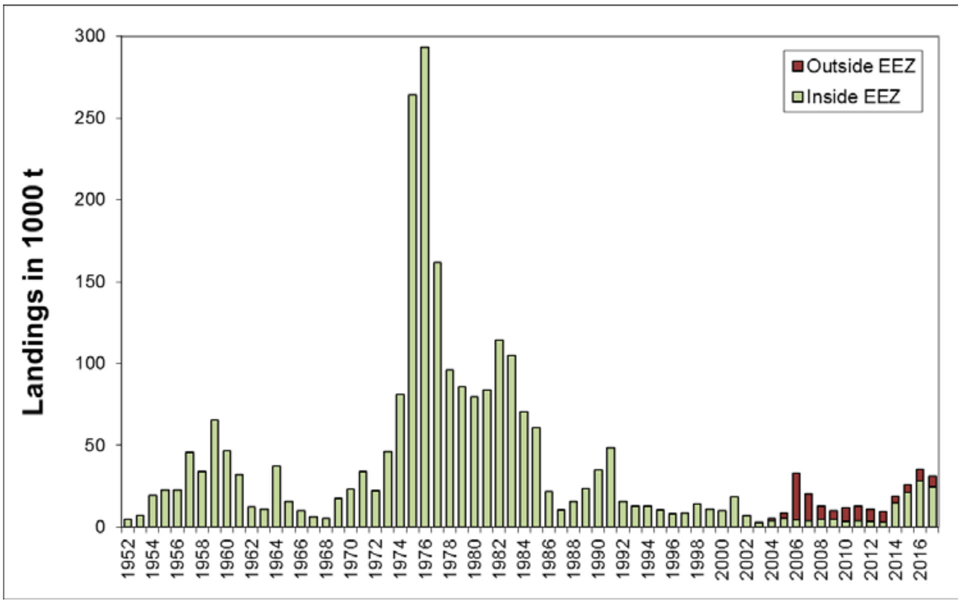


Figure 10. Historical catches of *S. mentella*.

1.7 Results

12 combinations of F and $B_{trigger}$

Following ICES WKGMSE 2013, and given that we use a stock-recruitment relationship which is independent of SSB, prob2, the probability for true SSB to be below B_{lim} at least once during the simulation period should be the appropriate quantity to use as far as B_{lim} is concerned. We also calculated the probability of SSB < $B_{trigger}$ in a given year (). The results are given in tables 1a–b.

Table 1a. Yield and spawning-stock biomass (SSB) for the harvest control rules (HCRs) examined.

HCR		Yield (kt)							SSB (kt)
$B_{trigger}$	Target F	2019	2020	2021	2022	2023	Average 2019–2028	Average 2019–2068	Average 2016–2068
315	0.06	52.9	55.1	56.8	57.8	58.2	56.4	48.1	778.7
315	0.08	70.2	71.5	72.7	72.8	72.3	69.7	54.5	661.5
315	0.10	87.0	87.0	87.3	86.2	84.2	81.0	59.0	573.4
450	0.06	53.2	55.0	56.7	57.8	58.2	56.4	48.1	778.0
450	0.08	70.1	71.5	72.6	72.9	72.2	69.6	54.3	662.2
450	0.10	86.8	86.8	87.1	86.0	83.8	80.7	58.6	578.2
600	0.06	53.0	54.9	56.7	57.8	58.1	56.5	47.8	781.7
600	0.08	70.0	71.4	72.5	73.0	72.1	69.6	53.6	673.7
600	0.10	86.9	87.1	87.1	86.0	83.9	80.7	57.0	598.4
800	0.06	51.3	53.9	56.1	57.4	58.0	55.9	46.4	800.0
800	0.08	67.9	69.8	71.4	72.0	71.3	68.3	51.4	708.4
800	0.10	83.7	83.7	84.4	84.9	83.9	79.6	54.6	645.0

Table 1b. Risk levels, realised F and TAC variability (50-year period) for runs 1–12. Risk levels above 0.05 are shown in red. Yields and biomass in kt. 5yr = 2019–2023, 10yr = 2019–2028, and 50yr = 2019–2068.

HCR		Prob(SSB< B_{lim})			Prob(SSB< $B_{trigger}$)			Realized F			Interannual TAC variation		
$B_{trigger}$	F	5 years	10 years	50 years	5 years	10 years	50 years	5 years	10 years	50 years	50 years	Precautionary to observed recruitment failure	Precautionary to 50% assessment bias over 10 years
450	0.06	0.000	0.00000	0.000	0.00000	0.00000	0.004	0.06000	0.06000	0.0600	12.9	YES	YES
450	0.08	0.000	0.00000	0.000	0.00000	0.00000	0.062	0.08000	0.08000	0.0793	13.3	YES	Only with 50 kt TAC cap
450	0.10	0.000	0.00000	0.000	0.00000	0.00000	0.214	0.10000	0.10000	0.0972	14.7	Not tested	Only with 50 kt TAC cap
600	0.06	0.000	0.00000	0.000	0.00092	0.00046	0.108	0.05999	0.05997	0.0590	15.0	YES	YES
600	0.08	0.000	0.00000	0.000	0.00088	0.00044	0.351	0.07996	0.07996	0.0759	17.4	YES	Only with 50 kt TAC cap
600	0.10	0.000	0.00000	0.000	0.00136	0.00178	0.578	0.09993	0.09981	0.0900	19.5	Not tested	Only with 50 kt TAC cap
800	0.06	0.000	0.00000	0.000	0.16394	0.09213	0.505	0.05864	0.05904	0.0552	20.9	YES	YES
800	0.08	0.000	0.00000	0.000	0.23538	0.27121	0.766	0.07765	0.07751	0.0683	23.1	YES	Only with 50 kt TAC cap
800	0.10	0.000	0.00000	0.000	0.36992	0.60061	0.897	0.09596	0.09405	0.0792	23.9	Not tested	Only with 50 kt TAC cap
315	0.06	0.000	0.00000	0.000	0.00000	0.00000	0.0001	0.06000	0.06000	0.0600	12.7	Not tested	YES
315	0.08	0.000	0.00000	0.000	0.00000	0.00000	0.0006	0.08000	0.08000	0.0800	12.1	Not tested	Only with 50 kt TAC cap
315	0.10	0.000	0.00000	0.000002	0.00000	0.00000	0.012	0.10000	0.10000	0.0998	11.8	Not tested	Only with 50 kt TAC cap

