ICES WKNPOUT Report 2018

ICES Advisory Committee

ICES CM 2018 / ACOM:38

Ref WGNSSK

Report of the Workshop for management strategy evaluation for Norway Pout (WKNPOUT)

26-28 February 2018

Copenhagen, Denmark

International Council for the Exploration of the Sea

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 for management strategy evaluation for Norway Pout (WKNPOUT). ICES WKNPOUT Report 2018 26-28 February 2018. ICES HQ, Copenhagen, Denmark. ICES CM 2018 / ACOM:38 Ref: WGNSSK. 96 pp.

DOI: http://doi.org/

Contents

Executive summary

The ICES Workshop for management strategy evaluation for Norway Pout (WKNPOUT) took place 26–28 February 2018 at ICES Headquarter chaired by Andrés Uriarte, Spain, with the assistance of ICES Secretariat. 12 participants, both scientific experts and stakeholders, from Denmark and Norway, attended the meeting. The group addressed the special request from the European Union and Norway to advise on the long-term management strategies of Norway Pout in ICES Subarea 4 (North Sea) and ICES Division 3.a (Skagerrak-Kattegat). The proposed management strategy is based on the ICES escapement strategy with the aim of achieving a high probability of having the minimum SSB required to produce MSY (B_{lim}) surviving to the following year. ICES was requested to evaluate:

1. Whether a management strategy is precautionary if the TAC is constrained with a lower bound in the range of 20 000 tonnes to 40 000 tonnes and an upper bound in the range of 150 000 tonnes to 250 000 tonnes, or another range suggested by ICES.

2. Whether such a strategy would be precautionary if the TAC constraints referred to in paragraph 1 are overridden by a constraint on the maximum value of fishing mortality (F_{cap}), and whether the application of the F_{cap} would allow a precautionary strategy with a higher minimum TAC than if the F_{cap} was not applied.

3. Whether a provision to override the minimum value of the TAC when the stock is forecast to be below some threshold value would allow a precautionary strategy with a higher minimum TAC than if the escape-clause was not included, and whether such a provision would provide any additional benefit to the inclusion of an F_{cap} as referred to in paragraph 2.

The alternative management procedures were tested in the framework of a management strategy evaluation (MSE) set up according to the assessment model SESAM adopted for Norway pout in the 2016 benchmark. One thousand simulations (replicates) were projected over 20 years for each of the different harvest control rules. Each replicate begins in the 2018 TAC year which starts in quarter 4 of calendar year 2017. Each replicate randomly draws a true state of the system (starting population, age and quarterly fishing patterns and series of past recruitments) from the joint distribution estimated by the last stock assessment. This is taken as the approach best reflecting the uncertainties in the SESAM assessment. An alternative reducing the uncertainty in the initial stock numbers, recruitment and exploitation pattern at the median estimate from the last assessment was also tested. The simulations were conditioned by a maximum realized level of fishing mortality the fishery can exert (assumed at 0.89; $F_{historical}$), which means that the full TAC will not be taken if the required F exceeds this value.

First the group tested whether the current ICES procedure for providing TAC advice for Norway Pout, based on an escapement strategy (the default method), was precautionary. Results showed that it is not precautionary (as tested with unconstrained levels of fishing mortality), because the probability of SSB falling below B_{lim} is higher than 5%. This is probably linked to cases of very high TAC and F when very high recruitments occur, in association with observation errors in the assessment. This called for modifying the default escapement strategy either by setting an upper F (F_{cap}) or including conditions on TACmin/ TACmax as explored here.

Concerning Request 1: The group tested HCR escapement strategies (as the default method) bounded by a combination of TACmin (at either 20, 30 or 40 kt) and TACmax (at 150 and 200 kt). Results show that these HCR were precautionary for TACmin at

20 kt for the two TACmax levels and for a TACmin at 30 kt when bounded by a TAC-max of 150 kt. They gave median and mean TACs (around 100–130 kt depending upon the rule) and realized catches around 110–115 kt, with TACs set at TACmin or at TAC-max around 20–24% and 36–46% of the cases respectively. In these cases, $F_{historical}$ was reached in 9–16% of the years, which makes the results sensitive to the assumption that the fishery will not exceed catches requiring F above $F_{historical}$. Other combinations based on higher values of TACmin or TACmax led to unprecautionary outcomes.

Concerning Request 2: The same combinations of TACmin and TACmax as for request 1 were explored with F_{cap} at either 0.3 or 0.4. Results showed that the inclusion of a F_{cap} increases the range of TACmin and TACmax combinations that are precautionary, reaching for F_{cap} up to a TACmin at 30 kt and a TACmax of 200 kt. On average, TACs become considerably lower when F_{cap} is applied (ranging between 72 and 97 kt depending upon the combinations) and realized catches did not exceed 92 kt. In general, TAC increases with increasing F_{cap}. The gain in average TAC by increasing TACmin or TAC-max is minimal, but the TAC constraints affect the probability of falling below B_{lim}. For these bounded rules, the probability of setting TAC at TACmin is very similar to the probability for HCRs without an F_{cap}, but the probability of reaching TACmax is considerably lower due to the application of the F_{cap}. The absolute changes in TAC between years are smaller with F_{cap} constraints as well, (in the order of 40 000 t) partly because of the lower TAC in general. Applying F_{cap} makes the HCR more robust to violations of the assumption of an $F_{historical}$, as the probability of reaching $F_{historical}$ becomes significantly lower than for the HCR without an F_{cap}.

The sensitivity of the performance of tested HCR to the alternative fittings of the stock recruitment relationship is minor, as shown for examples of rules with F_{cap}.

Concerning request 3, due to time limitation and little interest from stakeholders to override the TACmin, the group did not fully cover this request, but an exercise was made to find out if the TACmin of 40 000 t might become precautionary under an alternative configuration of the escapement policy. An escapement strategy with a TAC-min at 40 000 tonnes aiming at an escapement Biomass at 65 000 t instead of the current B_{lim} (39 450 t) would become precautionary with combinations of F_{cap} in the range 0.3 to 0.4 and TACmax in the range 150 to 200 kt. TACmin would be set in around 48% of the years, which gives a median TAC slightly above TACmin. Mean TAC for the three HCR is in the same order of size as for Request 2.

The Special request also asked ICES to evaluate whether the results of the MSE would be significantly changed if the TAC year were defined as 1 November to 31 October rather than a calendar year. The latter TAC year is applied to the EU Member States fishing in EU waters, while Norway uses the calendar year (January–December). Furthermore, ICES advice is produced based on a forecast from 1 October to 30 September, and ICES uses such forecast to advice management for the period 1 November- 31 October. The MSE adopted to answer the request follows the same practice. The WK has not compared the results of the MSE for the TAC year defined as 1 November to 31 October with those for a calendar year, as the latter would require a time shift in the assessment and forecast. The report includes some considerations on the current practice for advice and the TAC year.

1 Opening of the meeting

The ICES Workshop for management strategy evaluation for Norway Pout (WKNPOUT) took place 26–28 February 2018 at ICES Headquarter chaired by Andrés Uriarte, Spain, with the assistance of Sarah Millar and David Miller from ICES. 12 participants, both scientific experts and stakeholders, from Denmark and Norway, attended the meeting (see Annex 1). The group had to address the special request from the European Union and Norway to advise on the long-term management strategies of Norway Pout in ICES Subarea 4 (North Sea) and ICES Division 3.a (Skagerrak-Kattegat).

The meeting started on Monday 26 March at 09:30 hh at ICES Headquarters, by welcome of ICES and presentation of attendees; organising our agenda of work; reviewing the request and ToRs of the meeting and looking at the state of advances of the work foreseen to be carried out for the meeting (MSE software and some initial testing of rules).

The meeting had been prepared in advance by the group through a couple of Webex meetings which took place on Friday 9 and on Monday 19 February 2018 and by a lot of work of the modellers (from DTU Aqua) preparing the software for the Management Strategy Evaluation (MSE).

2 Adoption of the agenda

A draft agenda was circulated by the chair in advance to the group (see Annex 2) and it was tentatively adopted at the beginning of the group. The review and the discussion of the actual work and results moved the completion of some of the issues back in time in the agenda (see actual agenda in Annex 2 as well).

The Norway pout assessment uses the newly developed SESAM model, which is a stochastic state space model with quarterly time steps. To actually mimic the SESAM model and forecast, it was necessary to develop new software before and during the WKNPOUT meeting. The end result is considered as a well-tested software, however the process would have been more optimal if suitable MSE software was fully tested before the WK. As some improved configuration of the Operative Model were implemented during the WK meeting, much of the work actually ended up after the meeting, as well as much of the sensitivity analysis. It took an additional month after the meeting to complete the work foreseen by exchange of texts and draft results through the sharepoint, and be emails and Webex meetings (on the 21 and 23 March 2018).

3 Introduction

3.1 Terms of Reference: The request

On May 5th 2017, The European Union and Norway jointly requested ICES to advise on the management of Norway Pout in ICES Subarea IV (North Sea) and ICES Division IIIa (Skagerrak-Kattegat). The proposed management strategy is based on the ICES escapement strategy for Norway pout with the aim of achieving a high probability of having the minimum SSB required to produce MSY (Blim) surviving to the following year.

ICES is requested to evaluate:

1. Whether a management strategy is precautionary if the TAC is constrained with a lower bound in the range of 20,000 tonnes to 40,000 tonnes and an upper bound in the range of 150,000 tonnes to 250,000 tonnes, or another range suggested by ICES.

2. Whether such a strategy would be precautionary if the TAC constraints referred to in paragraph 1 are overridden by a constraint on the maximum value of fishing mortality (Fcap), and whether the application of the Fcap would allow a precautionary strategy with a higher minimum TAC than if the Fcap was not applied.

3. Whether a provision to override the minimum value of the TAC when the stock is forecast to be below some threshold value would allow a precautionary strategy with a higher minimum TAC than if the escape-clause was not included, and whether such a provision would provide any additional benefit to the inclusion of an Fcap as referred to in paragraph 2.

ICES is requested to indicate the results of the evaluation in a table that shows for the combination of parameter values selected for the evaluation:

• The average inter-annual TAC variation

• The average yield

• The average fishing mortality

• The average escapement biomass

• The probability that the stock falls below Blim in the year following the fishing year over a 20 year period.

ICES is additionally asked to indicate whether the results of the evaluation are significantly changed if the TAC year is defined as 1 November to 31 October rather than a calendar year.

To answer this request ACOM adopted the following resolution:

2018/2/ACOM38: The Workshop for management strategy evaluation for Norway Pout (WKNPout), chaired by Andrés Uriarte, Spain, will meet at the ICES Headquarter 26–28 February 2018 to (TORs):

1. Address the European Union and Norway joint request to ICES to advise on the management of Norway Pout in ICES Subarea 4 (North Sea) and ICES Division 3.a (Skagerrak-Kattegat). The proposed management strategy is based on the ICES escapement strategy for Norway pout with the aim of achieving a high probability of having the minimum SSB required to produce MSY (B_{lim}) surviving to the following year. (whereby the three requests listed above are required to abe addressed)...
2. Update reference points for the stock in light of the MSE results.

3.1.1 The fishery

A comprehensive description of the fisheries for Norway pout can be found in the Stock Annex for this stock:

(http://ices.dk/sites/pub/Publication%20Reports/Stock%20Annexes/2017/nop.27.3a4_SA.pdf).

3.1.2 Reference points

The reference points for Norway pout (Table 1.1.1) were estimated at the Benchmark meeting, WKPOUT in 2016 (ICES, 2016a).

As described in the benchmark report (ICES, 2016a), the WKPOUT benchmark group found that the stock–recruitment plot of Norway pout (Figure 1.1.1) equaled a type 5 Stock–recruitment. A type 5 stock is defined as "Stocks with no evidence that recruitment has been impaired or with no clear relation between stock and recruitment (no apparent S–R signal)." This implies that $B_{lim} = B_{loss}$, and hence the new B_{lim} value was set equal to B_{loss} taken from the 2016 benchmark assessment. According to the assessment result, this B_{lim} corresponds to 39 450 t (B_{loss} = SSB value in quarter 4, 2005).

Norway pout spawn in quarter 1, however the SSB reference point refers to the beginning of quarter 4. This timing was chosen to fit the escapement strategy (which aims a leaving a minimum biomass in the sea after the TAC has been caught) in combination with the (EU) TAC year from 1 November to the 31 October.

SSB estimated in quarter 4 was chosen as biomass reference point because this is closest to the beginning of the TAC year (1 November). By this timing the probability of SSB being below B_{lim} can then be evaluated immediately after the fishing season for which a TAC is being calculated.

ICES considers that the forecast period quarter 4 to quarter 3 in the following year in the assessment (October–September) can be used directly for management purposes to set the TAC for the period 1 November to the 31 October (ICES advice for Norway pout, 2017).

The short-term forecast is stochastic and allows the probability of SSB being below B_{lim} to be evaluated directly. By such an approach there is no need for a MSY $B_{escapement}$ and the value is not defined for this stock.

There is no stock recruitment model for scenario recruitment for SSB below B_{lim}, as there is no apparent stock-recruitment relation and as B_{lim} is chosen at B_{loss}. For the simulation of recruitments in the MSE it was decided to use a "Hockey-stick" SR-model with inflection point at the lowest observed SSB in quarter 1. That point corresponds to SSB (72 101 tonnes) in quarter 1 of 2005, the same year with the lowest SSB (B_{loss}) in quarter 4. The Hockey-stick model was only applied for SSB below B_{loss} (see section 4.2.2.6). In all simulations the risks have been assessed on the risk of SSB(q4) of falling below B_{lim} of 39 450 t (B_{loss} = SSB value in quarter 4, 2005), but for comparative purposes, the risk of SSB(q1) of falling below B_{lim} of 72 101 tonnes (B_{loss} in quarter 1 of 2005) has also been assessed.

The reference points are based on the benchmark assessment in 2016. The newest assessment gives a slightly different estimate of SSB (Figure 1.1.2). SSB in Q4 of 2005 (B_{loss}) is now estimated somewhat lower than in the benchmark assessment (*30 742 tonnes in the 2017 stock assessment vs 39 447 tonnes in the benchmark assessment*), while SSB in Q1 of 2005 (Hockey-stick inflection point) is now estimated quite close to the previous estimate. For the MSE, it was decided to maintain the reference points from the benchmark.

Regarding TOR 2 of WKNPOUT asking to update reference points for the stock in light of the MSE results, WKNPOUT has not considered necessary to update the reference points for Norway Pout agreed in the last benchmark. The scatter plot of stock and recruits was very similar in last assessment compared to the benchmark (ICES, 2017). And in spite of having found that the fitting of a SR relationship (with Eqsim) mostly selects a hockey stick with an inflection point well above the current B_{lim} of 72 101 tonnes (B_{loss} in quarter 1 of 2005, as adopted in the benchmark), the approach followed for the MSE has been that of testing the sensitivity of results on the probability of falling below current B_{lim} to the adoption of alternative SR relationships (according to the likelihood of their fitting) (see sections 4.5.2 and 5.5.2). The fitted Eqsim S-R relation was used to generate recruitments in a sensitivity analysis. This analysis used however the same B_{lim} (72 101 tonnes) in the evaluation of the probability of SSB below B_{lim}. The WK does not consider that there is a strong justification for changing the decision made by the benchmark of classifying the S-R as a “type 5 S-R” (no stock recruitment relation). Therefore, the biomass reference points remained unchanged.

Table 1.1.1 Norway pout in Subarea 4 and Division 3.a. Reference points, values, and their technical basis.

Framework	Reference point	Value	Technical basis	Source
MSY approach	MSY $B_{escapment}$	Not defined		
	F_{MSY}	Not defined		
	F_{cap}	Not defined		
Precautionary approach	B_{lim}	39 450 t, 4th quarter	$B_{lim} = B_{loss}$, the lowest observed biomass in 2005.	ICES (2016a)
	B_{pa}	65 000 t, 4th quarter	$= B_{lim}\ e^{0.3 \times 1.645}$	ICES (2016a)
	F_{lim}	Not defined		
	F_{pa}	Not defined		
Management plan	SSB_{MGT}	Not applicable		
	F_{MGT}	Not applicable		

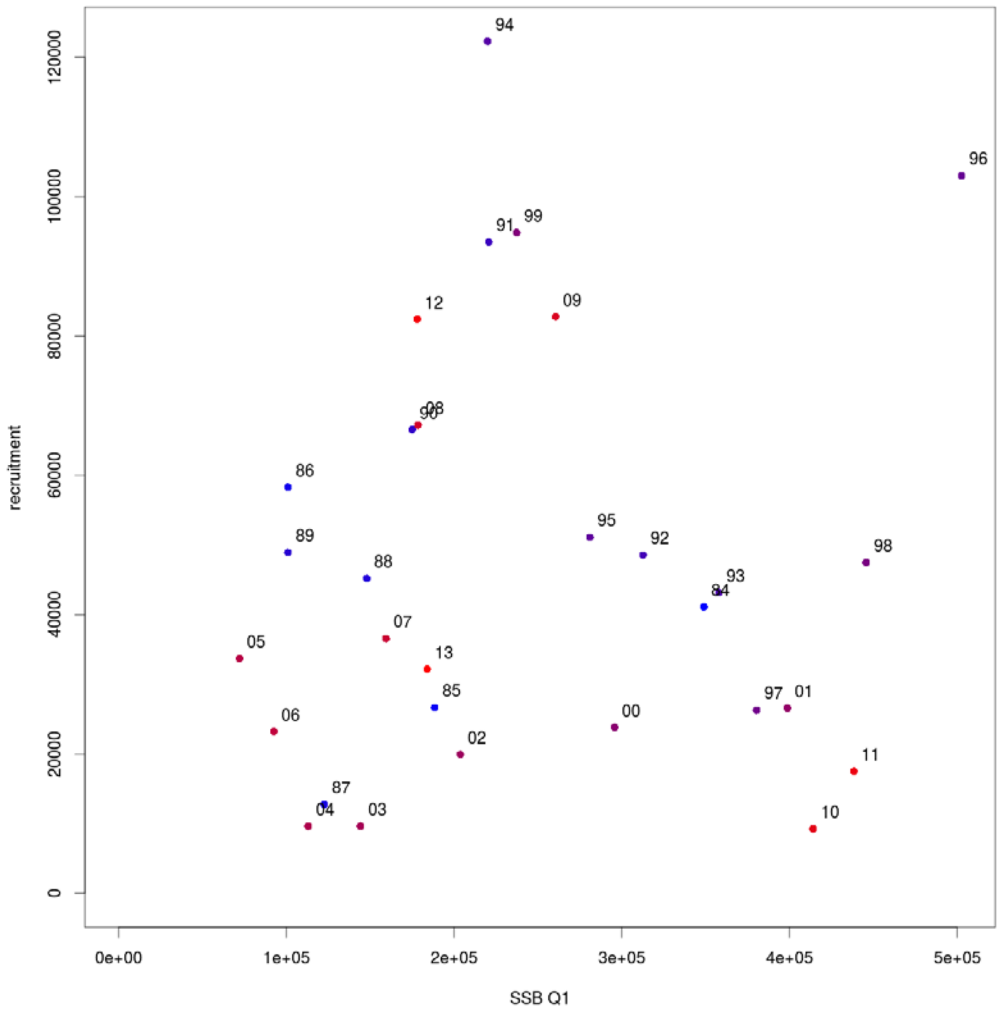


Figure 1.1.1. Stock–recruitment from SESAM. SSB in quarter 1. The corresponding plot with SSB in quarter 4 is similar. Colours are associated with the year (blue for earliest to red for most recent). The figure and figure caption are copied from the Benchmark meeting report (ICES, 2016a).

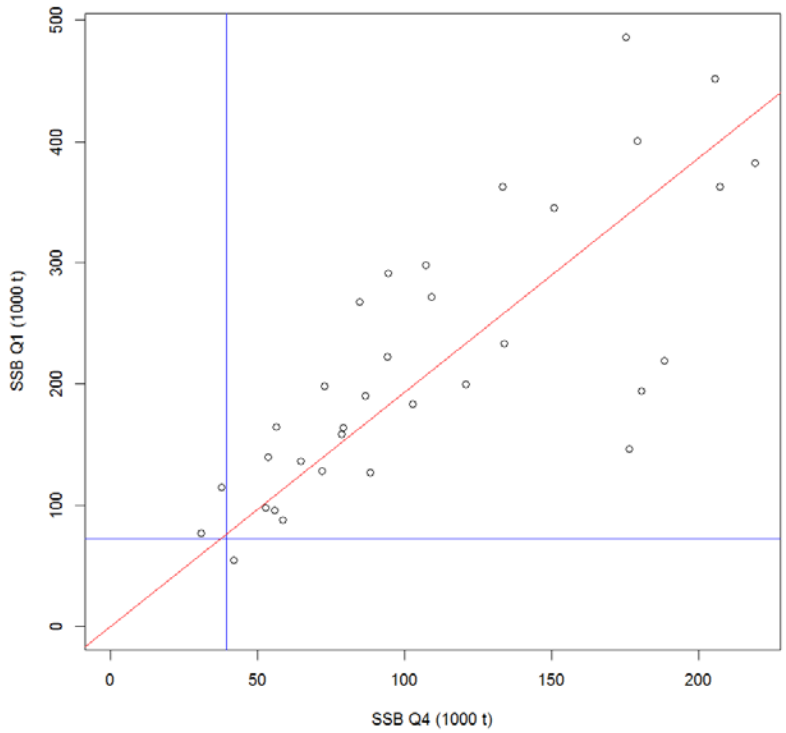


Figure 1.1.2. Relation between SSB in quarter 4 and SSB in quarter 1 with data from the 2017 SESAM assessment. The vertical blue line corresponds to B_{loss} (SSB in Quarter 4 of 2005 from the benchmark assessment, 39 450 t), and the horizontal line corresponds to the benchmark estimate of SSB in quarter 1 of 2005 (72 101 t). The red line shows the linear relation between the two measures of SSB.

3.2 Interpretation of the request (Approach taken by WKNPOUT)

ICES sent a letter to the EC officer on June 29 2017 with comments on the request to clarify the understanding and procedure to follow (Annex 3).

ICES understood the relevance of inviting stakeholders to actively participate in the process (ICES workshop), such that the potentially many options included in the proposal can be reduced concurrently in agreement with stakeholders and their preference can be duly taken into account to customize as much as possible the Harvest Control Rules (HCR) being tested to reply the request.

According to the letter with comments to the request and with the opinions of the scientific experts and stakeholders attending the meeting, the following interpretations and decisions were made to address the different points of the request:

Point 1 in the request:

This aims at evaluating "*Whether a management strategy is precautionary if the TAC is constrained with a lower bound in the range of 20,000 tonnes to 40,000 tonnes and an upper bound in the range of 150,000 tonnes to 250,000 tonnes, or another range suggested by ICES.*"

The group has applied the default escapement strategy this stock, to estimate SSB, yield and risk to B_{lim} (and other performance indicators) as function of combinations of maximum TAC in the range 150–250 kt (step 50 kt) and minimum TAC in the range 20–40 kt (step 10 kt). It is understood that when catch associated to the escapement strategy is lower or higher than the TACmin or TACmax values then these two bound levels of the TAC prevail and are adopted for management.

The 250 000 t as a maximum bound limit was finally omitted as this level has not been reached by the fishery for the last 25 years. Instead, maximum TAC in the range 150–200 kt were investigated.

Point 2 in the request:

This aims at evaluating "*Whether such a strategy would be precautionary if the TAC constraints referred to in paragraph 1 are overridden by a constraint on the maximum value of fishing mortality (Fcap), and whether the application of the Fcap would allow a precautionary strategy with a higher minimum TAC than if the Fcap was not applied.*"

The WG understood that the F_{cap} approach would override TAC constraints only for the upper bound of the TAC ranges (i.e. of the TACmax, for instance whenever the escapement strategy allows TACmax but the F_{cap} leads to a lower TAC, but not the reverse, if the F_{cap} applied to restrict the escapement strategy still allows a TAC above TACmax then TACmax would prevail), whereas the lower bound (TACmin) would be kept invariant whatever the actual level of the stock is.

For the options on TACmin and TACmax analyzed in point 1 the WK would estimate the SSB, yield and risk to B_{lim} as function of F_{cap} in the range 0.3–0.4(step 0.1). This range of F covers the upper mean range of estimated yearly Fs over the last 20 years and hence would be consistent with the normal fishing activity of the fleet over last decades.

Point 3 in the request:

This aims at evaluating "*Whether a provision to override the minimum value of the TAC when the stock is forecast to be below some threshold value would allow a precautionary strategy with a higher minimum TAC than if the escape-clause was not included, and whether such a provision would provide any additional benefit to the inclusion of an Fcap as referred to in paragraph 2.*"

WKNPOUT assumes that the stock threshold value referred to in request 3 will correspond with a spawning biomass level surviving at the end of the management year (SSB(q4)).

This request is not sufficiently clear, and several interpretations and alternative considerations were *a priori* considered: First the WK understood that this request should apply to HCR cases resulting non-precautionary (i.e. with risks of falling below B_{lim} above 5%). In those cases on the one hand, HCRs making the minimum TAC not applicable for a range of very low stock sizes, below some ad hoc defined threshold levels of biomass, were briefly discussed but not tested; On the other hand one alternative HCR allowing high TACmin (as suggested in this request) was devised and tested conditioned to be applied on escapement strategies seeking for a higher escapement biomasses.

This request was not sufficiently covered by the group, essentially for two reasons: **a)** Lack of time, as sufficient exploration of other HCR to answer request 1 and 2 left no time during the meeting to address this issue, and the work finally carried out was made entirely afterwards and **b)** a sufficient amount of HCR explored for the requests 1 & 2 were precautionary and in addition stakeholders at the meeting have a strong preference for adopting TACmin values in no case being overridden by any escape clause when the stock is forecast to be below some threshold value. This preference for TACmin prevailing at whatever low stock level made this request to become of low priority in comparison with the sensitivity analysis being carried out for the other HCRs. Some testing of an alternative HCR allowing a high TACmin (of 40 000 t) was tested and proved precautionary in section 3.4. The group acknowledge that this request was not sufficiently addressed.

The request on the Management Calendar year

ICES was additionally asked to indicate whether the results of the evaluation are significantly changed if the TAC year is defined as 1 November to 31 October rather than a calendar year.

This request was a bit confused, because currently management follows an intermediate calendar between the calendar year and 1 November–31 October calendar, as the TAC for the EU Member States fishing in EU waters applies from 1 November to 31 October, while for Norway is the calendar year (January–December). Furthermore, the ICES advice is produced based on a forecast from 1 October to 30 September.

The SESAM model and reference points are reviewed in a configuration with an assessment that does not follow the calendar year but includes quarter 4 of year y and quarters 1–3 of year y+1. The advice given by ICES, for the period 1 November to 31 October, follows this timing of the assessment. It is requested to make the MSE for an advice following the calendar year, and to evaluate if the results would have been significantly changed if the TAC year is defined as 1 November to 31 October. Such a request cannot be fulfilled, as it would require a new review of an assessment and reference points based on an assessment following the calendar year. Therefore there was not time to explore alternative setting of the forecast procedure to better coupling a natural year management calendar. Therefore the WG based its entire analysis on the current procedure followed by the ICES assessment WG, i.e. basing the HCRs on a forecast running from 1 October to 30 September for a TAC applicable over this period, even though it is admitted that it will be implemented from 1 November–31 October, in the assumption that this time shift does not cause a major deterioration in the quality of the advice. Therefore no particular answer was provided to this additional request. It should be understood that the entire MSE of the HCRs is conceived for a TAC running from 1 November–31 October.

The potential problems arising from the time shift between the forecast supporting the advice (from 1 October to 30 September) and the calendar of the management is preliminarily explored in section 4.5.1 of the report.

About TOR 2 of WKNPOUT asking to update reference points for the stock

In light of the MSE results, WKNPOUT has not considered necessary to update the reference points for Norway Pout agreed in the last benchmark (see section 2.1.1 and sections 4.5.2 and 5.5.2).

The default escapement strategy in use for this stock, unbounded by any TAC level or F_{cap} is not considered precautionary. Precautionary scenarios, applying a maximum value of fishing mortality (F_{cap}), is presented in Annex 4.

4 Methods

4.1 Overview

4.1.1 Management Strategy Evaluation conceptual overview

The MSE model projects the age structured population forward in the operating model, TAC year by TAC year, accounting for management advice (i.e. setting the TAC based on estimates from the assessment), fishing mortality, natural mortality, and recruitment (Figure 4.1). The true stock numbers on the log scale (logN) and the true exploitation pattern on the log scale (logE) in a given year of a given replicate represent the state of the system at that time and are referred to as the "true state". The true state has one value of logN and one value of logE per age per quarter of every year and replicate (later we assume that the true logE is constant across TAC years). "Estimated state" also refers to the state of the system in a given year and replicate; it has a median logN and logE for each age and quarter, but with uncertainty around those estimates, represented by a multivariate normal distribution.

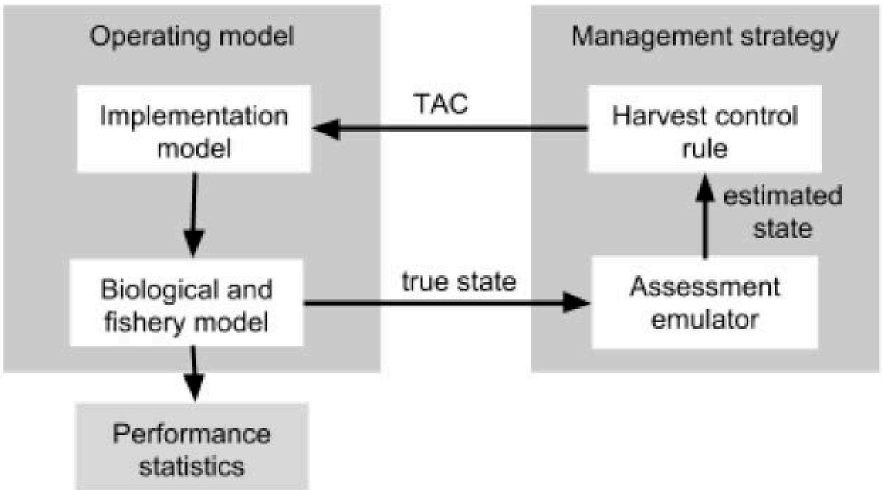


Figure 4.1. Conceptual overview of the management strategy evaluation modelling process (adapted from Punt *et al.*, 2014).

4.1.2 Seasonal and age structured model

The models used in this MSE are structured by quarters and by age as done in the SESAM model. Age groups are 0, 1, 2, and 3+. As the TAC year is shifted by a quarter from the calendar year, the quarters of the TAC year are quarters 4, 1, 2, and 3 of the calendar year (abbreviated q4, q1, q2, q3 henceforth). The escaped SSB is calculated in q4 after the TAC year. The state variables N (stock numbers) and E (exploitation pattern), are both structured according to season and age. E refers to the exploitation pattern before it is multiplied by an F multiplier to get the actual fishing mortality in an individual year.

4.1.3 Starting state

This MSE begins in the 2018 TAC year. Each replicate of the MSE randomly draws a true state of the system (logN and logE) from the joint distribution estimated by the last stock assessment (ICES, 2017). The random draw of the true state of logN at the end of the 2017 TAC year (q3 of the calendar year) is projected forward into the beginning of the 2018 TAC year (q4 of the 2017 calendar year) to start the time series for each replicate. Then, E is scaled so that any F multiplier will be equivalent to the mean fishing mortality for ages 1 and 2 (F_{bar}). We assume that this true E is constant across years within a replicate of the MSE, but varies among replicates.

4.2 Operating model

4.2.1 Biological and fishery model

Given a TAC estimated from the management strategy, the operating model simulates the fishery and the biological dynamics affecting the population. The TAC is taken from the true population. As the TAC is taken, the dynamics of the true population are simultaneously simulated quarter by quarter. Surviving N from q3 of the past TAC year are brought into q4 of this TAC year by applying natural mortality, fishing mortality from the last TAC year, and biological noise (i.e. process error). Process error was estimated by SESAM in the assessment and implemented as mortality of the stock numbers. Given, N(a,t), the number of fish of age a in quarter t, the mortality and process error are implemented as N(a,t+1)=N(a,t) * exp(-(F(a,t) + M(a)) + epsilon(a,t)), where epsilon(a,t) is a random draw from a normal distribution with mean 0 and standard deviation 0.1932362, as estimated by SESAM. A different random draw of process noise is applied to different age groups in a quarter. In q4, survivors are calculated by applying fishing mortality from this TAC year along with natural mortality and a new random draw of process noise. Then, the survivors from q4 increase in age as they move to q1. Survivors from q1 move to q2 and then from q2 to q3 by applying fishing mortality, natural mortality, and new random draws of process noise. Recruitment occurs in q3 based on the SSB from q1 (see details in recruitment section below). This concludes the end of the TAC year, so that the next TAC year can begin with a new assessment.

4.2.2 Conditioning (input variables)

The input variables used in the biological and fishery model are the same as those used in the forecast.

4.2.2.1 Initial population

The number of individuals at the end of the last TAC year (q3) as estimated by the SESAM assessment (ICES, 2017) are in Table 4.1.

Table 4.1 Median numbers of individuals (millions) in each age class at the end of the 2017 TAC year.

	q3
Age 0	37204.522
Age 1	23478.134
Age 2	1680.344
Age 3+	591.907

4.2.2.2 Natural mortality

The MSE assumes that natural mortality is constant.

Table 4.2 Natural mortality rates

	q1	q2	q3	q4
Age 0	0.29	0.29	0.29	0.29
Age 1	0.29	0.29	0.29	0.29
Age 2	0.39	0.39	0.39	0.39
Age 3+	0.44	0.44	0.44	0.44

4.2.2.3 Mean weights at age

The MSE assumes that stock and catch weights are constant. Catch weights were forecast by SESAM function "CWpredict" using the following generalized additive model (ICES WKPOUT Report, 2016):

$E((CW_{a,q,t})^{.5}) = \mu_{a,q} + s(\text{cohort}, a) + U_t$

where $\mu_{a,q}$ is a mean for each combination of quarter and age, s() is tensor product smoothing spline, and U_t are normal distributed random effects. The square root transform is used to achieve variance homogeneity in the residuals.

Table 4.3 Stock weights (grams)

	q1	q2	q3	q4
Age 0	0.000	0.000	4.000	6.000
Age 1	9.000	12.000	25.000	25.000
Age 2	25.000	25.000	40.000	40.000
Age 3+	40.000	50.000	60.000	58.000

Table 4.4 Catch weights (grams)

	q1	q2	q3	q4
Age 0	8.450772	12.258520	9.548663	10.33685
Age 1	10.041707	14.16063	25.421595	28.23879
Age 2	21.032165	23.52772	31.009362	36.37819
Age 3+	29.482098	29.47802	34.866770	41.92795

4.2.2.4 Proportion mature

The MSE assumes that the proportion of fish mature is constant.

Table 4.5 Proportion of fish mature

	q1	q2	q3	q4
Age 0	0.000	0.000	0.000	0.000
Age 1	0.200	0.200	0.200	0.200
Age 2	1.000	1.000	1.000	1.000
Age 3+	1.000	1.000	1.000	1.000

4.2.2.5 Fishing mortality

The exploitation pattern for each replicate of the simulation is drawn from a lognormal distribution with medians in Table 4.6. After a true exploitation pattern is drawn for a replicate at the beginning of the simulation, it stays constant throughout the 20 years of simulation. Thus, it does not follow the random walk as it does in the SESAM assessment model.

Table 4.6 Median exploitation pattern (and 95% CI) estimated from the last assessment, corresponding to annual F_{bar}(1–2) of 1.

	q1	q2	q3	q4
Age 0	0.000 (0, 0)	0.000 (0, 0)	0.000 (0, 0)	0.020 (0.004, 0,078)
Age 1	0.003 (0.001, 0.009)	0.042 (0.013, 0.102)	0.099 (0.031, 0.256)	0.558 (0.317, 0.816)
Age 2	0.008 (0.002, 0.02)	0.117 (0.035, 0.276)	0.271 (0.085, 0.61)	0.902 (0.513, 1.23)
Age 3+	0.002 (0, 0.012)	0.051 (0.01, 0.21)	0.048 (0.009, 0.204)	0.020 (0.005, 0.062)

4.2.2.6 Recruitment

If SSB in q1 is above 72 101 tonnes (the assumed break point in a hockey stick model), then recruitment is a single random sample from the estimated recruitment from previous years. From the last SESAM assessment, the median estimates of past recruitment are (39327.660, 25995.035, 51421.822, 10650.346, 43493.304, 44963.595, 56956.836, 93069.810, 49615.290, 42409.190, 125023.123, 48916.188, 102916.248, 21844.957, 38451.252, 89932.999, 23589.333, 24010.888, 18658.985, 8078.843, 7729.149, 29561.302, 20929.877, 31002.677, 50187.333, 71259.853, 6803.348, 11409.491, 53020.413, 14139.579, 85274.074, 34096.557, 79343.026, 37204.457). These values have uncertainty in the estimates from SESAM. The uncertainty is estimated as a multivariate normal distribution, so that correlations among estimates of logR, logN, and logE are accounted for. Each replicate of the MSE differs slightly in the set of recruitment estimates that are resampled from. This step occurs at the same time that the initial true state is sampled as described in section 3.1.3.

If SSB in q1 is below 72 101 tonnes, then recruitment is sampled from a line going from the origin (0,0) to (72 101, 43 861.38). The value 43 861.38 is the mean of past recruitment estimates. The line has slope a = 43 861.38/72 101. The sample has lognormal truncated error. A random sample, z, from the standard normal N(0,1) is truncated to be between -2 and 1.5. This truncation is based on the distribution of historical standard normal deviations from the mean log recruitment. The formula for recruitment is R = a*SSB*exp(z*sdlogR), where sdlogR is the standard deviation of the log of estimated past recruitment, 0.770491. The new recruits are log transformed and put into the true logN structure in age 0, q3.

4.2.3 Implementation model

The TAC is taken from the true population, but only up to an F_{bar} equal to $F_{historical}$. This is a form of implementation error conveying the maximum effort that the fleet has historically implemented. As Norway pout is not a schooling species, increasing effort is required to implement a larger F_{bar} and this effort is limited by the fleet. We assume that $F_{historical}$ cannot exceed 0.89, which is the 97.5% percentile of F estimated by SESAM in the last 20 years. This value is the upper confidence interval of the 2013 F_{bar}, calculated in the 2017 assessment.

4.3 Management strategy

4.3.1 Assessment emulator

For each replicate, in each year of the MSE, a stock assessment is emulated to produce an estimated state. The assessment emulator adds observation error to the true state to get an estimated state. In this MSE, the assessment emulator randomly draws an estimated state (log N and log E) from a multivariate normal distribution which has the true state as the mean and a variance-covariance matrix as estimated by SESAM (Figure 4.2) (ICES, 2017). Moving from the log scale to the natural scale, this implies that estimated N and E are log-normally distributed with the true N and E as the median. Then, the estimated E is scaled so that any F multiplier is the estimated F_{bar}.