All tested HCRs without assessment bias (runs 1–12) are considered precautionary, as the probability (prob2) of SSB < B_{lim} in any year is zero or extremely low. Trigger points of 800 kt (all three F values) and 600 kt (F = 0.08 and 0.10) will result in high probability of SSB < $B_{trigger}$ in any given year and will thus cause larger interannual variation in TAC as target F often will change between consecutive years. Trigger points of 450 and 315 avoid this issue, and a trigger point of 450 seems plausible. The effect of trigger point on mean yield is rather small.

Concerns with bias in assessments and advice in the short term

The present assessed stock size is somewhat above the long-term mean stock size obtained using the recruitment function. Thus, recommended catches in the short term (3–5 years) will be higher than the mean long-term catches (about 55, 70 and 85 kt, respectively for F = 0.06, 0.08 and 0.10 compared to the long-term mean of 40 kt). These catch levels are also well above the TAC 2018 = 32.658 kt, which is the highest since the early 1990s. Given this and the uncertainty in current stock size (see ICES, 2018a; b), this gives reason for concern, and a plausible approach could be to introduce a cap on the TAC in the harvest control rule. We decided to investigate how fixed TACs of 50 and 60 kt would work in a 10-year period, assuming that the current stock size is overestimated by 25 or 50% (runs 13–16). The values for the cap were chosen based on the catches during historical time periods where the stock was not depleted by the fishery. The results (5% percentile of SSB for each run) are shown in Figure 11, compared to the reference points B_{lim} and B_{pa}. Obviously, the reference points would also be reduced if the present stock size is an overestimated (not necessarily by the same percentage), but for discussion let us use the reference points already obtained. For simplicity and robustness we use catches fixed at the TAC ceiling for these runs, any simulation which is precautionary here would also be precautionary if the HCR actually resulted in lower catches in some years.

Run 13: -25%, fixed TAC 50 kt

Run14: -50%, fixed TAC 50 kt

Run 15 -25%, fixed TAC 60 kt

Run16: -50%, fixed TAC 60 kt

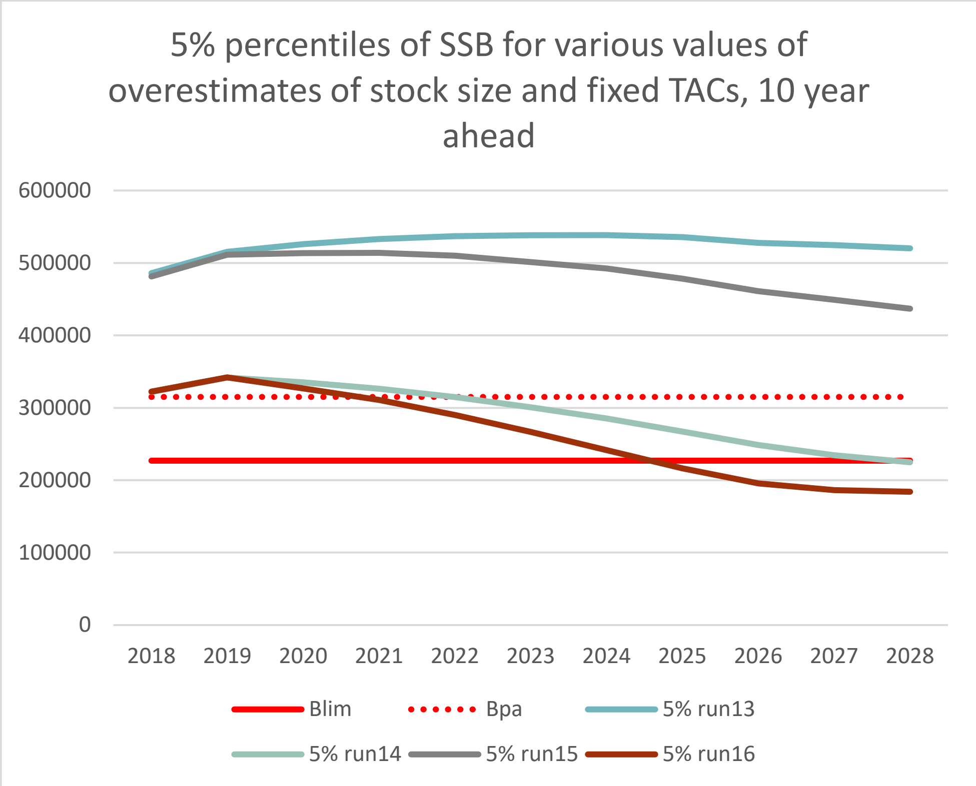


Figure 11. 5% percentiles of SSB for various values of overestimates of stock size and fixed TACs, 10 year ahead, vs. reference points.

The results show that for a 25% underestimation, either a 50 kt or a 60 kt cap would leave the stock well above B_{lim} after 10 years. However, the for a 50% underestimate the 95% percentile falls below B_{lim} for the 60 kt cap, but remains above it for the 50 kt cap. Based on this a cap of 50 kt on TAC does seem reasonable and would for a 10-year period even cover the situation with a 50% overestimate. 50 kt in 2019 corresponds to an increase of about 50% from the 2018 TAC. Obviously, a low F would be an alternative approach to increasing the precautionarity of the HCR, but this would come at the expense of lowered catch in all years. In contrast a cap on TACs only impacts on the catch in the years with highest catch, and we consider this to be a better compromise to achieve good overall yield while minimizing the risk of stock collapse.