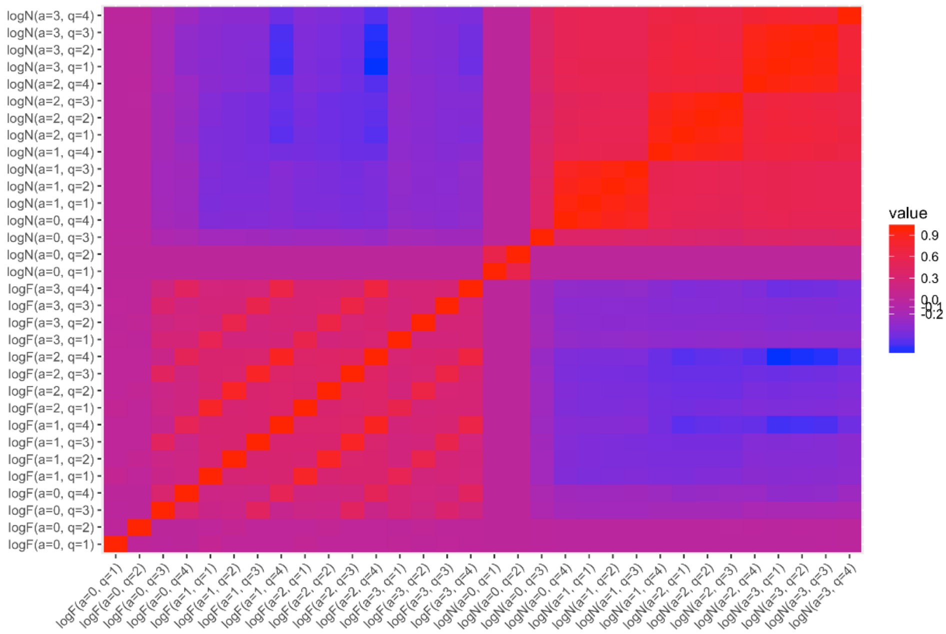


Figure 4.2 Correlation matrix of logN and logF estimated by SEASAM.

A real stock assessment would be done with SESAM, a seasonal extension of a state-space assessment model named SAM (Nielsen and Berg, 2014). As a state-space model, SAM assumes that the true underlying state is observed with some error. The observations are generated by the standard catch and survey equations. Stock numbers N follow the standard dynamics, but with the addition of process error. The formula for updating the stock number is the same as in our operating model, N(a,t+1) = N(a,t) * exp(-(F(a,t) + M(a)) + epsilon(a,t)), where epsilon(a,t) is normally distributed with mean zero. SAM assumes that the exploitation pattern (i.e. selectivity) follows a random walk through time.

4.3.2 Stochastic forecast

A stochastic forecast is performed on the estimated state to suggest a TAC. The goal of a forecast in an escapement strategy is to find the TAC that gives 5% probability of the escaped SSB (i.e. the surviving SSB after the TAC is taken) being below B_{lim}. A Norway pout stochastic forecast (as described in ICES WKPOUT Report 2016 and used in this MSE) begins by drawing K (K = 1000 in this MSE) values of the true state of logN, logE, and logR from their estimated joint distribution - a multivariate normal distribution which is estimated by the SESAM assessment model. Then, the exponential of the simulated values is taken, to put the set on the natural scale (N, E, and R). The Norway pout stochastic forecast applies the same F multiplier to all simulated sets of N, E, and R and then calculates the escaped SSB for each set using the standard stock equation with randomly sampled process errors. It adjusts the F multiplier until 5% of the escaped SSBs are below B_{lim}. Each simulated set produces a different TAC because they are using the same F multiplier. When the 5% target is reached, the forecast returns the median TAC produced when the F multiplier is applied to the simulated sets. By performing the forecast in this way, it might not precisely achieve the goal of finding the single TAC that gives 5% probability of escaped SSB being below B_{lim}. That precise goal requires searching over values of the TAC and examining the resulting distribution of escaped SSB. For each simulated set of N, E, and R and an examined TAC, a search would be performed to find the F multiplier to give the TAC that is being examined. Thus, each simulated set would have a unique F multiplier for an examined TAC. Then, the escaped SSB would be calculated for each simulated set. The examined TAC would be adjusted up or down until 5% of the simulated sets give escaped SSB below B_{lim}. Therefore, a typical forecast algorithm would have an outer search over TACs and nested searches over F multipliers for each simulated set, but the Norway pout forecast is faster because it only searches over F multipliers. Any potential bias that could be introduced by taking this shortcut is not intuitively obvious.

In this seasonal model with four quarters of the TAC year, calculating the escaped SSB for a given set of N, E, and F multiplier follows the same quarter by quarter population dynamics as described above in the operating model, starting with q3 of the previous TAC year. However, the forecast projects the population one quarter further, to predict the escaped SSB at the beginning of q4 of the following TAC year because the distribution of escaped SSB in q4 is what is examined to assess the 5% risk in the forecast. This is generally how a basic escapement strategy sets a TAC, by choosing the TAC that gives a 5% risk of the escaped SSB being below B_{lim}. In Norway pout, the escaped SSB is defined as those in q4 after the TAC year.

The suggested TAC from the stochastic forecast can be modified to account for more complex HCRs, as described below in the Harvest Control Rules section. The modified TAC from the HCR will be applied to the true population, potentially with implementation error, as described above in the Implementation Model section.

4.3.3 Harvest Control rules

4.3.3.1 Unbounded escapement strategy

In the unbounded escapement strategy, the TAC suggested by the stochastic forecast is used.

4.3.3.2 Bounding the escapement strategy with TACmax & TACmin

To bound the escapement strategy with TACmax and TACmin, the TAC suggested by the stochastic forecast is compared to the TACmax and TACmin. If it is above TACmax, then the final TAC is TACmax. If it is below TACmin, then the final TAC is TACmin. The MSE considered TACmax values of 100 000, 150 000, and 200 000 tonnes. The MSE considered TACmin values of 20 000, 30 000, and 40 000 tonnes.

4.3.3.3 Bounding the escapement strategy with TACmax & TACmin & F_{cap}

To bound the escapement strategy by F_{cap}, the stochastic forecast is performed in a way that it only searches for F_{bar} values up to F_{cap}. Then the TAC suggested by the stochastic forecast can subsequently be bounded by TACmax and TACmin. The MSE considered F_{cap} values of 0.3 and 0.4. An F_{cap} of 2.0 is considered to have negligible effect and is used in the software to keep the optimizer in a well-behaved region.

4.3.3.4 Adding safe guard clauses at low SSB values (TOR3)

As explained in section 3.5.2 request 3 aimed at searching for alternative HCR that would allow high TACmins be precautionary thanks to the inclusion of some "provision to override the minimum value of the TAC when the stock is forecast to be below some threshold value".

Some potential HCR trying to address this issue were initially considered by modifying the general HCRs outlined before, as follows:

a) Adding a clause such that in case of expecting the biomass, falling below $B_{lim}/2$ at the end of the management year then the TAC would be set to 0
b) Or for the cases with F_{cap} allowing the F_{cap} to override the TACmin whenever perceived required F for TACmin would exceed F_{cap}
c) Or modifying the general HCR by setting a higher escapement SSB(q4) target than the current B_{lim}, as for instance to Bpa (65 000t), to test if such modification would make a TACmin as high as 40 000 t precautionary

However only option ***c*** above was covered, the others were not addressed simply because of lack of time and the low priority given to them (see section 3.2 for the justification). The methodology for option ***c*** is the same as described above for the HCR covered to answer requests 1 and 2, except for adopting as the escapement target the biomass at Bpa (65 000t) in quarter 4.

4.4 Replication in simulations

For each HCR tested, we ran 1000 replicate simulations forward for 20 years. Throughout the simulations, different random numbers are generated for process error, recruitment, and observation error for each replicate, TAC year, and quarter as applicable. Random numbers drawn for initial conditions (N, E, and R) are unique across replicates except as described in the sensitivity test described below. The same random seeds were used across HCRs to ensure that differences in the outcomes were due to differences in the HCRs rather than in the random numbers. Any potential noise from small sample sizes should disappear as replication goes to infinity, but some may still be present with 1000 replicates.

4.5 Sensitivity tests

4.5.1 Ignoring uncertainty in initial N, E, and R

The models described above assume that the last assessment had uncertainty in the exploitation pattern, the past recruitment, and in the stock numbers in q3 of 2017; they use this uncertainty to randomly draw the exploitation pattern, recruitment, and starting stock numbers in the operating model. To test the sensitivity to this assumption, we instead set the true initial stock numbers, past recruitment events, and exploitation pattern equal to the median estimates from the last SESAM assessment (ICES, 2017). We did this for 1000 replicates. The replicates began with the same initial stock numbers in q3 of 2017 at the median estimate, but diverged in q4 of the 2018 TAC year due to process error affecting survival. As in all the MSE simulations, the replicates differed from each other at each time point in their random draws of observation error in the assessment emulator, process error, and recruitment. As in all the simulations, the true exploitation pattern was held constant across years, so the replicates in this set of simulations always had the same true exploitation pattern as each other, but different estimated exploitation patterns. See figure 4.3 for more explanation of stock recruitment in this sensitivity test. We refer to this sensitivity test as the "one initial population version", but it has these differences in E and R as well.

We applied this operating model to all of the HCRs described above, including all combinations of the TACmin, TACmax, and F_{cap} described in Table 5.1, except scenarios 18 to 21.

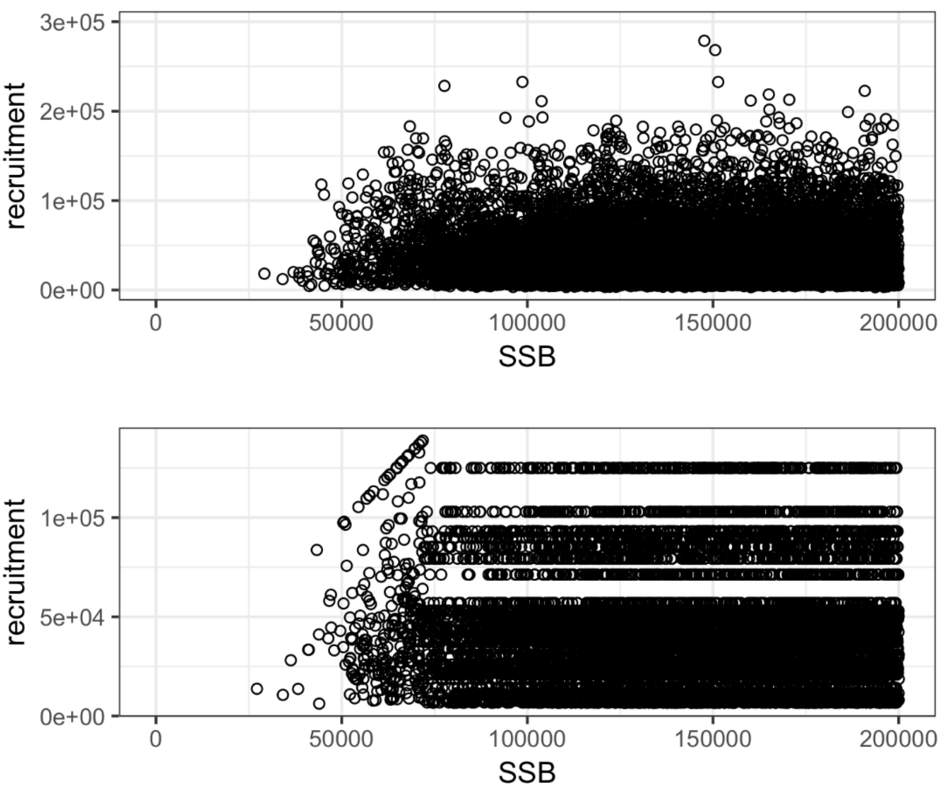


Figure 4.3. Each panel shows the stock recruitment simulations from 1000 replicates over 20 years with F_{cap} = 2, TACmin = 0, and TACmax = 500 000. In each panel, if SSB in q1 was greater than B_{lim}, then recruitment was drawn as a random sample from the estimated past recruitment. In the lower panel, for all replicates, estimated past recruitment was taken to be the median of the SESAM estimates. In the upper panel, each replicate took estimated past recruitment from the distribution of SESAM estimates with uncertainty.

4.5.2 An alternative stock recruitment relationship

We did a sensitivity test in which we used a different stock recruitment relationship in the operating model. In this sensitivity test, the stochastic forecast still used the same stock recruitment relationship as before, as described in the recruitment section of the operating model description above. The new stock recruitment relationship is based on models fit using the eqsr_Buckland function from the MSY package in R. We used the function to fit 1000 models of types "Ricker", "Segreg", and "Bevholt" to the past estimates of recruitment and SSB. Then to simulate recruitment in the operating model, we randomly selected one of these 1000 models and input the estimated SSB. For comparison, the rest of the MSE settings followed the unbounded escapement strategy.

4.6 Performance Statistics

The following table (Table 4.7) lists the performance statistics that were calculated. For every year of every replicate, F_{bar} is calculated by averaging fishing mortality across quarters and across ages 1 and 2; then the median and mean are calculated across replicates and years. SSB for comparison with B_{lim} is always calculated in q4, as this is the escaped SSB targeted by the forecast. Two types of risk (1 and 3) were calculated as described in ICES WKGMSE Report 2013. Additionally, since SSB in q1 determines recruitment, the risk of this SSB dropping below the break in the assumed hockey-stick model (at 72 101 tonnes = B_{loss} in quarter 1 of 2005) was calculated (see section 2.1.2). The percent of years where implementation error ($F_{historical}$) is creating an upper limit on the fishery is reported as at F_{hist}; this is the percent of years where the real catch (TRC) is predicted to be lower than the TAC.

Table 4.7 Performance statistic descriptions

F_{bar}.median	Median true $F_{age\ 1\text{-}2}$
F_{bar}.mean	Mean true $F_{age\ 1\text{-}2}$
SSB.median	Median SSB in quarter 4 in tonnes
SSB.mean	Mean SSB in quarter 4 in tonnes
risk3.short.Q4	Probability of SSB in quarter 4 is below B_{lim}. The maximum risk in one of the years 2018–2022 is used (ICES Risk type 3)
risk1.long.Q4	Probability of SSB in quarter 4 is below B_{lim}. The average risk in the years 2023–2037 is used (ICES Risk type 1)
risk3.long.Q4	Probability of SSB in quarter 4 is below B_{lim}. The maximum risk in any of the years 2023–2037 is used (ICES Risk type 3)
risk1.long.Q1	Probability of SSB in quarter 1 is below the inflection point in the Hockey-stick SR applied. The average risk in the years 2023–2037 is used (ICES Risk type 1)
atFhist	the probability (%) that the TAC will require a true $F_{age\ 1\text{-}2}$ higher than $F_{historical}$ to be taken.
TAC.median	Median TAC in tonnes
TAC.mean	Mean TAC in tonnes
TRC.median	Median Total Realized Catch weight in tonnes (catch taken with a true F capped at $F_{historical}$)
TAC.change	Mean change in TAC from one year to the next, calculated from the absolute values of (TAC_{y+1} - TAC_y)
TAC.changeRel	Mean change in relative TAC from one year to the next, calculated from the absolute values of (TAC_{y+1} - TAC_y)/ TAC_y
atTACmin	Probability (%) that TAC is set at TACmin
atTACmax	Probability (%) that TAC is set at TACmax
YearToRecover. Q4	Average number of years it takes to rebuild SSB in quarter 4 to above B_{lim} (for cases whereSSB has been below B_{lim})
YearToRecover. Q1	Average number of years it takes to rebuild SSB in quarter 1 to above the inflection point in the Hockey-stick SR applied.

4.7 Implementation

The MSE simulation code was written in R. We ran the simulations on a high performance computing cluster with 20 cores. The 1000 replicates were run in parallel across the cores. Performing the simulations for one HCR took approximately 2 hours.

5 Results

This section presents the results from a high number of scenarios (Table 5.1), testing the effect of combinations of F_{cap}, TACmin and TACmax. Scenario 1 presents the default escapement strategy, Scenarios 2–9 present combination of TACmin and TACmax to answer request 1, while scenarios 10–28 also include the use of F_{cap} to answer request 2. Sensitivity analyses with respect to recruitment are done in scenario 29–30.

Table 5.1. Overview of HCR tested. Scenario 1–30 use $B_{escapement}$ at B_{lim}, while Scenario 41–44 $B_{escapement}$ at B_{pa}.

	Scenario	F_{Cap}	TACmin	TACmax	$F_{historical}$
Default Escapement, Scenario 1	1	5.0	0	5000000	5.0
Request 1, Scenario 2	2	2.0	0	150000	0.89
Request 1, Scenario 3	3	2.0	0	200000	0.89
Request 1, Scenario 4	4	2.0	20000	150000	0.89
Request 1, Scenario 5	5	2.0	20000	200000	0.89
Request 1, Scenario 6	6	2.0	30000	150000	0.89
Request 1, Scenario 7	7	2.0	30000	200000	0.89
Request 1, Scenario 8	8	2.0	40000	150000	0.89
Request 1, Scenario 9	9	2.0	40000	200000	0.89
Request 2, Scenario 10	10	0.3	0	150000	0.89
Request 2, Scenario 11	11	0.4	0	150000	0.89
Request 2, Scenario 12	12	0.3	0	200000	0.89
Request 2, Scenario 13	13	0.4	0	200000	0.89
Request 2, Scenario 14	14	0.3	20000	150000	0.89
Request 2, Scenario 15	15	0.4	20000	150000	0.89
Request 2, Scenario 16	16	0.3	20000	200000	0.89
Request 2, Scenario 17	17	0.4	20000	200000	0.89
Request 2, Scenario 18	18	0.3	30000	100000	0.89
Request 2, Scenario 19	19	0.4	30000	100000	0.89
Request 2, Scenario 20	20	2.0	30000	100000	0.89
Request 2, Scenario 21	21	0.3	30000	150000	0.89
Request 2, Scenario 22	22	0.4	30000	150000	0.89
Request 2, Scenario 23	23	0.3	30000	200000	0.89
Request 2, Scenario 24	24	0.4	30000	200000	0.89
Request 2, Scenario 25	25	0.3	40000	150000	0.89
Request 2, Scenario 26	26	0.4	40000	150000	0.89
Request 2, Scenario 27	27	0.3	40000	200000	0.89
Request 2, Scenario 28	28	0.4	40000	200000	0.89
Sensitivity, Eqsim Scenario 29	29	0.3	20000	150000	0.89
Sensitivity, Eqsim, Scenario 30	30	0.4	30000	200000	0.89
Request 3, Scenario 41	41	0.3	40000	150000	0.89
Request 3, Scenario 42	42	0.4	40000	150000	0.89

	Scenario	F_{Cap}	TACmin	TACmax	$F_{historical}$
Request 3, Scenario 43	43	0.3	40000	200000	0.89
Request 3, Scenario 44	44	0.4	40000	200000	0.89

5.1 Default Escapement strategy.

The ICES TAC advice for Norway pout is based on an escapement strategy with no upper limit on TAC and no F_{cap} applied. This HCR is not tested precautionary using the methods describe in this report, as the probability of quarter 4 SSB is below B_{lim} is higher than 5%. Figure 5.1 shows the historical assessment and the scenario projections, and Table 5.2 presents the performance statistics. See Table 3.6.1 for description of each of the variable names used in Table 5.2.

Table 5.1. Performance statistics for request default escapement strategy. An $F_{historical}$ at 5.0 and a TACmax at 5 million tonnes were applied for technical reasons (tolerance in optimizations). (Probabilities and risks are in percentages).