The above results show that a cap of 50 kt is precautionary to a 50% bias in the assessment. The question then arises of if it necessary to maintain precautionarity. We present results below of an analysis of the risks involved in running a HCR with different F levels without such a cap.

F levels of F = 0.06, F = 0.08 and F = 0.1 were simulated forwards with no cap on catches for 10 years assuming a 50% bias in assessment, with a $B_{trigger}$ at 450 kt. The mean and 95th percentile of the F = 0.1 run went below B_{lim} in the ten year period. The F = 0.08 mean remained above B_{lim}, but the 95th percentile went below B_{lim}, and is therefore not precautionary to the investigated 50% bias in the assessment. We note in passing that the yield-per-recruit based F_{MSY} of 0.0825 is higher than the F = 0.08, and would also therefore fail this precautionarity test. Finally, the F = 0.06 level remained above the B_{lim} value, both the mean and the 95th percentile.

We therefore conclude that a cap on TACs is required for precautionarity to uncertainty in the level of assessment at the level investigated here if the target F is 0.08 or above, but that a target F of 0.06 is precautionary with or without a cap on TAC.

Handling of recruitment failure

Reducing F when incoming recruitment is low – does the harvest control rule need extra precaution for that?

In order to investigate how this would have worked out during the recruitment failure around year 2000 (8 consecutive years with bad recruitment), we did a run starting in 1992, with the number at age taken from the assessment. We then ran simulations to 2018, using the same recruitment, weight, mortality and maturity at age as in the assessment time series, but applying the HCR and using the same (2015–2017 average) exploitation pattern as in the other forward simulations. We tested this for F19+ = 0.06 and 0.08 and $B_{trigger}$ = 450 kt. The results are shown in Figure 12, with 5% and 50% percentiles for both rules, showing that for both those rules the 5% percentile of SSB is above B_{pa} in all years. This indicates that at least for this rule no extra precaution for incoming bad recruitment of the kind we have experienced, is needed. However, in the event that a period of poor recruitment were to be observed we would expect that further investigation would be conducted at that time. Given that multiple years of observation would be needed to confirm such a period of recruitment failure and the usual practice of periodically reviewing and revising HCRs, we do not find the need to add anything on this topic to the HCR.

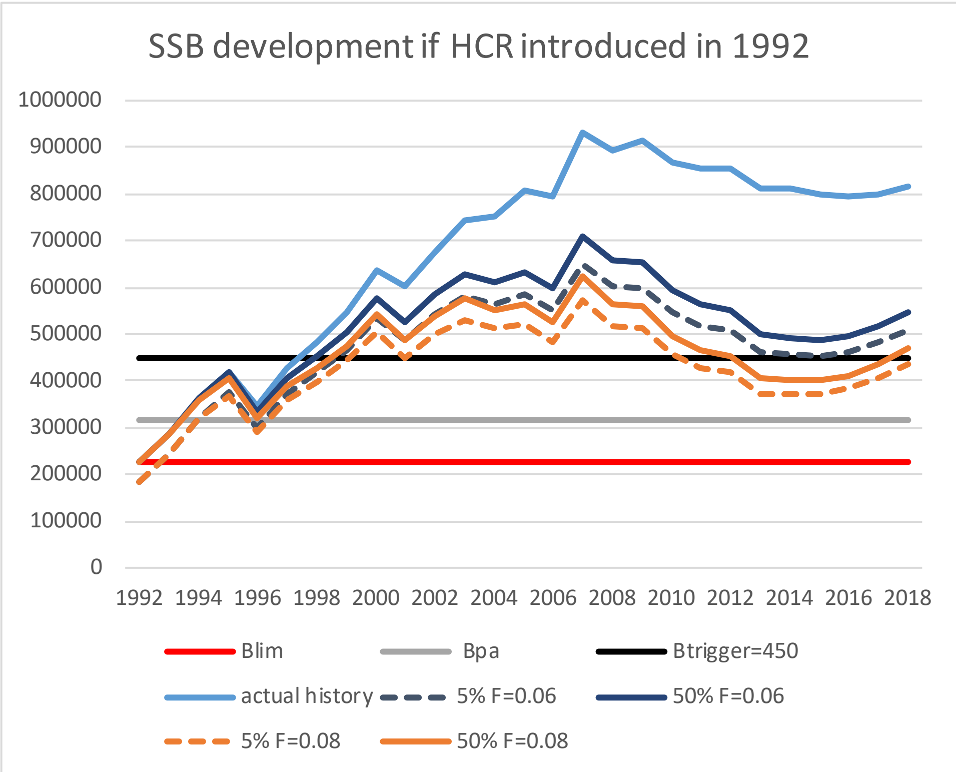


Figure 12. SSB development if HCR introduced in 1992, stock dynamics including recruitment unchanged.

1.8 Factors not taken into account in the simulations

The simulations assume that weight at age 19+ is constant, implying that the internal age structure in the 19+ group is not taken into account. As some growth in weight is thought to occur past age 19, this will bias the results, indicating optimal yield at slightly higher F value than would be the case if growth in weight past age 19 had been accounted for. Also, possible maternal effects favouring older spawners is not considered. The latter issue was discussed by WKREDMP 2014 (ICES, 2014). Both of these factors would indicate preferring HCRs with lower F values where performance is otherwise similar.

During WKREDMP in 2014 and also this year, the relationship between recruitment and "mature biomass of at least a given age" was examined with the data available. Since "19+" is the plus group reported in the current analytical assessment and is therefore available for the entire time period 1992–2017, we report this here. At WKREDMP in 2014, it was found that using 19+ biomass ($R^2 = 0.77$) instead of "gonad based SSB" ($R^2 = 0.42$) gave a much clearer signal between adult biomass and recruitment. However, the difference between using SSB19+ and total SSB is now in the other direction, as illustrated when comparing figures 3 and 13.

The level of possible bias in the assessment is unknown and this limits the level of detail in the precautionary measures that could be evaluated to account for this. If more information were to become available on quantifying the uncertainty around the assessment then it would be possible to re-visit this work.