	Default escapement strategy
F_{cap}:	5.0
TACmin	0
TACmax	5000000
F_{bar}.median	0.494
F_{bar}.mean	0.952
SSB.median	89310
SSB.mean	104769
risk3.short.Q4	9.2
risk1.long.Q4	9.2
risk3.long.Q4	10.4
risk1.long.Q1	7.3
atFhist	4.7
TAC.median	98696
TAC.mean	181387
TRC.median	98696
TAC.change	208235
TAC.changeRel	22494.9
atTACmin (TAC at zero tonnes)	18.0
atTACmax	0.0
YearToRecover.Q4	1.20
YearToRecover.Q1	1.79

The default escapement strategy produces a probability of SSB in quarter 4 being less than B_{lim} at around 10% which is above the precautionary ICES criteria (of a probability of 5%). The higher probability is probably linked to the very high scenario F applied in some combinations with very high recruitments and the observation errors (from the assessment emulator). $F_{historical}$ was set to 5.0 (as an arbitrarily high value of F) however this value is reached in 4.7% of the cases (row atFhist in Table 5.2). Such a high F used for forecast would probably be capped if it was observed in the ICES advice (see e.g. the ICES advice from October 2015, where an Fcap at 0.6 was applied), but for this scenario the unconstrained value was used. The results show however that ICES should explore the use of an upper F (F_{cap}) for forecast (as shown in Annex 4) or inclusion of conditions on TACmin/ TACmax as explored below, or mixing both approaches, in order to make the escapement strategy precautionary.

The variability in the projections regarding recruitment and SSB in quarter 4 is coherent with past observed variability in the assessment (Figure 5.1). The variability in realized F and catches are far higher than in the past. The latter is simply due to the fact that the unconstraint escapement strategy rule can allow catches well above those observed in the past in cases when good recruitments are predicted to increase the population to high levels. In those cases, the required F to produce those catches fall outside assessment ranges. This derives from the approach followed here of testing the rule unconditional on the past historical assessed F values.

Figure 5.1 presents both the median of SSB for quarter 4 of 2017 and the median scenario SSB for the same period. The two SSB values are not the same. The median of stock numbers (N) at age drawn from a log-normal distribution will give an unbiased estimate of the median N. However the median of the sum of log-normal distributed N at ages (weighted by mean weight at age and proportion mature, as used for the calculation of SSB) will not be the same as the sum of median N at ages (weighted by mean weight at age and proportion mature, as used for the calculation of SSB) as used for the calculation of SESAM SSB. This discrepancy is an application of Jensen's inequality (Jensen, 1906).

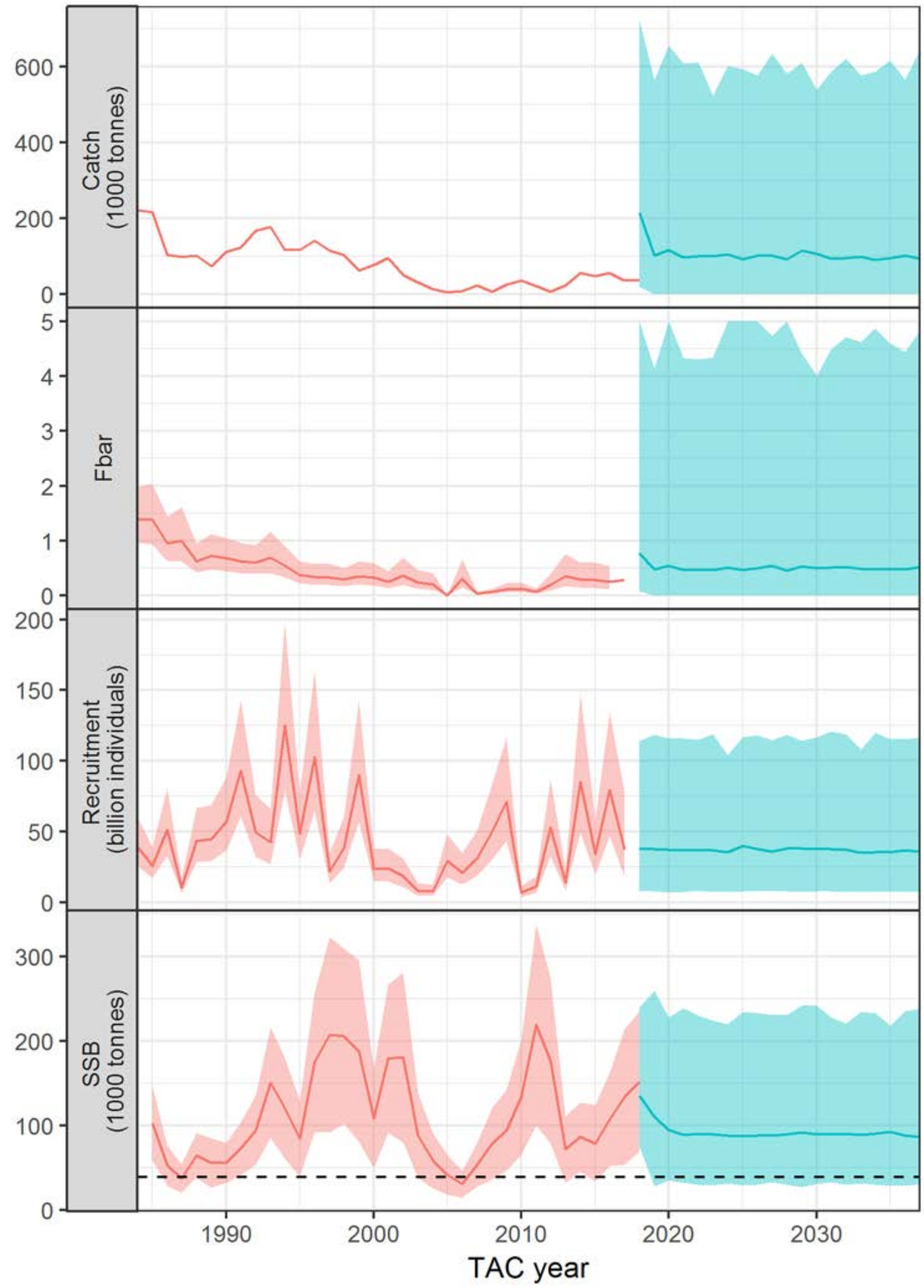


Figure 5.1. Summary result from the SESAM assessment of Norway pout (in red) and scenario values using the (un-constrained) escapement strategy as presently applied by ICES (in green). Catch is catch weight by TAC year, F_{bar} is the average of quarterly F_{age1-2} within a TAC year, Recruitment is stock number at age zero in the beginning of quarter 3, and SSB is SSB in the beginning of quarter 4. The lines show the median value and the shaded area the 5th and 95th percentiles.

5.2 Escapement strategies with TACmax & TACmin, but no F_{cap} (request 1)

The long-term (2023–2037) performance statistics for scenarios with TACmin in the range 20 to 40 kt in combination with TACmax in the range 150 to 200 kt are presented in Table 5.3. The performance statistics show that scenarios with a TACmin at 20kt in combination with a TACmax up to 200 kt, or with TACmin at 30kt in combinations with TACmax up to 150 kt, are considered precautionary using the ICES criteria. ICES considers an HCR as precautionary if the probability of SSB being below B_{lim} is less than 5%. For this stock this refers to the SSB in quarter 4 and a B_{lim} at 39 450 t. In this section, this probability is referred to as p(SSBq4 < B_{lim}).

Figure 5.2 shows p(SSBq4 < B_{lim}) for the individual years in the scenarios. The probability is low (~0) in the first two years after which the probability increases and reaches a plateau. The asymptotic probability is achieved rather quickly as expected for this short lived species, however there is still some variation in the annual p(SSBq4 < B_{lim}) in the expected equilibrium period (i.e. long-term). This is probably because the number of replicates in the MSE (1000) is not sufficient to reach equilibrium for this stock with a high process noise and with individual pools of recruitments for each replicate.

The default ICES approach for MSE is to use the maximum of annual probabilities (ICES risk type 3) in an equilibrium period (row "risk3.long.Q4" in Table 5.3) as basis to evaluate if a HCR is precautionary. We have, however, chosen to use the average p(SSBq4 < B_{lim}) ("risk1.long.Q4" in Table 5.3) equivalent to the ICES risk type 1 as basis for the evaluation. Risk type 1 and type 3 should be the same in equilibrium provided sufficient number of replicates (ICES, 2013). By using the average probability over 15 years we assume that we get a more accurate probability than using the maximum probability within any of the years.

For the precautionary scenarios (Scenario 2–6 in Table 5.3) the largest median TAC is obtained by using the 150 kt TACmax and no TACmin, while the highest mean TAC is obtained by TACmax at 200 kt and TACmin at 20 kt. For scenarios with a TACmin (scenario 4–6), the median TAC is quite the same (124–130 kt) for the three scenarios, but highest for the combination with lowest TACmax and TACmin at 20kt. The mean TAC is highest for the TACmax at 200 kt and TACmin at 20 kt. The lowest absolute change in TAC between years is obtained by TACmax at 150 kt in combination with TACmin at 30 kt.

The large differences in median and mean TAC for a given scenario indicate a skewed distribution of TAC, which is also clearly seen from Figure 5.3. The distributions of TAC are clearly bimodal, with a high probability of being constrained by TACmin or TACmax. For scenarios with no TACmin (scenario 2–3), the probability of a zero TAC is around 16%, and is almost the same irrespective of the TACmax values tested. With TACmin (scenario 4–6) the probability of setting the TAC at TACmin is 21–24% (row atTACmin in Table 5.3) and the probability of reaching TACmax (row atTACmax in Table 5.3) is in the range 36–46%, lowest for the highest TACmax.

The variation in TAC from one year to the next (Figure 5.4) show for scenarios without a TACmin, the maximum change is defined by the TACmax value. That means that the TAC can change from e.g. 150 kt in one year to zero TAC in the next, or from zero TAC to TACmax in following years. With TACmin the maximum change becomes constrained by TACmax minus TACmin.

The distribution of F (Figure 5.5) shows a distribution constrained by $F_{historical}$ (0.89) which is reached in around 9% (see also row atFhost in Table 5.3) of the cases for TACmax at 150 kt and around 17% for TACmax at 200kt. The application of TACmin barely affects these probabilities.

The distribution of SSB (Figure 5.6) is quite the same for all the scenarios. The highest median and mean SSB is obtained with no TACmin and TACmax at 150 kt. With application of TACmin, the median and mean SSB are around 1% higher when TACmin at 20kt is applied compared to the TACmin at 30kt.

Table 5.2. Performance statistics for request 1. Shaded scenarios have more than 5% probability of SSB in Quarter 4 being below B_{lim} (row "risk1.long.Q4" > 5%). (Probabilities and risks are in percentages)

	scen 2	scen 3	scen 4	scen 5	scen 6	scen 7	scen 8	scen 9
F_{cap}	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
TACmin	0	0	20000	20000	30000	30000	40000	40000
TACmax	150000	200000	150000	200000	150000	200000	150000	200000
F_{bar}.median	0.359	0.423	0.367	0.429	0.391	0.452	0.423	0.485
F_{bar}.mean	0.387	0.441	0.420	0.475	0.444	0.500	0.473	0.529
SSB.median	118036	111597	116075	109826	114750	108491	113070	106990
SSB.mean	151692	142050	149571	139969	147773	138229	146196	136598
risk3.short.Q4	2.2	2.3	3.1	3.2	3.6	3.9	5.1	5.2
risk1.long.Q4	2.3	2.4	3.7	3.9	4.9	5.1	6.2	6.4
risk3.long.Q4	3.1	3.4	5.2	5.8	6.5	6.9	8.3	8.7
risk1.long.Q1	1.9	2.0	3.4	3.5	4.6	4.8	6.0	6.3
atFhist	9.2	16.7	9.1	16.4	9.6	16.9	11.1	18.2
TAC.median	132170	125961	130054	123299	128023	122024	126556	120333
TAC.mean	96908	115542	99597	118119	101179	119789	103419	121895
TRC.median	117236	112466	115297	110791	114184	109715	113003	108671
TAC.change	49911	69988	44950	64608	41826	61308	38496	57656
TAC.changeRel			1.0	1.3	0.7	0.9	0.6	0.7
atTACmin	13.7*	14.2*	20.8	21.7	24.6	25.4	28.2	29.2
atTACmax	46.5	36.9	46.0	36.4	45.7	36.2	45.3	35.9
YearToRecover.Q4	1.20	1.21	1.37	1.36	1.47	1.48	1.61	1.60
YearToRecover.Q1	1.66	1.65	1.79	1.79	1.90	1.89	2.00	2.00

***probability of zero TAC**

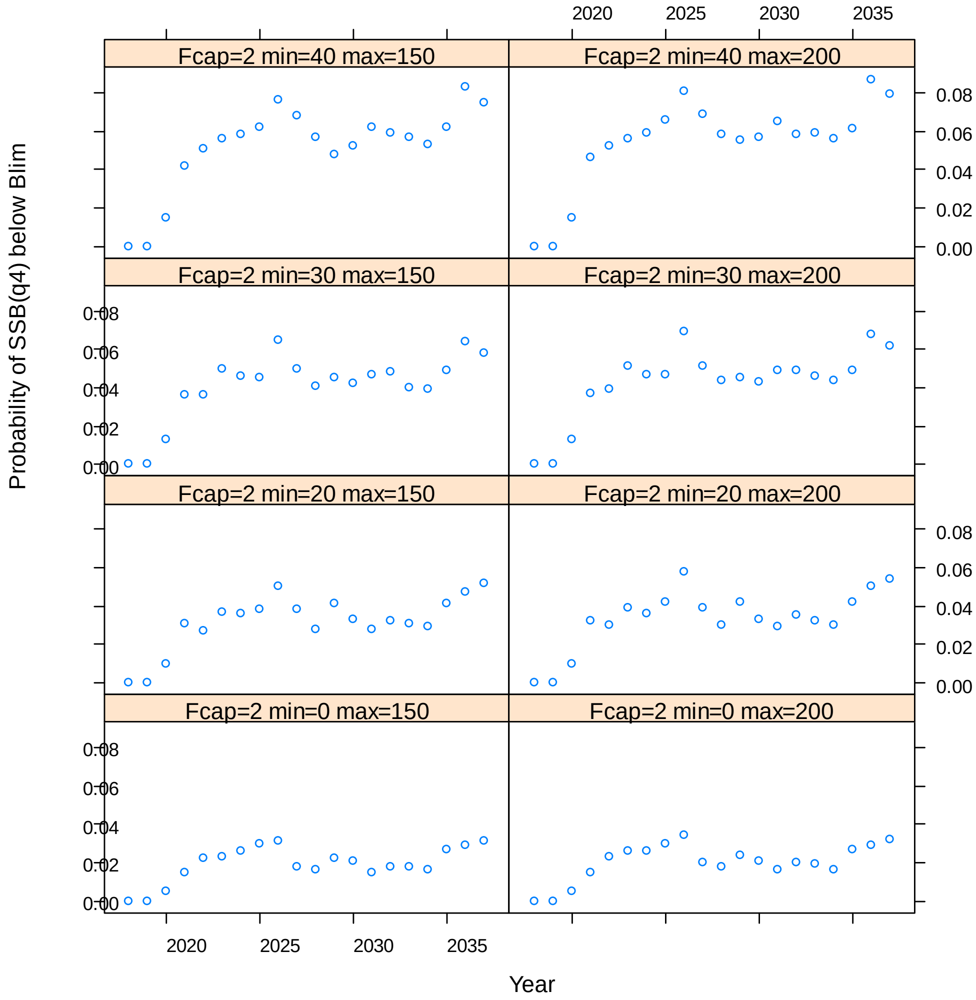


Figure 5.2. Request 1. Probability of SSB in quarter 4 is below B_{lim} by year and scenario. The headings for each subplot give the scenario values of F_{cap}, TACmin and TACmax.

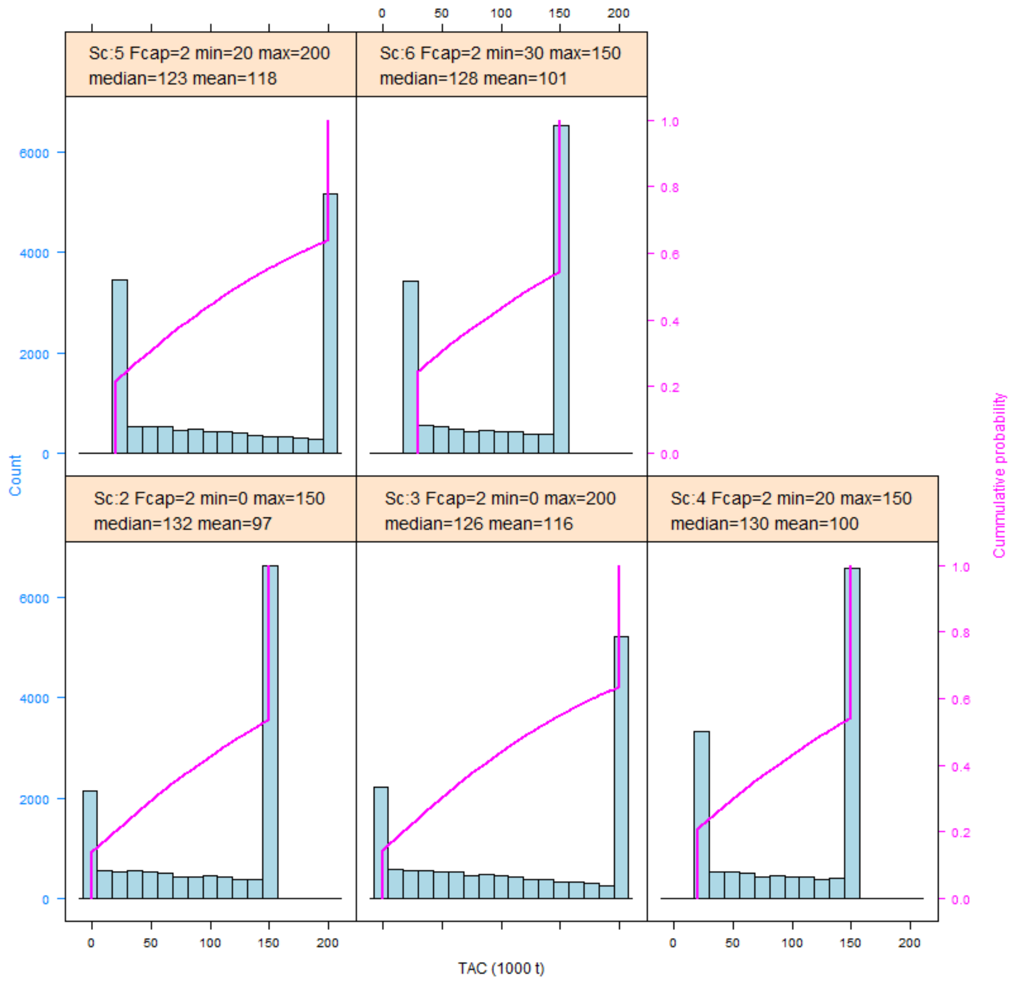


Figure 5.3. Request 1. Long-term distribution of TAC from individual years and replicates, including the cumulative probability, by scenario 2–6.