One could obviously also go beyond the text of the request and create and evaluate a much wider range of possible HCRs, with different trigger points and target F levels. The working group has mostly restricted itself to those actually requested, with the exception of adding an extra (lower) trigger point following the re-estimation of stock reference points as described above.

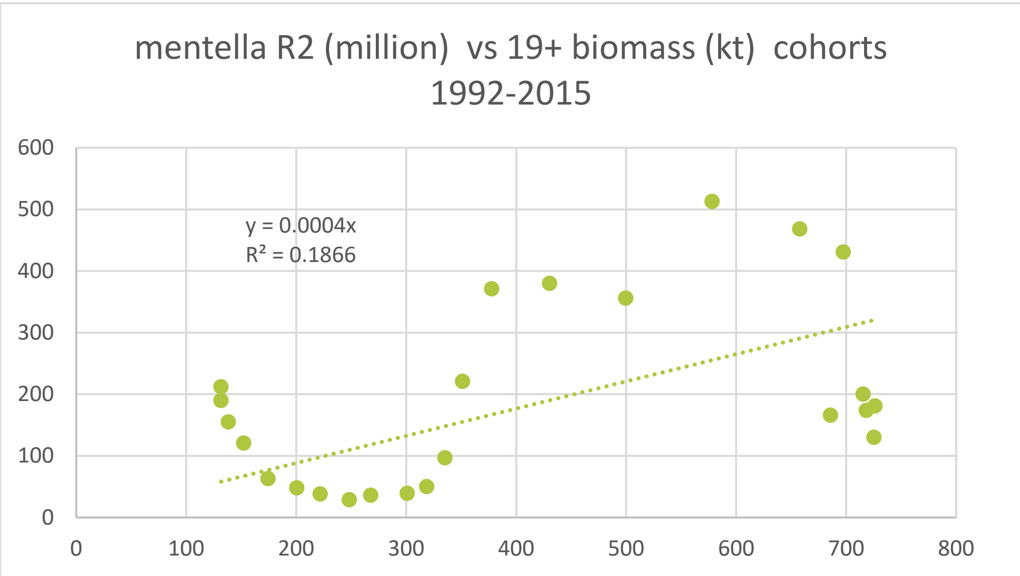


Figure 13. Spawning–stock biomass (SSB) age 19+ (bottom) vs. age 2 recruitment during 1992–2015 as estimated by the last ICES assessment (ICES, 2018a).

1.9 Discussion and conclusions

ICES considers that, for a long-lived, slow growing, late maturing stock any management action will take longer than five years before changes in the biomass are likely to be detected. Therefore, ten years seem to be a more sensible time span to assess the impact of a harvest control rule. The life-history characteristics of this stock also make it vulnerable to overfishing, and once overfished, the recovery might take decades. ICES therefore recommends applying a rather conservative management approach.

Based on the HCRs specified in the request, the simulations made and the discussion above, ICES recommends a harvest control rule with $B_{trigger}$ = 315 kt (based on the B_{pa} value) and F19+ = 0.06 or 0.08, with a cap on TAC of 50 000 t.

After the next surveys on mature *S. mentella* in the Norwegian Sea, which are planned for 2019 and 2022, there should be more info on absolute stock size and it should be possible to include this survey in the assessment model, also it will be known whether the good year classes after 2003 have recruited to the mature stock. Thus, the HCR should be re-evaluated in 2023. We note that more advanced and flexible MSE tools (e.g. full assessment model, more flexible recruitment patterns,…) are in development at IMR, and should be in place for such a re-evaluation.

ICES also recommends a re-evaluation of the HCR be part of the research around any future period of poor recruitment.

ICES recommends that at the next benchmark a thorough evaluation be made as to the appropriate ICES category (1 or 2) for this stock.

1.10 References

ICES. 2013. Report of the Workshop on Guidelines for Management Strategy Evaluations (WKGMSE) , 21 - 23 January 2013, ICES HQ, Copenhagen, Denmark. ICES C.M. 2013/ACOM:39, 121 pp.

ICES. 2014. Report of the Workshop on Redfish Management Plan Evaluation (WKREDMP), 20–25 January 2014, Copenhagen, Denmark. ICES CM 2014/ACOM:52. 269 pp.

ICES 2018a. Report of the Benchmark Workshop on Redfish Stocks (WKREDFISH). 29 January – 2 February 2018, Copenhagen, Denmark. ICES CM 2018/ACOM:34, 174 pp.

ICES 2018b. Report of the Arctic Fisheries Working Group (AFWG), 18–24 April 2018, JRC, Ispra, Italy. ICES CM 2018/ACOM:06, 857 pp.

Skagen, D.W., and Aglen, A. 2002. Evaluating precautionary values of fishing mortalities using long-term stochastic equilibrium distributions. Annex 7(WD 7) in: ICES Study Group on the Further Development of the Precautionary Approach to Fishery Management, Copenhagen 2-6 December 2002. ICES C. M. 2003/ACFM:09. 144 pp.

Annex 1: Request

Request to ICES for evaluation of harvest control rule for Sebastes mentella:

"Norway and Russia ask ICES to evaluate the following set of harvest control rules for *Sebastes mentella* in ICES areas 1 and 2:

All combinations (total 3x3=9) of the following elements:

Fishing mortality (F19+) of 0.06, 0.08 and 0.10

Trigger points of 450, 600 and 800 kt

In addition, for a trigger point of 450 kt and F19+=0.08, a rule with the following additional clause should be tested: "Reduction of F by 50% or no reduction of F if the average strength at age 2 for the year classes which are 3-12 years old in the first year for which the TAC advice is given, is below 100 million individuals."

In all cases F should be reduced linearly towards F=0 at SSB=0 if SSB in the first year for which TAC advice is given, is below the trigger point.

If none of the rules are found to be precautionary, rules with additional values of F and trigger point should be investigated in order to find rules which are precautionary.