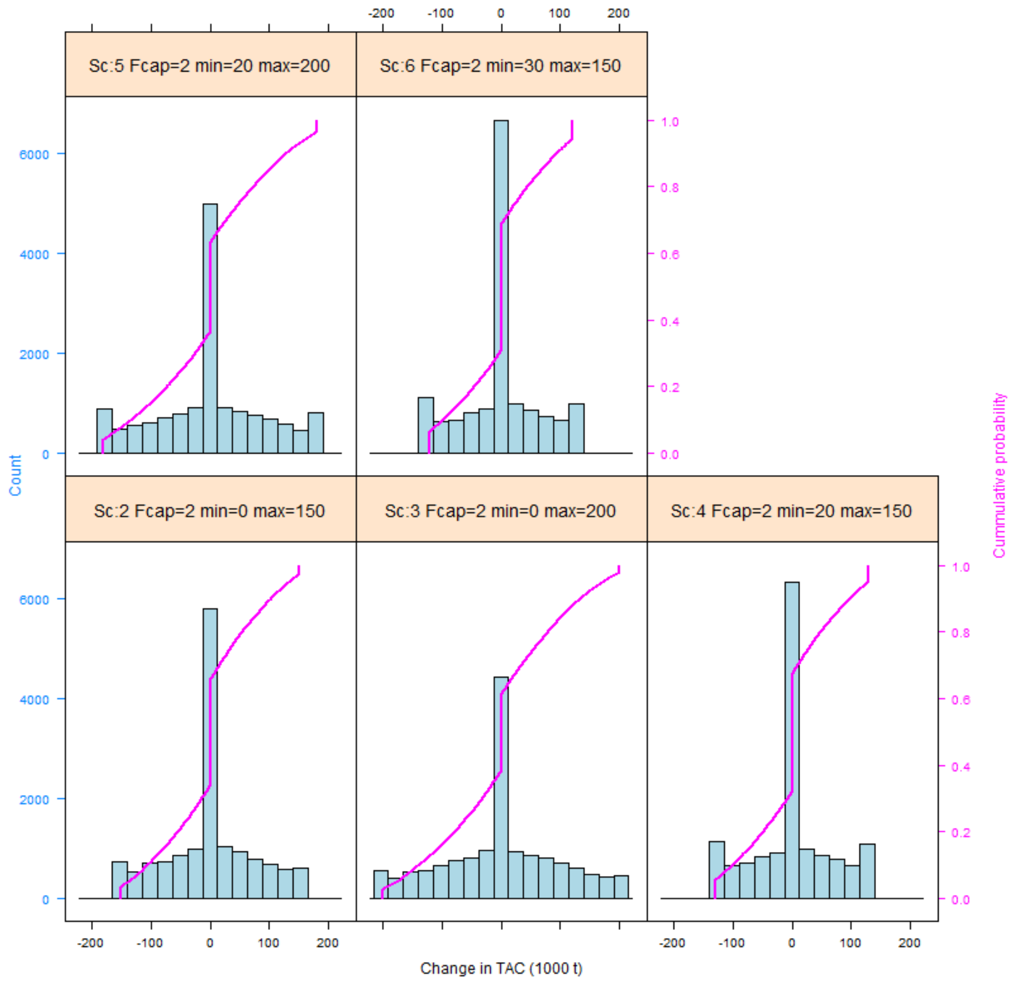


Figure 5.4. Request 1. Long-term distribution of TAC changes from one year to the next, including the cumulative probability, by scenario 2–6.

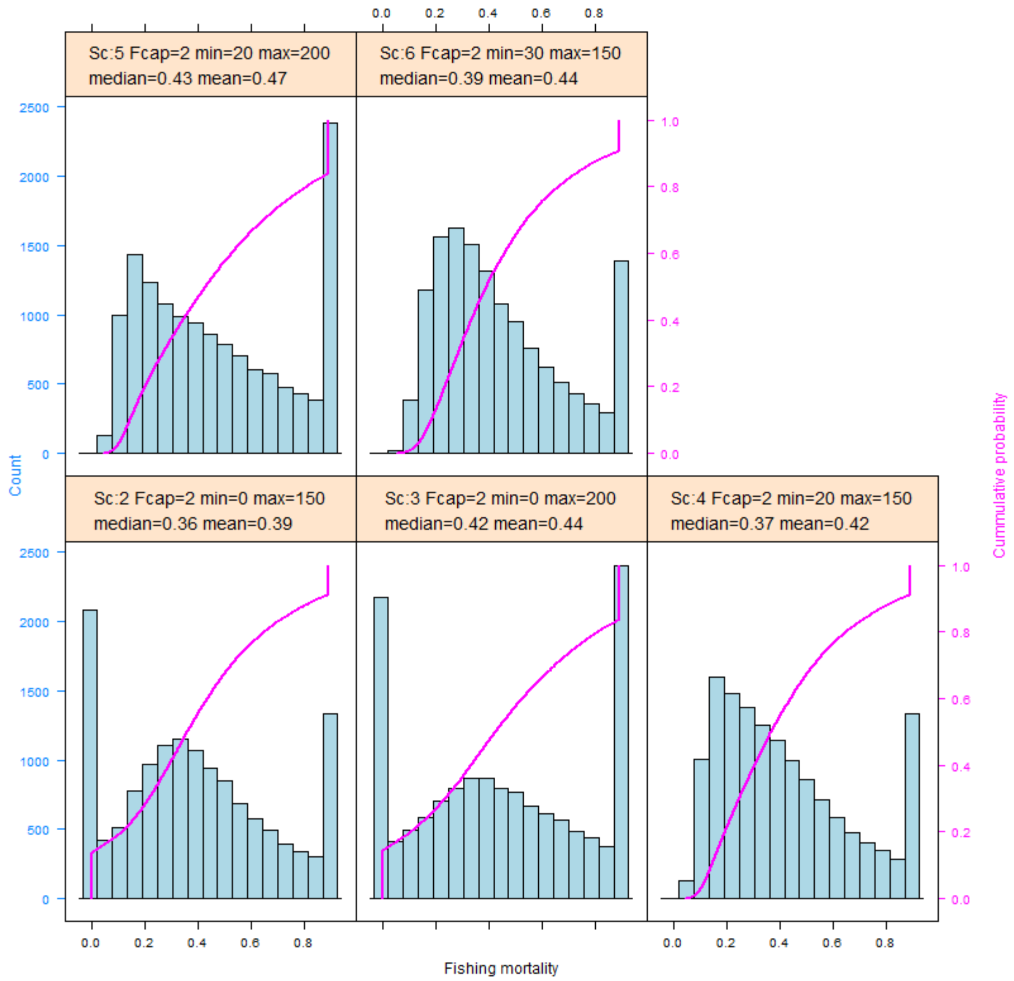


Figure 5.5. Request 1: Long-term distribution of true F for individual years and replicates, including the cumulative probability, by scenario 2–6.

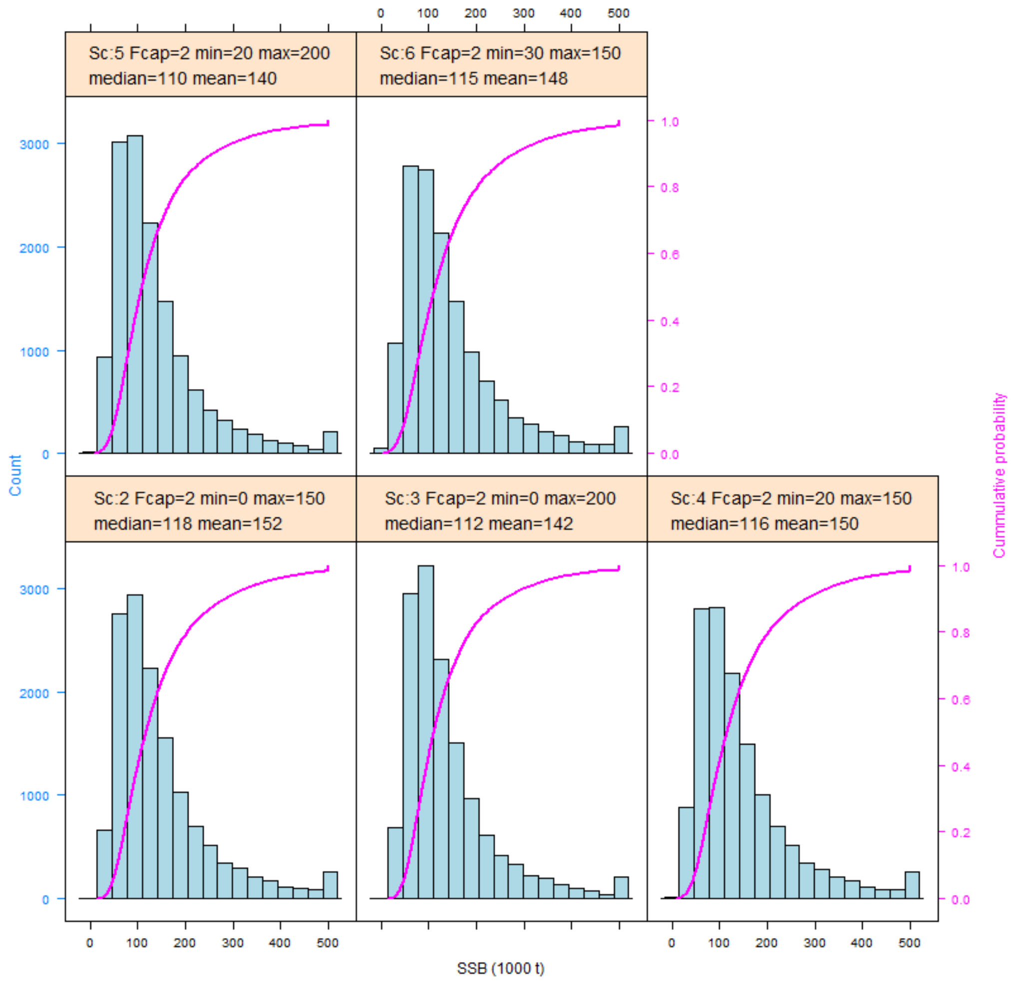


Figure 5.6. Request 1: Long-term distribution of true SSB(quarter 4) from individual years and replicates, including the cumulative probability, by scenario 2–6. SSB larger than 500 kt have been truncated on the plot.

5.3 Escapement strategy with TACmax & TACmin and F_{cap}. (Request 2)

Scenarios with TACmin at 30kt or lower, in combination with F_{cap} 0.3 to 0.4 and TACmax less than 200 kt have a p(SSBq4 < B_{lim}) below 5% for the both short (row risk3.short.Q4 of Table 5.4) and long term (row risk1.long.Q4). The probability of SSB(quarter 1) is below the hockey-stick reflection point from the SR is less than 5% (row risk1.long.Q1**)** for these combinations. This shows that the target of having SSB in quarter 4 above B_{lim} is sufficient to maintain SSB at spawning time in quarter 1 such that recruitment is not impaired by low SSB. Scenarios with TACmin at 40 kt resulted in long term (row risk1.long.Q4) above 5% and therefore they are not considered precautionary according to the ICES criteria.

Median TAC decreases by increasing TACmin (Figure 5.7), but the loss is less than 5% for the range of applied TACmin. The relative losses are slightly higher when mean TAC is used (Figure 5.8). The highest median TAC is obtained by the highest F_{cap}, while an increase in TACmax from 150 kt to 200 kt gives actually a small reduction in median TAC.

The absolute change in TAC from one year to the next (row TAC.change in Table 5.4) increases by increasing TACmax (Figure 5.9) as the TAC span increases. Likewise, the change in TAC decreases by increasing TACmin.

The distribution of TACs for selected precautionary scenarios (Figure 5.10) shows a bimodal distribution with the highest frequencies of TAC at TACmin and TACmax.

The variation in TAC from one year to the next (Figure 5.11) show that the full range (TACmax-TACmin) is obtained, even though an F_{cap} is applied. F_{cap} minimizes the TAC variation, but the effect is not large. Scenario 16 (F_{cap}=0.3) has, for example, a higher frequency of small changes than scenario 17 (F_{cap}=0.4). The same is shown in row TAC.changes (absolute change in TAC from one year to the next) in Table 5.4 where TAC.changes increases from 47kt for scenario 16 to 55 kt for scenario 17.

As expected, TACmax is reached more often in combinations with the lowest TACmax and the highest F_{cap} (Figure 5.12). The probability of a TAC at TACmin increases with an increasing TACmin, while the values of F_{cap} and TACmax have a limited influence on this probability.

F_{cap} at 0.3 and 0.4 reduce the probability of reaching $F_{historical}$ to levels of 0.7–3.6% for precautionary scenarios (row atFhist in Table 5.4). This low probability makes the MSE robust to the assumption of an $F_{historical}$ at the presently chosen value (0.89). The use of F_{cap} gives also a unimodal distribution of true F (Figure 5.13), and probably also a more stable fishing effort.

The distributions of SSB (Figure 5.14) seem similar for the scenarios. Mean SSB is larger than the median SSB (Figure 5.15). The highest SSB is from scenarios with the lowest F_{cap} and lowest TACmax. The increase in TACmin from 0 to 40kt reduces SSB by around 5%.

The past (assessment) and future (scenario) values of Catch, F and SSB are plotted for one example of the precautionary scenarios (Figure 5.16). Scenario F and TAC are considerably higher than the recent historical values. This difference is due to the assumption that the TAC will actually be taken (up to F at $F_{historical}$), while TAC has been far from limiting in the fisheries for most of the recent years. The 90% confidence limits of scenario catches are defined by the TACmin and TACmax.

Table 5.3. Performance statistics for request 2. Shaded scenarios have more than 5% probability of SSB in Quarter 4 being below B_{lim} (row "risk1.long.Q4" > 5%) (none). (Probabilities and risks are in percentages)

	scen 14	scen 15	scen 16	scen 17	scen 18	scen 19	scen 21
F_{cap}	0.3	0.4	0.3	0.4	0.3	0.4	0.3
TACmin	20000	20000	20000	20000	30000	30000	30000
TACmax	150000	150000	200000	200000	100000	100000	150000
F_{bar}.median	0.278	0.325	0.289	0.348	0.270	0.294	0.298
F_{bar}.mean	0.304	0.358	0.314	0.379	0.297	0.329	0.327
SSB.median	130447	123637	129123	120708	133303	129159	128784
SSB.mean	162631	156007	159630	151028	167883	164517	160648
risk3.short.Q4	2.6	3.0	2.6	3.0	3.2	3.2	3.3
risk1.long.Q4	3.2	3.6	3.2	3.6	4.0	4.3	4.1
risk3.long.Q4	4.6	4.8	4.6	4.9	5.4	5.7	5.5
risk1.long.Q1	2.9	3.2	2.9	3.2	3.8	4.0	3.9
atFhist	0.7	2.2	0.9	3.1	0.8	1.4	1.1
TAC.median	72265	89742	71968	88465	72941	91547	71706
TAC.mean	77453	87686	81944	95345	69002	74100	79107
TRC.median	72240	89162	71885	87797	72930	91003	71692
TAC.change	39497	42865	47072	54578	23265	22868	36744
TAC.changeRel	0.8	0.8	0.8	1.0	0.4	0.4	0.6
atTACmin	19.1	20.1	19.2	20.4	22.6	22.9	22.9
atTACmax	14.3	23.4	6.7	12.4	32.9	45.8	14.2
YearToRecover.Q4	1.41	1.39	1.41	1.39	1.49	1.49	1.52
YearToRecover.Q1	1.73	1.77	1.73	1.76	1.84	1.87	1.84

Table 5.4 (continued). Performance statistics for request 2. Shaded scenarios have more than 5% probability of SSB in Quarter 4 being below B_{lim} (row "risk1.long.Q4" > 5%). (Probabilities and risks are in percentages).

	scen 22	scen 23	scen 24	scen 25	scen 26	scen 27	scen 28
F_{cap}	0.4	0.3	0.4	0.3	0.4	0.3	0.4
TACmin	30000	30000	30000	40000	40000	40000	40000
TACmax	150000	200000	200000	150000	150000	200000	200000
F_{bar}.median	0.346	0.307	0.369	0.318	0.372	0.328	0.395
F_{bar}.mean	0.382	0.338	0.404	0.355	0.410	0.366	0.432
SSB.median	122030	127501	119407	127046	120278	125938	117590
SSB.mean	154085	157681	149131	159008	152514	156032	147565
risk3.short.Q4	3.4	3.3	3.4	4.2	4.6	4.2	4.6
risk1.long.Q4	4.6	4.2	4.7	5.3	5.7	5.3	5.9
risk3.long.Q4	5.9	5.5	6.1	7.3	7.9	7.2	8.0
risk1.long.Q1	4.3	3.9	4.4	5.1	5.6	5.2	5.8
atFhist	2.7	1.4	3.6	2.4	4.1	2.6	5.0
TAC.median	89236	71367	88057	71361	88230	71029	87130
TAC.mean	89391	83574	97107	81325	91539	85743	99189
TRC.median	88574	71364	87449	71344	87686	71017	86617
TAC.change	40005	44260	51715	33638	36872	41063	48369
TAC.changeRel	0.6	0.7	0.7	0.5	0.5	0.5	0.6
atTACmin	23.8	23.1	24.2	28.1	27.6	28.4	28.0
atTACmax	23.2	6.6	12.3	14.1	23.0	6.6	12.2
YearToRecover.Q4	1.50	1.52	1.48	1.62	1.58	1.61	1.58
YearToRecover.Q1	1.86	1.84	1.85	1.91	1.95	1.91	1.95

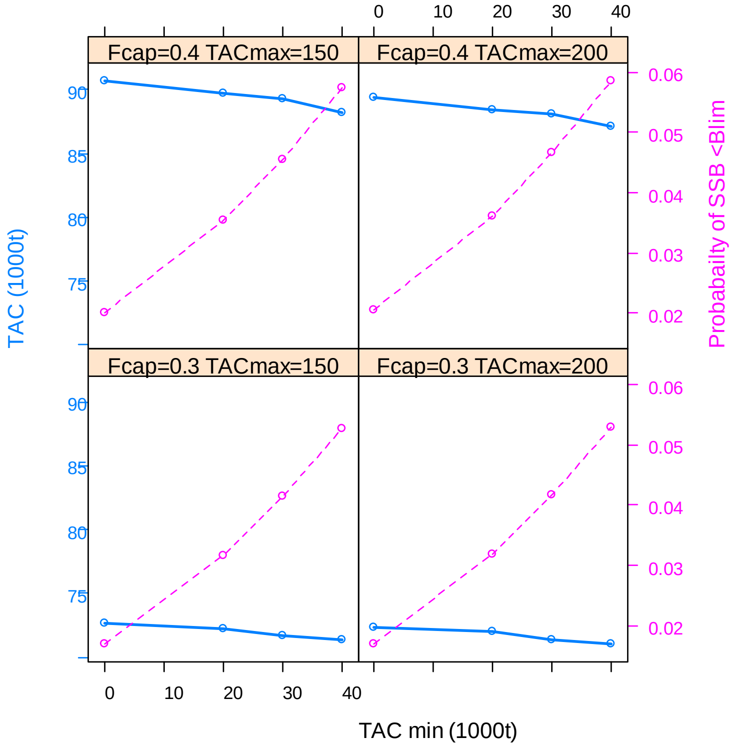


Figure 5.7. Median TAC and probability of SSB below B_{lim} by combinations of F_{cap}, TACmax and TACmin. Request 2.

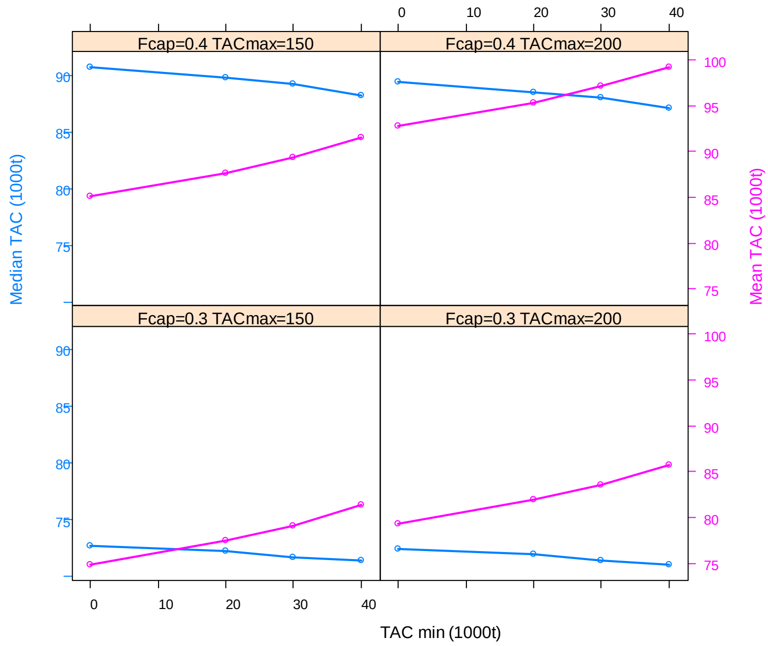


Figure 5.8. Median and mean TAC by combinations of F_{cap}, TACmax and TACmin. Request 2.

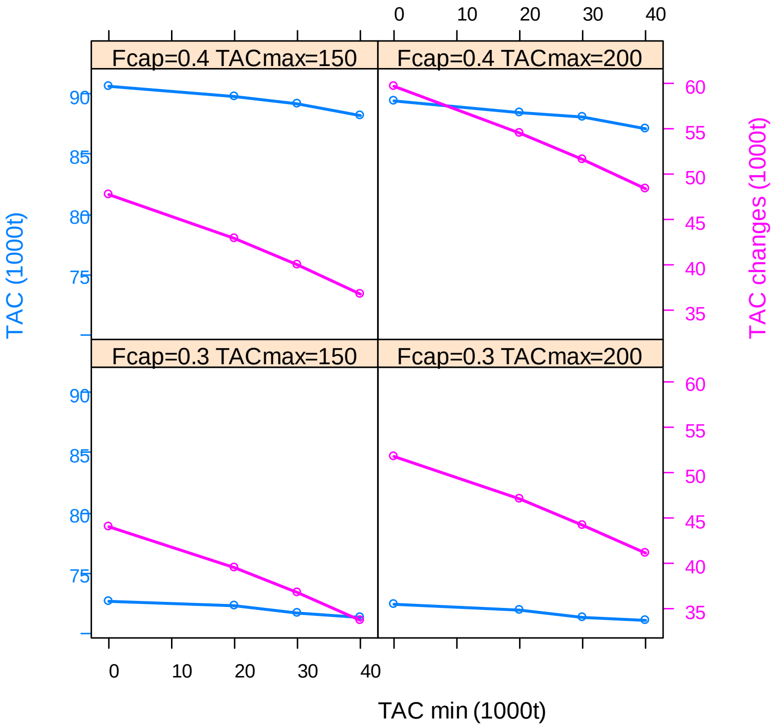


Figure 5.9. Request 2: Median TAC and absolute change in TAC from one year to the next by combinations of F_{cap}, TACmax and TACmin.

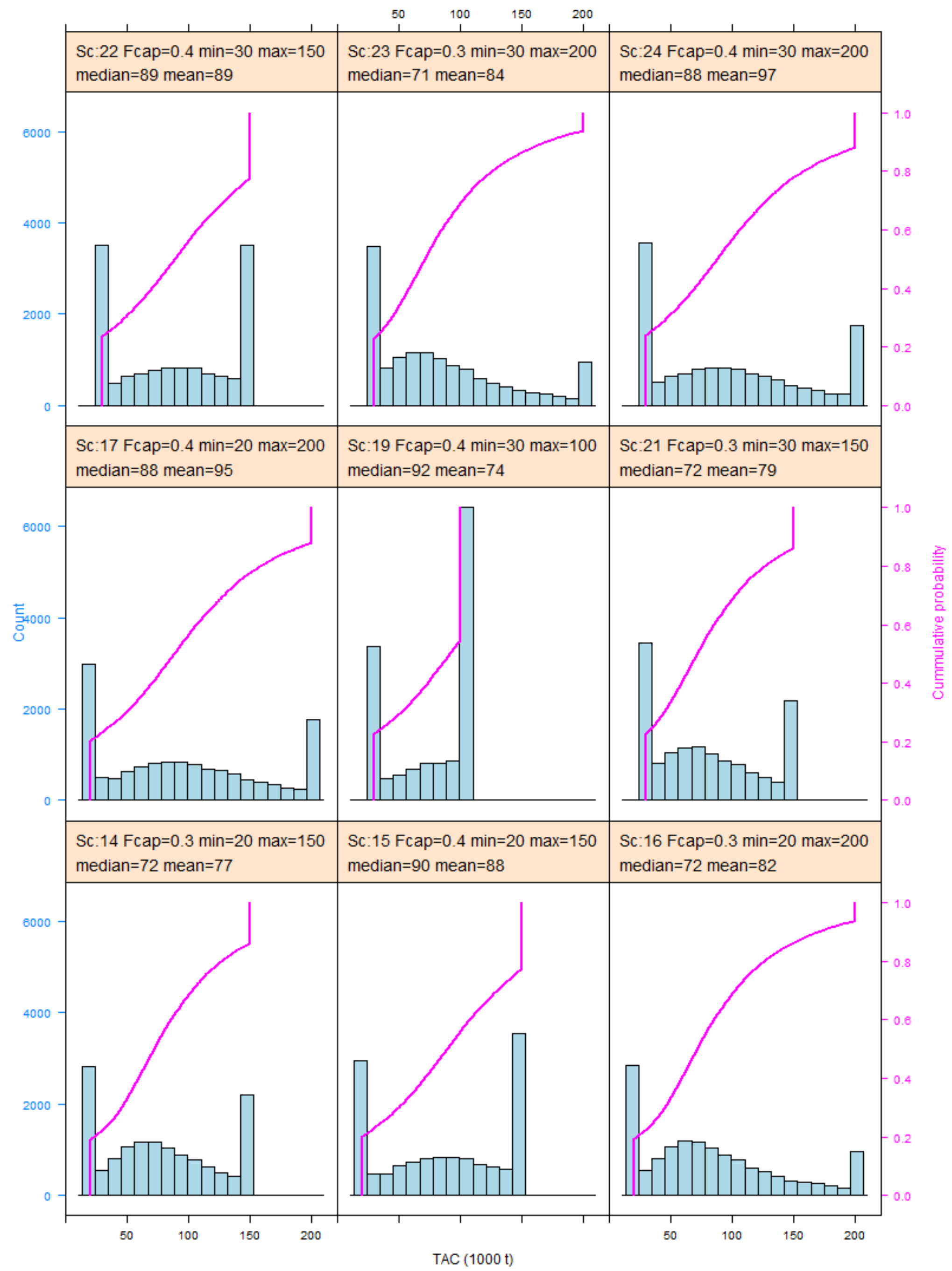


Figure 5.10. Request 2: Long-term distribution of TAC for individual years and replicates, including the cumulative probability, by precautionary scenarios.

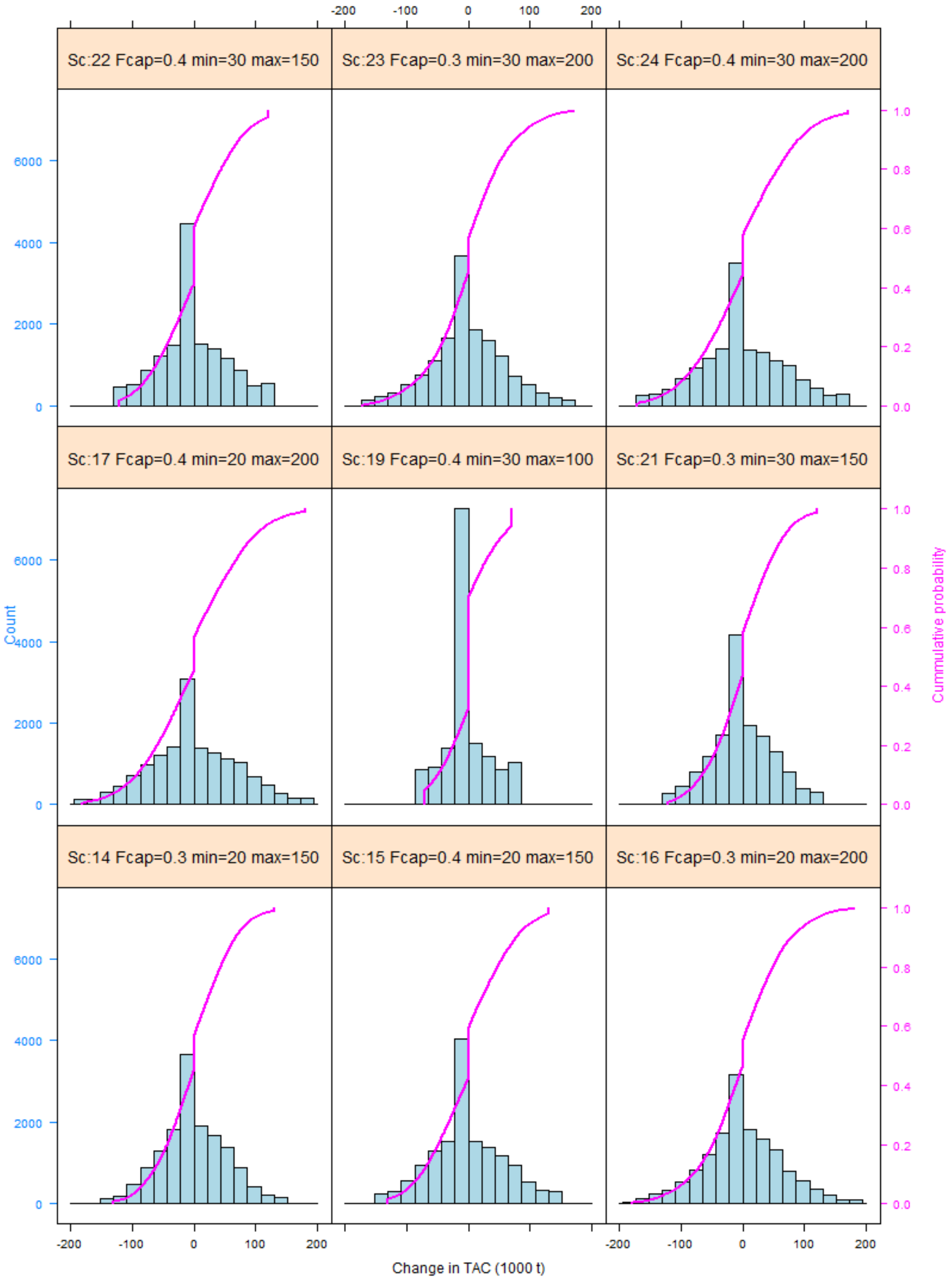


Figure 5.11. Request 2. Long-term distribution of TAC changes from one year to the next, including the cumulative probability.

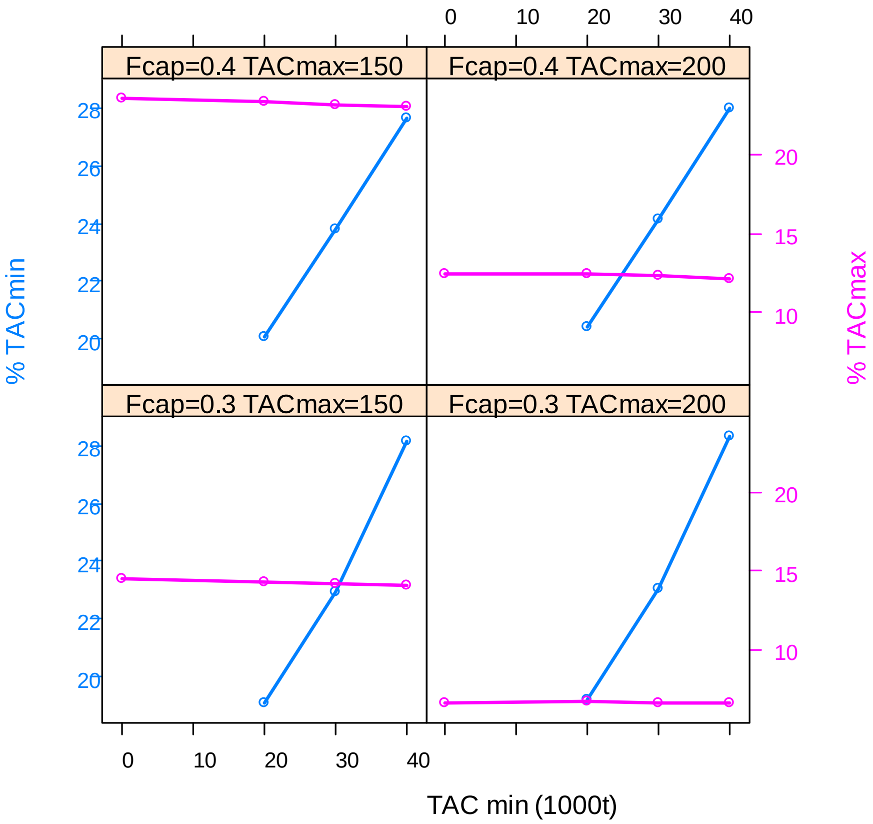


Figure 5.12. Request 2: Percentage of TACs reaching TACmin or TACmax.

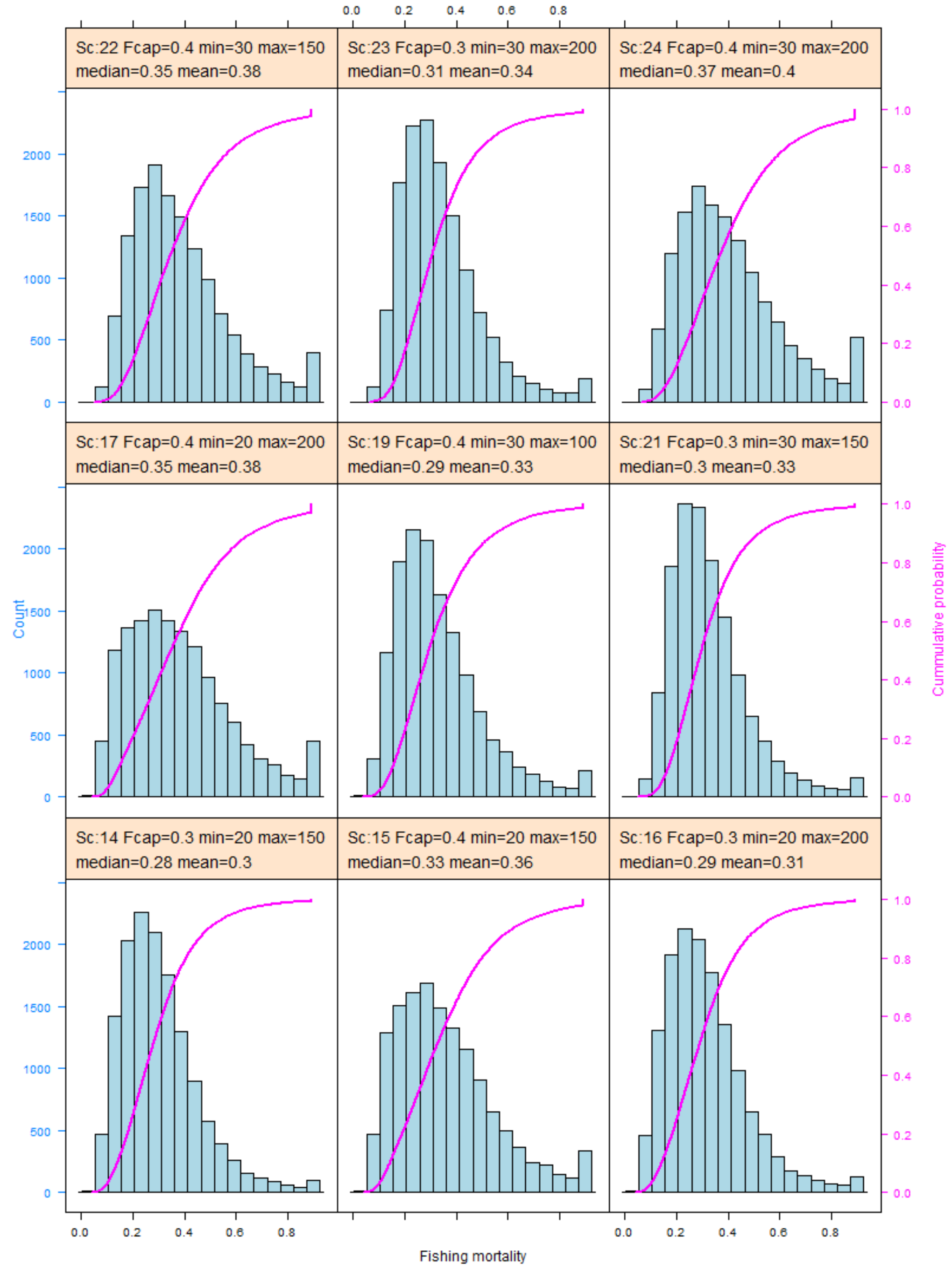


Figure 5.13. Request 2: Long-term distribution of true F from individual years and replicates, including the cumulative probability.

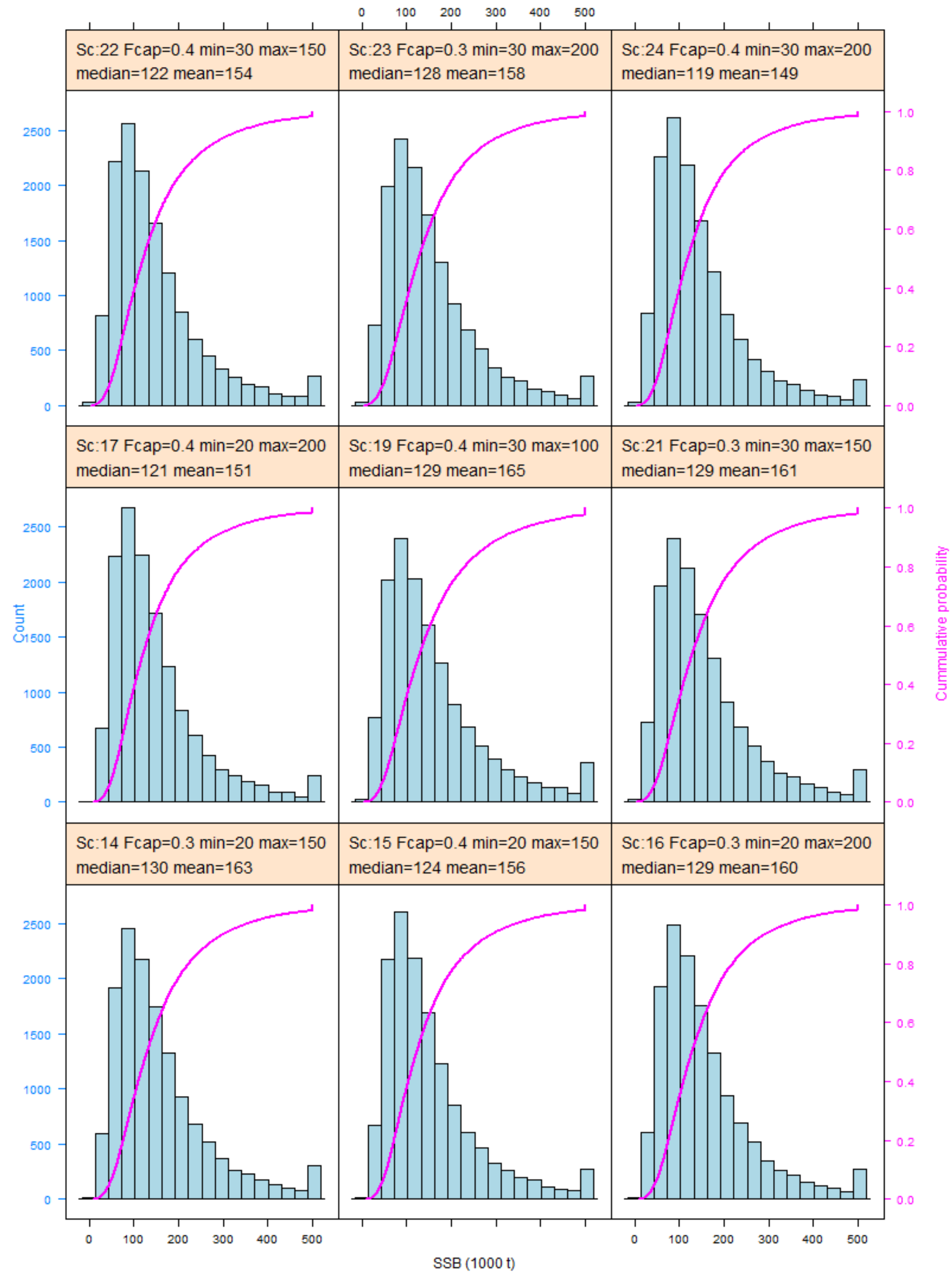


Figure 5.14. Request 2: Long-term distribution of true SSB (Quarter 4) from individual years and replicates, including the cumulative probability.