For all simulations, ICES is asked to assess the consequences through calculating the following performance indicators (expected values):

Annual yield during each of the next 5 years

Medium term yield, represented as average yield during the next 10 years

Long term yield and SSB, represented as average during the next 50 years

Probability that SSB falls below $B_{trigger}$, B_{pa} and B_{lim}, in a 5, 10 and 50 year period

Realised average fishing mortality

Annex 2: List of participants

Name	Institute	Country	e- mail
Daniel Howell (Chair)	Institute of Marine Research	Norway	daniel.howell@hi.no
Bjarte Bogstad	Institute of Marine Research	Norway	bjarte.bogstad@hi.no
Daisuke Goto	Institute of Marine Research	Norway	daisuke.goto@hi.no
Hannes Höffle	Institute of Marine Research	Norway	hannes.hoffle@hi.no

Annex 3: Reviewers' comments

Colm Lordan, Marine Institute, Galway Ireland & ACOM Vice-Chair

The Workshop, chaired by Daniel Howell, Norway, and attended by three invited external experts, addressed the following three terms of reference (ToRs) by correspondence.

a) Evaluate, according to ICES guidelines, whether the assessment for this stock should be treated as a Category 1 or Category 2 (relative) assessment. Determine appropriate reference points for the stock depending on the category chosen.

b) Address the request from Norway and Russia for an evaluation of a set of proposed harvest control rules for *Sebastes mentella* in ICES areas 1 and 2 (reb.27.1-2).

c) Draft advice for the reb.27.1-2 stock (including catch scenarios for those HCR considered precautionary).

ToR a.

The WK did not address whether the stock should be treated as Category 1 or 2; *WKREBMSE is reluctant to recommend reclassifying this stock as category 2.*

However, a Category 1 assessment would need and uncertainty cap on the TAC of 50kT and a second proviso to monitor SSB and *"and the HCR be revised if the stock is seen to be declining significantly more rapidly than projected"*, although this is not defined. This approach seems reasonable.

Regarding the biomass reference points the AFWG revised B_{lim} = 227kt based on the revised treatment of the mean weighst 19+ group in the assessment which is appropriate. Regarding the B_{pa} value of 315 kt which is derived with $\sigma = 0.2$, the justification for this value is not provided and it seems that the σ for the final SSB estimate from the assessment would be more appropriate.

Fishing mortality reference points were derived from a yield per recruit rather than EQSIM. There is no information in the report why this was the case. The F0.1 for 19+ group it amounts to 0.084 and to 0.074 for ages 12–18y, the F_{max} are 0.236 (19+ group) and 0.207 (12–18y), respectively. There is no firm conclusion from the ADG on F reference points. Given the uncertainties in the assessment, it would seem to this reviewer that F0.1 would be a candidate as a proxy for F_{MSY}.

ToR b.

The MSE was carried out using a shortcut procedure (i.e. the assessment was not included) using the PROST software. Given the concerns expressed about bias in the assessment it would be preferable to simulate include and test the assessment within the Management Strategy. Instead, a sensitivity analysis was carried out including up to 50% bias and the TAC cap of 50kT. While this approach seems sensible, it could also be considered arbitrary; efforts should be made to address this more objectively in the future.

Regarding the inclusion of stochastic in the assessment this is not overly clear or transparent what values were used and what was their basis? It is stated that the main sources of uncertainty are included but this is not properly documented and discussed in the report.

There appears to be strong auto-correlation in recruitment. There is also no apparent S-R relationship in the relatively short time series (given the longevity of the species). The choice of a log-normally distributed recruitment function is considered appropriate. However, given the history of several years of very weak recruitment in a row it should be highlighted the MSE would need to be re-evaluated if this were to occur again notwithstanding the WKs exploration of this.

The statement "*We stress that because of the suspected bias in the assessment model, a HCR without a cap on annual TAC cannot be said to have been assessed as sustainable in this workshop*." To some extent undermines the risk estimates presented in Table 1b.

Figure 11 confuses me, the runs with 60kT appear to have a higher SSB, maybe there is a labelling issue or I misunderstand.

The rational for the recommendation of a harvest control rule with B_{pa}=450kt and F19+=0.06 or 0.08, with a cap on TAC of 50 000 t should be justified better in the conclusion. Why was a Btrigger of 450kt used rather than the new value? Why have the Fs of 0.06 and 0.08 been recommended rather than an F of 0.084 (F0.1) which has some scientific basis. Why 50Kt cap not 60kT cap? It is important that we guide the managers with a clear rational for choosing options given that all options could be considered precautionary based on the results of the MSE which may not fully account for bias.

ToR C

The Summary sheet as drafted has only included a 50kT cap option while other options could be considered precautionary. It would be useful for the ADG to consider if other options are also needed.

Overall Conclusion

Some clarifications are needed to justify the choice of 450kt as the $B_{trigger}$ and the calculation of the new B_{pa}. There is a risk's to biomass reference points that are presented in the analysis are likely under-estimates, that said the absolute level of bias in the assessment is not known. The rational for the recommendations need a stronger scientific justification.

Review by Sarah B. M. Kraak, Thuenen Institute of Baltic Sea Fisheries

The Workshop, chaired by Daniel Howell, Norway, and attended by three invited external experts, addressed the following three terms of reference (ToRs) by correspondence.

d) Evaluate, according to ICES guidelines, whether the assessment for this stock should be treated as a Category 1 or Category 2 (relative) assessment. Determine appropriate reference points for the stock depending on the category chosen.

e) Address the request from Norway and Russia for an evaluation of a set of proposed harvest control rules for *Sebastes mentella* in ICES areas 1 and 2 (reb.27.1-2).

f) Draft advice for the reb.27.1-2 stock (including catch scenarios for those HCR considered precautionary).

ToR a.

Regarding the first ToR the WK did not actually evaluate whether the assessment should be treated as a Category 1 or Category 2 assessment. Instead the WK argued that since the assessment was recently benchmarked by ICES as a Category 1 stock, it was reluctant to recommend reclassifying this stock as Category 2. The rest of section 1.3. in the WK report is devoted to stating that there is substantial uncertainty about the absolute biomass level and arguing that therefore a cap on annual TAC of 50 kt should be added to the HCR. Thus, while the observation that the absolute biomass level is highly uncertain might have led to the reclassification of the assessment to Category 2, the WK decided not to recommend doing so. It does not seem relevant to the question of the first ToR that the WK proposed a cap on the TAC; it rather seems to imply that the WK apologises for the fact that they were reluctant to recommend reclassification and offers this precaution instead. Nevertheless, later in the report the WK argues why the cap should be used.

Regarding the question of the reference points under the first ToR, the WK argues that since at AFWG 2018 the SSB time series was revised from the one established at the Benchmark earlier this year (WKREDFISH 2018), using the same approach for determining the reference points as in the Benchmark (WKREDFISH 2018) would give Blim=227kt and Bpa=315 kt, and the WK suggests these values as reference points for this stock. In the absence of a legal requirement for Fmsy, the WK nevertheless estimated F0.1 and Fmax for the 19+ group as well as for ages 12-18y. The reviewer agrees with these conclusions.

ToR b.

The WK used the PROST software for making long-term stochastic simulations, as was done for this stock during WKREDMP in 2014. The software seems to have some limitations: apparently it was not possible to simulate the requested HCR clause that stipulates to reduce F at low recruitment. It also seems that it was not possible to run simulations of HCRs with capped TACs, or else this reviewer does not understand why these simulations were not carried out. Instead, the WK ran simulations with fixed

TACs. The reviewer considers that software should be chosen that is capable of running the scenarios requested and recommends that in the future the chosen software should be suitable to run the requested scenarios.

With regards to the operating model the reviewer has the following comments.

- Uncertainty could have been considered (in the form of added stochasticity or through sensitivity analyses) with respect to weight-at-age, natural mortality, and maturity. Also the decision to set weight-at-age in the stock and in the catch as equal could have been evaluated. Nevertheless, the WK might not have had the resources for such extensive studies.
- From the document it was not entirely clear whether and how stochasticity was implemented with regards to recruitment.
- The assumption of a constant selection pattern based on the recent (2015-2017) average could have been evaluated. Given that the fishery changed substantially as recently as 2014, it is not clear to the reviewer why the assumption of no change should be valid. A sensitivity analysis with several selectivity patterns of the recent past (e.g. those that occurred during 1992-2014) would have been interesting. But again, the WK might not have had the resources for such extensive studies.

Despite the critical remarks in the bullet points above, the reviewer does not think they severely jeopardize the ability to draw conclusions from the simulations.

Regarding the observation model the reviewer considers that caution is needed when running simulations without full feedback. The WK should have argued why it was thought sufficient to include only assessment error and deal with assessment bias in the form of assuming overestimations of initial stock size by 25% and 50%. In a full-feedback simulation study, Kraak et al. (2008) have shown that assessment bias can emerge with changing sign in the course of the simulated time period (in their case shrinkage of F to the recent past caused cyclic under- and overestimates of F and appropriate management decisions were always lagging behind). The WK should have argued why phenomena like this are thought not to occur in this case. Moreover, why were the scenarios with bias (in the form of assuming overestimations of initial stock size by 25% and 50%) only run with the 10-year fixed TACs at 50 kt and 60 kt and not with the normal HCRs? **The reviewer recommends running the 12 HCRs with the two levels of overestimation of initial stock size. Without this, it is not clear whether the recommendation of using a cap at 50 kt is too conservative; after all, it is not known how the normal 12 HCRs, without cap, perform under these assumptions of overestimation – perhaps they perform fine and a cap is not necessary.**

It was not clear whether the consequences of overestimations of initial stock size by 25% and 50% were investigated by running simulations where the initial stock size was reduced to 80% and 67% respectively of the stock size in the default scenario (implying that the default scenario overestimated actual stock size by 25% and 50% respectively) or whether it was reduced to 75% and 50% of the default stock size. The reviewer does not agree with using the same reference points in the scenarios with overestimated stock size (the WK also acknowledges that the reference points would also be reduced if the present stock size is an overestimate, but for discussion they use the reference points already obtained – this seems to be a conservative choice).

A reality check was conducted, which the reviewer commends.

With regards to the presentation of the results, it is not entirely clear what is presented in the Table 1a (e.g. Y19, Y19-68). Furthermore, the reviewer considers that the following sentence does not accurately describe what is seen in the Table. "Trigger points of 800 kt (all three F values) and 600 kt (F=0.08 and 0.10) will result in high probability of SSB< Btrigger in any given year …".

Given that the WK concludes that they recommend using the HCRs with a cap of the TAC to 50 kt, it is a mystery to this reviewer why the WK did not run simulation scenarios with such a cap. The only explanation that comes to mind is that the software was not able to do this. **If, however, the software is capable of doing this, the reviewer suggests that the 12 HCRs (or a well-argued subset) are simulated (i) with and (ii) without a cap for the situations of (a) no bias, (b) 25% overestimation, and (c) 50% overestimation (resulting in 2 x 3 x 12 scenarios).** The reader is told that the catches in the short term, namely 3-5 years, were higher than the mean long-term catches. That implies that perhaps in year 6 the catches were already low, and this makes it illogical that a fixed high TAC of 10 years was considered to mimic a cap of the TAC if the software was not able to run proper capping. Perhaps a shorter period for the fixed TAC should be chosen.

The reviewer commends the taken approach to simulate the effect of recruitment failure, given that the software was not able to simulate the requested clause in the HCRs. The reviewer accepts the conclusion that at least for these rules (F19+=0.06 and 0.08 and $B_{trigger}$=450 kt) no extra precaution for incoming bad recruitment of the kind we have experienced, is needed.