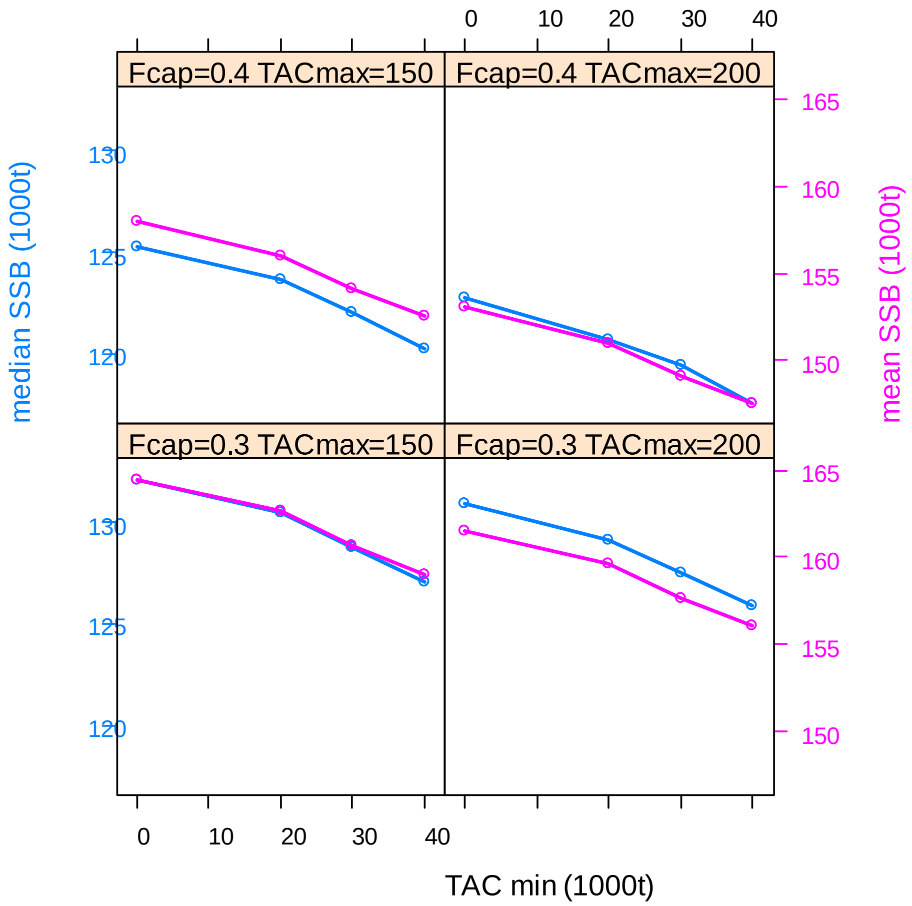


Figure 5.15. Request 2: Median and mean true SSB by combinations of F_{cap}, TACmax and TACmin.

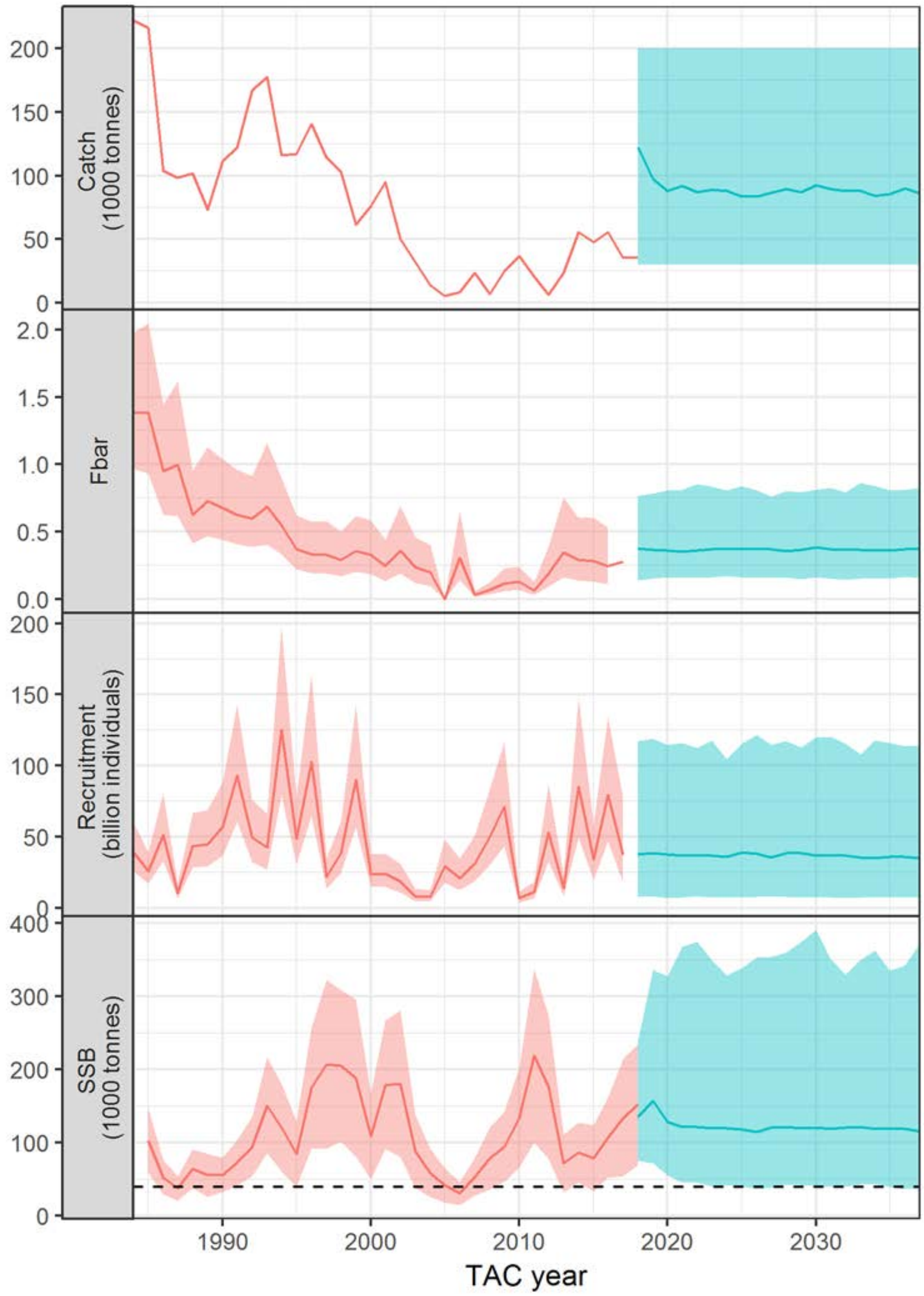


Figure 5.16. Request 2: Summary result from the SESAM assessment of Norway pout (in red) and scenario values using one example scenario (scenario 24: (F_{cap} = 0.4, TACmin = 30kt and TAC-max = 200 kt) (in green). The lines show the median value and the shaded areas the 5th and 95th percentiles (the shaded area in the catches scenario is defined by TACmin and TACmax).

5.4 Adding safe guard clauses at low SSB values (Request 3)

The scenarios evaluated for request 1 and request 2 show that a TACmin of maximum of 30 kt would be considered precautionary. As an example of an HCR which would allow a higher TACmin, the target quarter 4 $SSB_{escapement}$ is increased from B_{lim} (39 450 t) to 65 000 t (which is Bpa). This HCR will in general leave a higher biomass after the TAC has been taken and thereby reduce the probability of SSB below B_{lim} in Quarter 4. The scenarios evaluate TACmin at 40 kt, in combination with F_{cap} in the range 0.3–0.4 and TACmax in the range 150–200 kt.

Three out of the four scenarios are considered precautionary given the ICES criteria of less than 5% probability of SSB below B_{lim} (row risk1.long.Q4 of Table 5.5). Scenario 44 with F_{cap} = 0.4 and TACmax = 200kt is not considered precautionary.

The mean TAC (~ 80 kt) for these scenarios are generally in the same order of size as the mean TAC from other scenarios with application of an F_{cap} in the range 0.3–0.4. Median TAC (~47kt) is however considerably lower. Around half of the TAC is set close to TACmin (Figure 5.17), which also explains the rather low median TAC.

To conclude: targeting a higher $SSB_{escapement}$ allows a 40 kt TACmin, which is the minimum can be fished every year, however the median TAC becomes close to the TACmin, even though the mean TAC is in the same order of size as other scenario presented with application of F_{cap}.

Table 5.4. Performance statistics for request 3 ($SSB_{escapement}$ = 65kt). Shaded scenarios have more than 5% probability of SSB in Quarter 4 being below B_{lim} (row "risk1.long.Q4" > 5%). (Probabilities and risks are in percentages)

	scen 41	scen 42	scen 43	scen 44
F_{cap}	0.3	0.4	0.3	0.4
TACmin	40000	40000	40000	40000
TACmax	150000	150000	200000	200000
F_{bar}.median	0.286	0.303	0.296	0.321
F_{bar}.mean	0.323	0.347	0.334	0.368
SSB.median	132145	129132	130727	126216
SSB.mean	162723	159469	159697	154653
risk3.short.Q4	3.6	3.7	3.6	3.9
risk1.long.Q4	4.87	4.95	4.89	5.02
risk3.long.Q4	6.3	6.3	6.4	6.4
risk1.long.Q1	4.77	4.84	4.79	4.90
atFhist	2.1	3.0	2.4	3.9
TAC.median	47753	46387	46278	44197
TAC.mean	75843	80923	80323	88537
TRC.median	47753	46357	46278	44166
TAC.change	34013	37221	41546	48775
TAC.changeRel	0.5	0.6	0.6	0.7
atTACmin	47.9	48.4	48.3	49.1
atTACmax	14.4	22.4	6.6	12.5
YearToRecover.Q4	1.61	1.62	1.62	1.61
YearToRecover.Q1	1.91	1.90	1.91	1.90

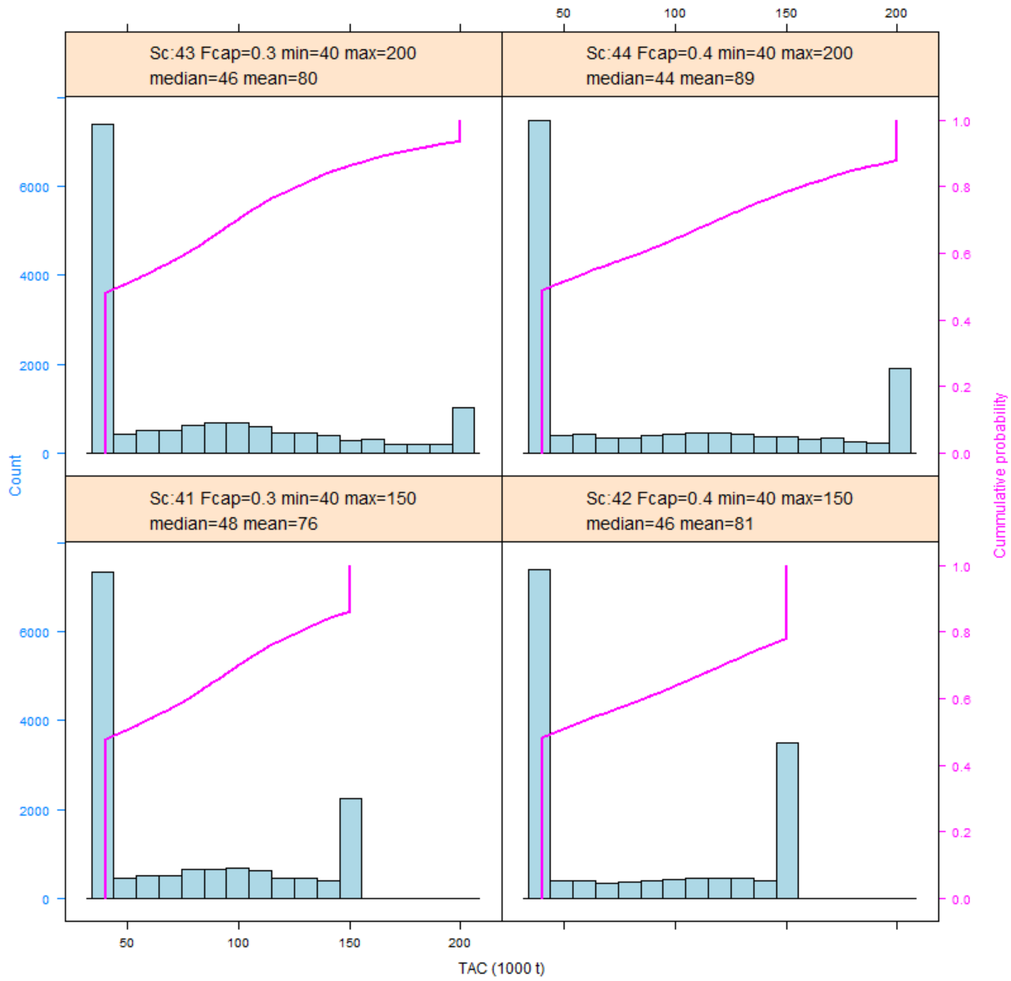


Figure 5.17. Request 3, ($SSB_{escapement}$ = 65kt): Long-term distribution of TAC for individual years and replicates, including the cumulative probability, by scenarios.

5.5 Sensitivity tests

5.5.1 Ignoring uncertainty in initial N, E, and R (i.e. one initial population)

When uncertainty in initial N, E, and R is ignored (i.e. starting the MSE from one true state, or here called the one population approach), the performance statistics for the scenarios tested before to answer requests 1 and 2 are quite similar (Table 5.6 and Table 5.7) (notice that scenarios 18 & 19 were not repeated here in the sensitivity analysis, as unnecessary). The plotted ratios (Figure 5.19) of selected values estimated by the one population and the 1000 population approach (e.g. TAC.median from scenarios using one initial population divided by TAC.median from scenarios using a thousand populations) show that the main difference between the two approaches appears in the probability of SSB below B_{lim}. This probability (risk1.long.Q4) is lower when the one population approach is used. This might be linked to a 4-5% higher recruitment (rec.median) when the one-population approach is used, but it might also be due to the reduced variability in the exploitation patterns (E). The higher recruitment might allow a generally higher F_{bar}. The TAC and the probability of setting TAC at TACmin are generally highest with the 1000 population approach.

Looking at the summary results from one scenario (scenario 24, Figure 5.18), they seem quite similar to the result from scenario 24 using the 1000 population approach (Figure 5.16). The main difference is a higher 90th percentile of recruitment in the 1000 population scenario.

The pattern of the probability of SSB below B_{lim} using the one population approach (Figure 5.20) is similar to the pattern from scenarios using a thousand populations (Figure 5.2), but the probabilities are generally lower when the uncertainty is ignored in the initial population.

To conclude, performance statistics like TAC, TRC and SSB for individual scenarios differ by generally less than 2–3% between methods using one initial population or a thousand populations. The probability of SSB being below B_{lim} is typically 1–2 percentage points higher when the thousand population approach is used. This means that some of the scenarios become not precautionary (i.e. get a higher than 5% probability) when the 1000 population approach is applied.

Table 5.5. Sensitivity test, one initial population. Performance statistics for request 1. Shaded scenarios have more than 5% probability of SSB in Quarter 4 being below B_{lim} (row "risk1.long.Q4" > 5%). (Probabilities and risks are in percentages).

	scen 2	scen 3	scen 4	scen 5	scen 6	scen 7	scen 8	scen 9
F_{cap}	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
TACmin	0	0	20000	20000	30000	30000	40000	40000
TACmax	150000	200000	150000	200000	150000	200000	150000	200000
F_{bar}.median	0.387	0.453	0.389	0.454	0.409	0.473	0.440	0.503
F_{bar}.mean	0.406	0.460	0.435	0.491	0.457	0.513	0.484	0.540
SSB.median	115260	109706	113987	108450	112905	107520	111651	105681
SSB.mean	142985	134239	141261	132550	139949	131086	138118	129302
risk3.short.Q4	1.6	1.7	2.5	2.7	3.3	3.4	4.0	4.5
risk1.long.Q4	1.7	1.8	2.8	3.0	3.8	4.1	5.1	5.4
risk3.long.Q4	2.4	2.4	4.0	4.2	4.7	5.3	6.3	6.8
risk1.long.Q1	1.4	1.5	2.6	2.8	3.7	4.0	5.2	5.4
atFhist	9.5	18.2	9.5	18.1	9.6	18.2	10.7	19.3
TAC.median	128355	123461	127323	122353	125925	120282	123900	118591
TAC.mean	97260	115350	99942	118184	101626	119704	103583	121630
TRC.median	115839	112279	115365	111541	114191	109926	112634	108661
TAC.change	51780	72573	47004	67506	43982	64236	40617	60783
TAC.changeRel	8622.7	10535.0	1.0	1.3	0.7	1.0	0.6	0.8
atTACmin	0.0	0.0	19.6	20.4	23.4	24.3	27.1	27.9
atTACmax	45.7	35.5	45.4	35.4	45.1	35.2	44.8	34.8
YearToRecover.Q4	1.17	1.16	1.31	1.31	1.41	1.42	1.54	1.51
YearToRecover.Q1	1.62	1.62	1.72	1.71	1.78	1.77	1.86	1.86

Table 5.6. Sensitivity test, one initial population. Performance statistics for request 2. Shaded scenarios (none) have more than 5% probability of SSB in Quarter 4 being below B_{lim} (row "risk1.long.Q4" > 5%). (Probabilities and risks are in percentages).

	scen 14	scen 15	scen 16	scen 17	scen 21	scen 22
F_{cap}	0.3	0.4	0.3	0.4	0.3	0.4
TACmin	20000	20000	20000	20000	30000	30000
TACmax	150000	150000	200000	200000	150000	150000
F_{bar}.median	0.284	0.338	0.292	0.358	0.301	0.357
F_{bar}.mean	0.307	0.367	0.316	0.386	0.327	0.388
SSB.median	129509	121759	128720	119609	128595	120869
SSB.mean	155285	148260	153177	144308	153983	146959
risk3.short.Q4	2.2	2.2	2.2	2.4	2.8	2.8
risk1.long.Q4	2.3	2.6	2.3	2.7	3.3	3.6
risk3.long.Q4	3.4	3.7	3.4	3.8	4.7	4.7
risk1.long.Q1	2.1	2.4	2.1	2.5	3.0	3.5
atFhist	0.6	2.3	0.7	3.1	0.7	2.4
TAC.median	71046	88218	70803	87339	70616	87763
TAC.mean	75911	87072	79261	93441	77558	88732
TRC.median	70971	87951	70785	87013	70586	87477
TAC.change	39949	44292	46195	55241	37276	41400
TAC.changeRel	0.8	0.9	0.8	1.0	0.6	0.6
atTACmin	17.6	18.7	17.7	18.9	21.6	22.5
atTACmax	11.4	20.9	4.7	9.5	11.4	20.7
YearToRecover.Q4	1.31	1.32	1.30	1.30	1.39	1.42
YearToRecover.Q1	1.68	1.69	1.69	1.69	1.74	1.75