Regarding the "factors not taken into account in the simulations" in section 1.10, the reviewer considers that the effects of these could have been explored in sensitivity analyses, although the WK might not have had the resources for such extensive studies.

Despite the above criticism of the scenarios with fixed TAC, if indeed the software is unable to run scenarios with a cap, the reviewer considers that the conclusions leading to the recommendation for using a cap may be correct although it cannot be excluded that the recommendation is too conservative (unless runs are carried out with the normal HCRs, without cap or fixed TAC, with assumptions of overestimation).

ToR c.

The WK does not explicitly answer this ToR. It recommends using a HCR with B_{pa}=450kt, without arguing why 450 kt is chosen rather than the new reference point of 315 kt, and F19+=0.06 or 0.08, with a cap on TAC of 50 kt.

Conclusion

The reviewer considers that the conclusions and recommendations can be accepted **provided that simulations of the 12 normal HCRs, without capped or fixed TAC, are carried out with the assumptions of 25% and 50% overestimation of the initial stock size**. Else it cannot be excluded that the WK's recommendation of capping is too conservative (perhaps capping is not necessary). If possible, the scenarios should also be run with proper capping (with each of the 12 HCRs and at each of the two levels of overestimation). The reviewer also recommends that at the next evaluation of the HCRs (in 2023 or earlier) the other criticisms are accommodated and software is chosen with which the actual requested scenarios can be run.

Reference

Kraak, S. B. M., Buisman, F. C., Dickey-Collas, M., Poos, J. J., Pastoors, M. A., Smit, J. G. P., Oostenbrugge, J. A. E. van, and Daan, N. 2008. The effect of management choices on the sustainability and economic performance of a mixed fishery: a simulation study. *ICES Journal of Marine Science*, 65, 697-712. https://doi.org/10.1093/icesjms/fsn045

Annex 4: Responses to Advice Drafting Group for *S.mentella*

Weight at age 19+

The change is essentially from using varying weight at age in the catch to using a method using both catch and survey data, which turned out to be fairly constant. The change in overall biomass results from using catch and survey data instead of simply catch data.

[From AFWG 2018, section 6.6.]

In earlier assessment, weight-at-age in the stock was set equal to the weight-at-age in the catch. This turned out to be problematic because of important fluctuations in reported weight-at-age in the catch that cannot be explained biologically (i.e. these are noisy data). In 2015, it was advised to either use a fixed weight-at-age for the 19+ group, or use a modelled weight-at-age based on catch and survey records (Planque 2015). The second option was chosen. Weight-at-age in the population was modelled for each year using mixed-effect models of a von Bertalanffy growth function (in weight). In 2018 an attempt was made to model weight-at-age for each cohort (rather than each year of observation). This showed that the growth function in nearly invariant between cohorts. As a result, it was decided to use a fixed (i.e. common to all years) weight-at-age as input to the Statistical Catch-at-age model. The observed and modelled weight-at-age are presented in Table 6.7 as well as Figures 6.5 and 6.6.

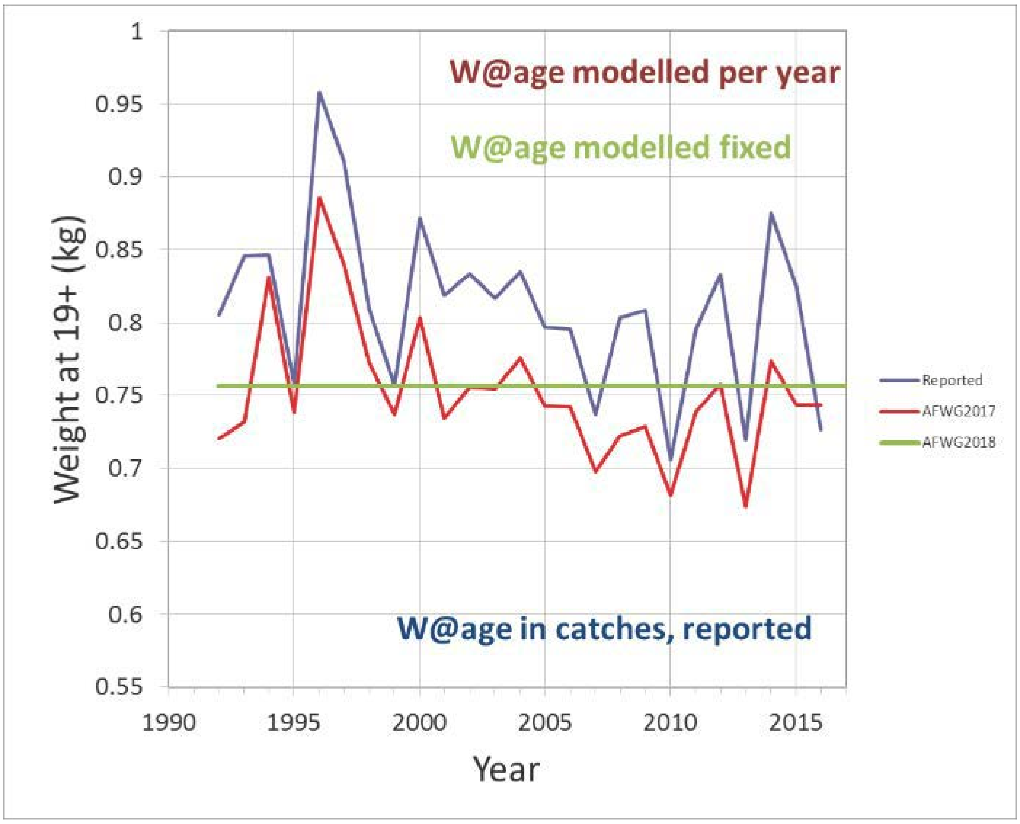


Figure 6.6. S. mentella in subareas 1 and 2. Weight-at-age 19+ as reported from catches (blue) or modelled from catches and survey observations (red) using a mixed effect model (Figure 6.5). The weights at age used in the assessment were based on the fixed effects model and are therefore the same for every year

Brtigger

Btrigger of 450kt and 600kt were tested against the recruitment failure scenario – and the 450kt mattered for F=0.08. Setting Btrigger of 315kt has not been evaluated against this scenario. It would be precautionary for F=0.06 (since this didn't hit the 450kt Btrigger, so it wouldn't hit the 315kt one either), but may or may not be precautionary for F=0.08.

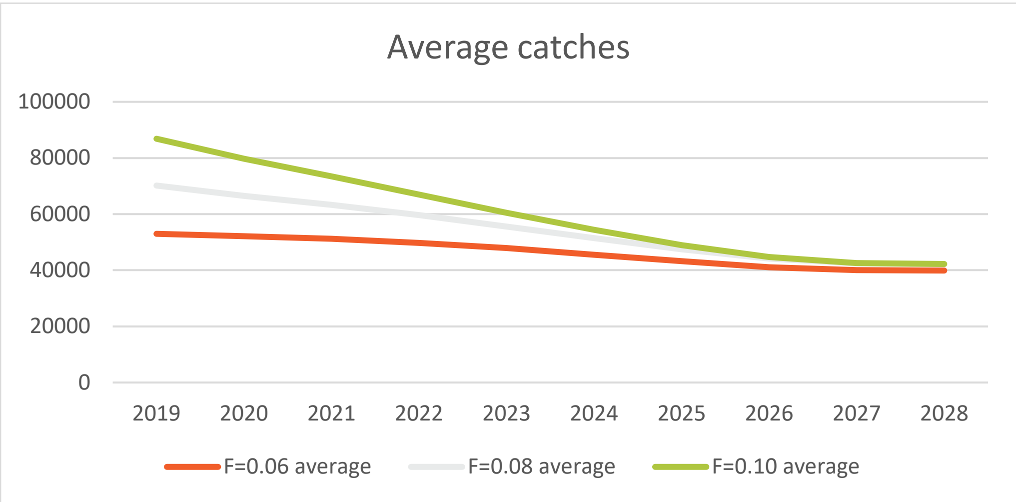
Bpa -> Blim multiplier

The sigma was available from the model, but was significantly less than the ICES standard 0.2, and expert judgement (i.e. Bjarte who was running the model) believed that the model estimate was artificially low and decided to use the ICES default.

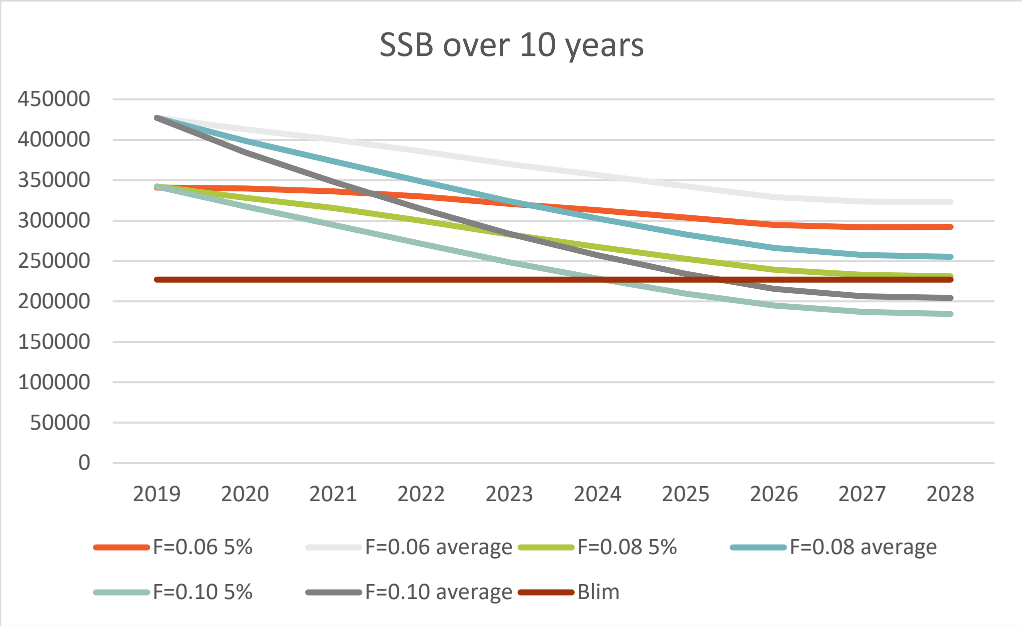
Cap on TAC is sufficient, but is it necessary?

Bjarte ran these by doubling the catch and halfing the Btrigger in the model – which approximates to overestimating the stock by a factor of two (i.e. having half the real stock we think we have). He kept the intermediate year catch as it was, which I think is wrong, so the SSB trend lines here should probably be shifted down by around 35kt at the start, somewhat less by the end.

Average catches: these are what the rule says, the model then applies double these in order to approximate having half the stock we think we have.



SSB trends. I think that these should all be shifted downwards c. 25kt to account for bias in stock size in the intermediate year, but that is not going to be run now



Conclusions:

The F=0.1 with no cap is not robust to a 50% bias in assessment

The F=0.08 goes just below Blim (with the adjustment mentioned above), so is "almost precautionary" to a 50% bias. Note that any higher bias and this would rapidly become unprecautionary, and there was nothing special about the 50% we chose to look at. Given the potential bias in assessment I would not say that F=0.08 and no cap could be certified as precautionary.

The F=0.06 is largely unffected by the cap. The initial catches are barely above 50kt in any case, and so this HCR is precautionary to this level of bias with and without the cap. It would also be precautionary to significantly higher levels of bias.

Suggestions:

1. We keep the cap. It does no harm at the F=0.06 level, and gives a level of precautionarity at the F=0.08 level.
2. We revise the cap text to say that we recommend a cap of 50kt if the chosen HCR gives catches of significantly higher than this level in the first years. The reccomended rule then becomes "F=0.06 with no cap" OR "F=0.08 with a 50kt cap".