Table 5.7 (cont.) Sensitivity test, one initial population. Performance statistics for request 2. Shaded scenarios (none) have more than 5% probability of SSB in Quarter 4 being below B_{lim} (row "risk1.long.Q4" > 5%). (Probabilities and risks are in percentages).

	scen 23	scen 24	scen 25	scen 26	scen 27	scen 28
F_{cap}	0.3	0.4	0.3	0.4	0.3	0.4
TACmin	30000	30000	40000	40000	40000	40000
TACmax	200000	200000	150000	150000	200000	200000
F_{bar}.median	0.309	0.375	0.322	0.380	0.331	0.400
F_{bar}.mean	0.336	0.407	0.353	0.414	0.362	0.434
SSB.median	127810	118817	127266	119640	126322	117357
SSB.mean	151851	142965	152297	145133	150104	141094
risk3.short.Q4	2.8	2.9	3.3	3.8	3.3	4.1
risk1.long.Q4	3.3	3.8	4.3	4.8	4.4	5.0
risk3.long.Q4	4.6	4.8	5.6	6.2	5.6	6.6
risk1.long.Q1	3.1	3.6	4.4	4.9	4.4	5.0
atFhist	0.9	3.2	1.6	3.4	1.8	4.3
TAC.median	70368	86818	70278	87032	70064	86107
TAC.mean	80883	95035	79620	90637	82921	96907
TRC.median	70357	86595	70253	86624	70047	85804
TAC.change	43490	52215	34400	38224	40512	48927
TAC.changeRel	0.6	0.7	0.5	0.5	0.5	0.6
atTACmin	21.7	22.9	26.7	26.3	26.8	26.8
atTACmax	4.6	9.4	11.3	20.5	4.5	9.4
YearToRecover.Q4	1.40	1.44	1.54	1.56	1.53	1.55
YearToRecover.Q1	1.74	1.75	1.83	1.84	1.83	1.83

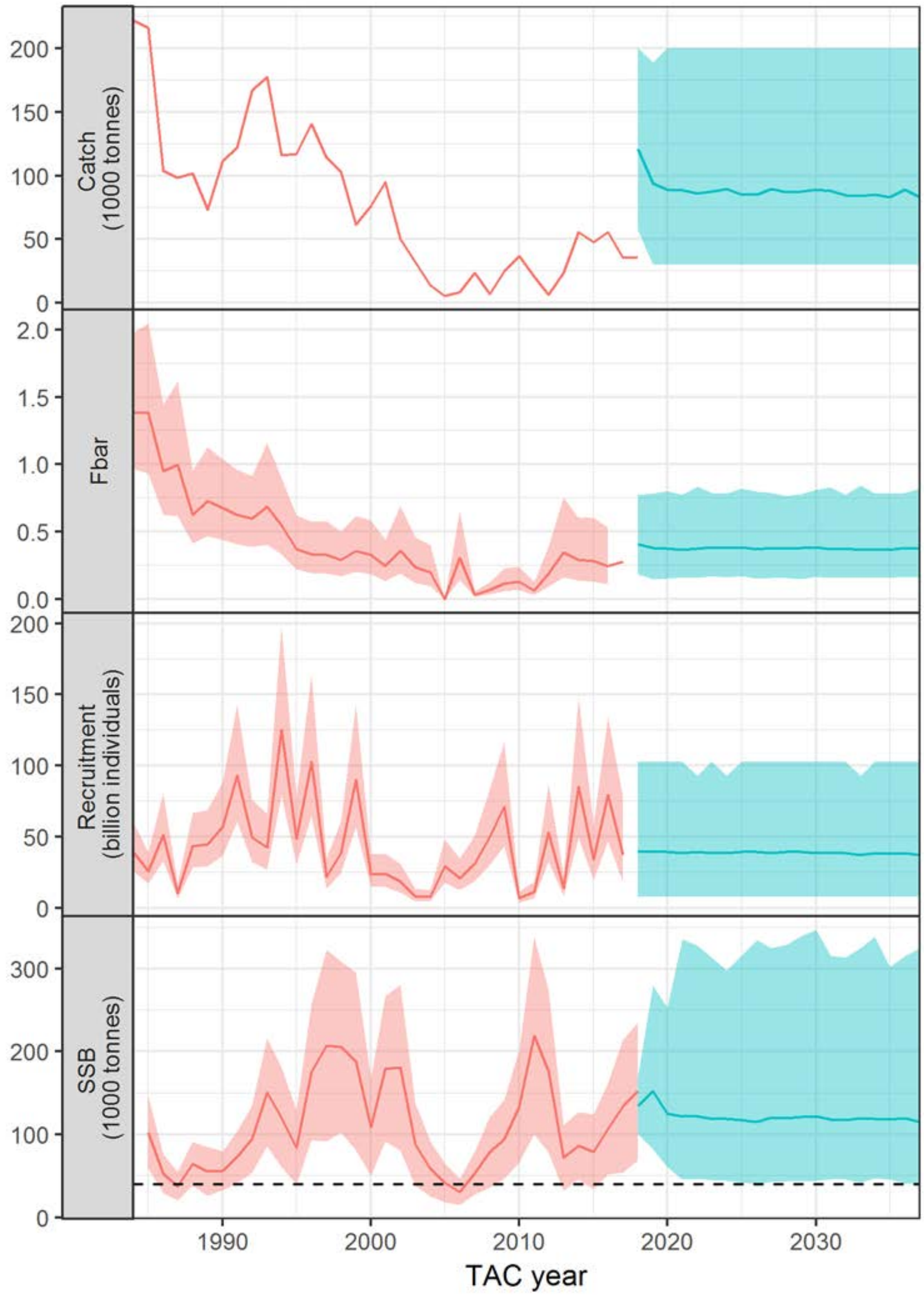


Figure 5.18. Sensitivity test, one initial population. Summary result from the SESAM assessment of Norway pout (in red) and scenario values using one example scenario (scenario 24: (F_{cap} = 0.4, TACmin = 30kt and TACmax = 200 kt) (in green). The lines show the median value and the shaded areas the 5th and 95th percentiles.

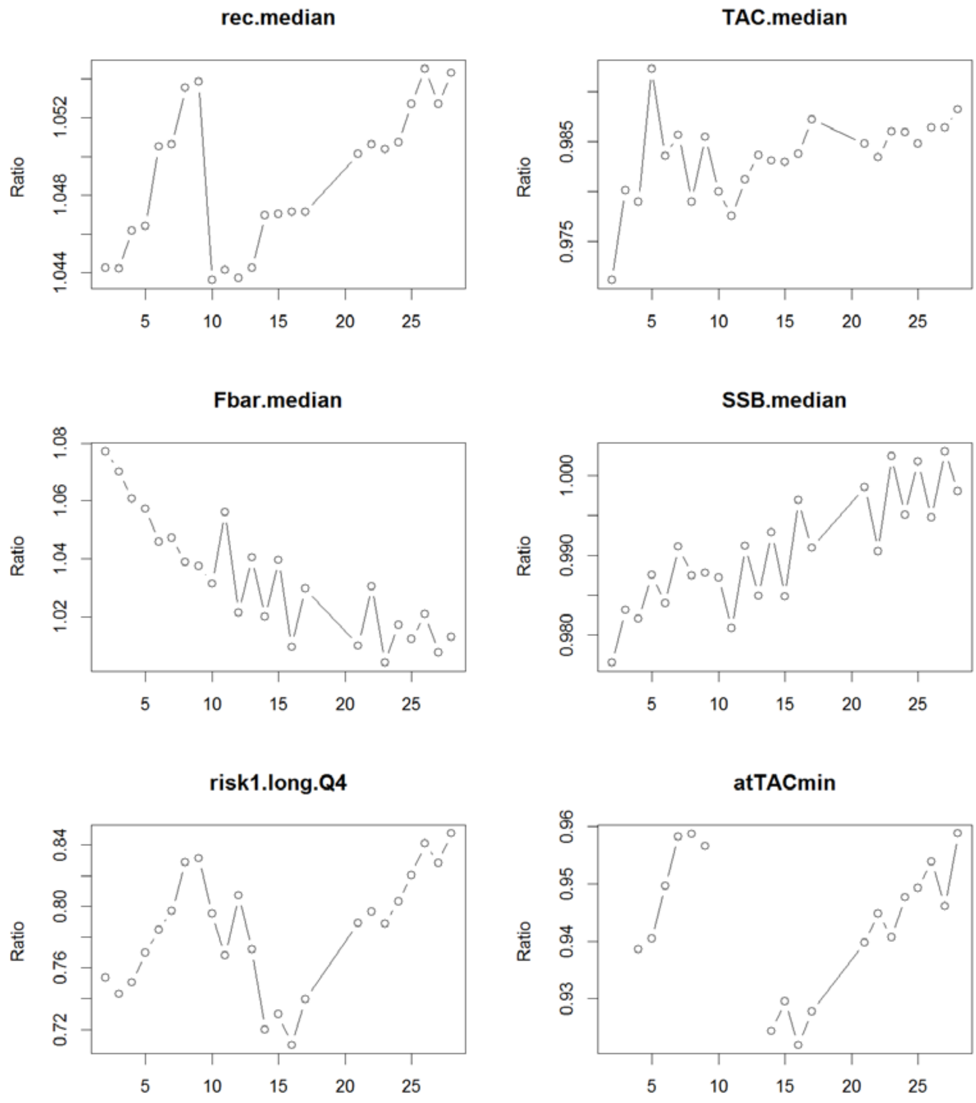


Figure 5.19. Sensitivity test, one initial population. The ratio of values estimated from scenarios using one initial populations and values estimated from scenarios using thousand populations (y-axis) by scenario (x-axis). Ratios below 1 indicate that the performance statistic is lower in the one population version than in the 1000 population version.

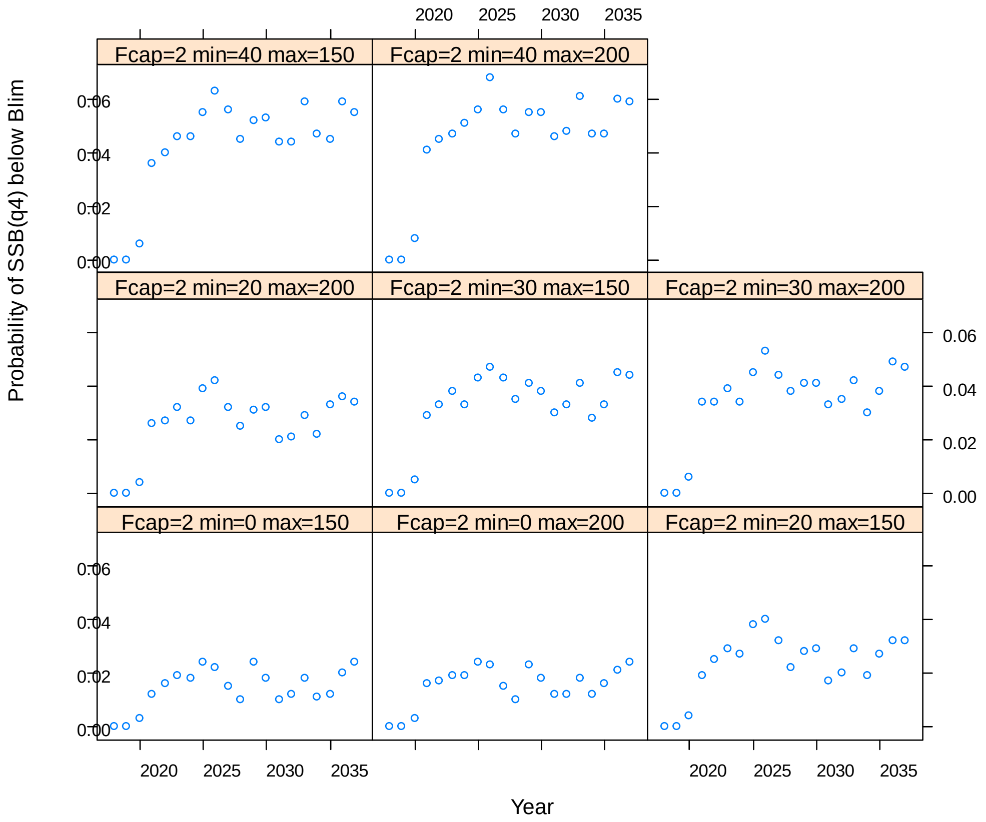


Figure 5.20. Sensitivity test, one initial population. Probability of SSB in quarter 4 is below B_{lim} by year and scenario. The headings for each subplot give the scenario values of F_{cap}, TACmin and TACmax.

5.5.2 An alternative stock recruitment relationship

This sensitivity test evaluates the effect of using an Eqsim fit of the stock recruitment relationship (SR) for two precautionary scenarios, one with a lower (3.2%) probability of SSB below B_{lim} (scenario 14) and one (scenario 24) with a higher (4.7%) probability. The Eqsim fit of SR estimates the highest weight for segmented regression (i.e. hockey stick 53%), some weight on the Ricker SR (38%) and low weight on the Beverton-Holt SR (9%) (Figure 5.21). Notice that the hockey stick fitting resulted in an inflection point around 190 000 t, well above the current B_{lim} 72 101 tonnes (B_{loss} in quarter 1 of 2005, as adopted in the benchmark). Given the consistent results of the last assessment with those managed during the benchmark, the WK did not change the B_{lim} reference point recently adopted in the last benchmark, but this result called for a sensitivity analysis on alternative SR fitting.

The performance statistics for scenario 14 and 24 and for the same two scenarios using the Eqsim SR (scenario 29 and 30 respectively) (Table 5.8) show a limited effect of the S-R. SSB and TAC are approximately 1–4% lower with the Eqsim SR.

The probability of SSB being below B_{lim}, row risk1.long.Q4 in Table 5.8, increases from 3.2% in scenario 14 to 3.3% is scenario 29, and from 4.7% in scenario 24 to 5.9% in scenario 30. Applying the ICES criteria, scenario 30 is not precautionary as the probability exceeds 5.0%. The criteria for precautionarity $p(SSBq4 < B_{lim})$ was applied in all four scenarios, but B_{lim} would probably have to be redefined if the Eqsim SR were determined to be the most appropriate. The most recent benchmark concluded however to use B_{loss} as B_{lim} as there is no apparent stock recruitment relationship.

Recruitment with Eqsim SR (Figure 5.22) has a higher 90th percentile and a marginal lower 50th percentile compared to recruitment with the default SR (Figure 5.16). The distributions of recruitment (Figure 5.23) are almost identical for scenario 14 (default SR) and scenario 29 (Eqsim SR). Comparing scenario 24 and 30, the Eqsim SR has a higher frequency of both the low and high recruitments. The lower frequency of the lower recruitments using the default SR might be due to the truncated noise function used to draw recruits when SSB is below B_{lim}, whereas Eqsim uses a non-truncated random noise for all SSB values.

To conclude, the sensitivity analysis shows that the application of an Eqsim fitted SR gives a marginally lower SSB and TAC compared to the values using the benchmark SR. The probability of SSB being below B_{lim} is only marginally increased, therefore sensitivity is minor.

Table 5.7 Performance statistics for sensitivity analysis.

	scen 14	scen 24	scen 29 scen 14 & Equsim	scen 30 scen 24 & Eqisim
F_{cap}	0.3	0.4	0.3	0.4
TACmin	20000	30000	20000	30000
TACmax	150000	200000	150000	200000
F_{bar}.median	0.278	0.369	0.279	0.370
F_{bar}.mean	0.304	0.404	0.305	0.406
SSB.median	130447	119407	127564	114767
SSB.mean	162631	149131	164676	148568
risk3.short.Q4	2.6	3.4	2.3	3.6
risk1.long.Q4	3.2	4.7	3.3	5.9
risk3.long.Q4	4.6	6.1	4.0	7.4
risk1.long.Q1	2.9	4.4	3.3	5.8
atFhist	0.7	3.6	0.7	4.2
TAC.median	72265	88057	70236	83326
TAC.mean	77453	97107	76408	93837
TRC.median	72240	87449	70221	82974
TAC.change	39497	51715	38972	49854
TAC.changeRel	0.8	0.7	0.7	0.7
atTACmin	19.1	24.2	18.7	25.6
atTACmax	14.3	12.3	13.7	11.2
YearToRecover.Q4	1.41	1.48	1.68	1.90
YearToRecover.Q1	1.73	1.85	1.91	2.17

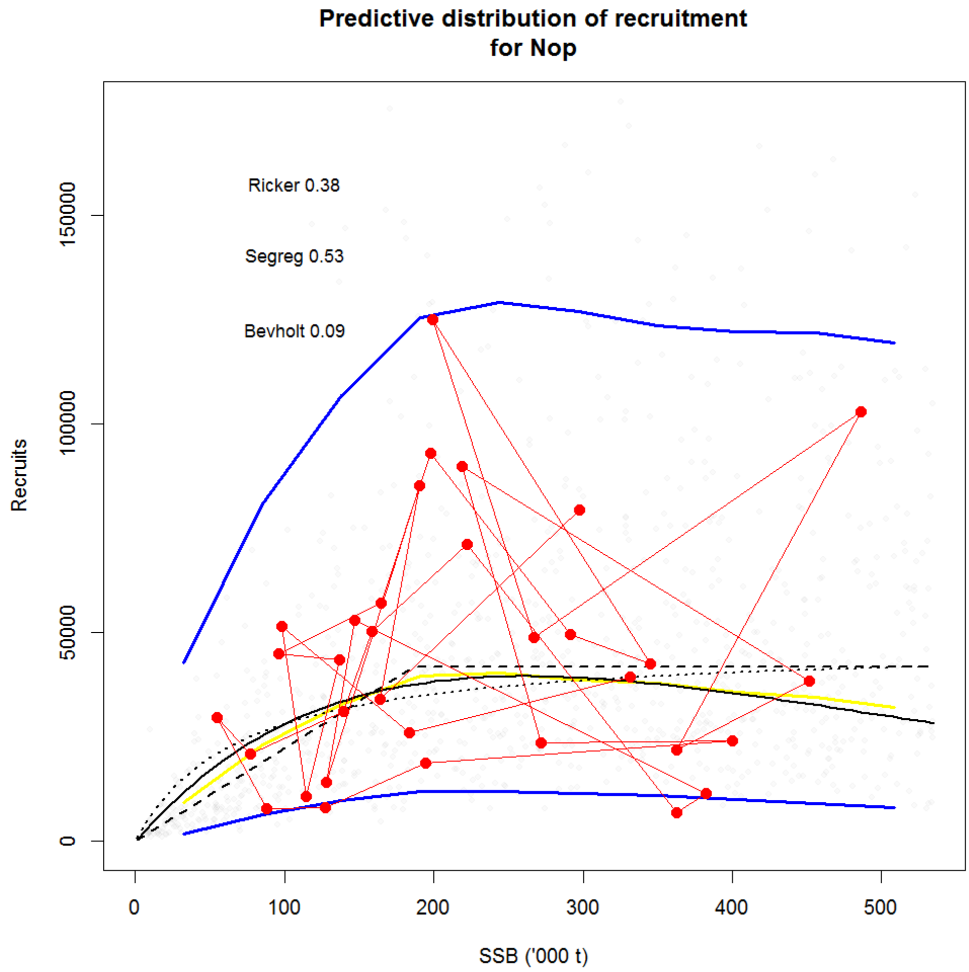


Figure 5.21. Fitted Eqsim SR model for Norway pout.

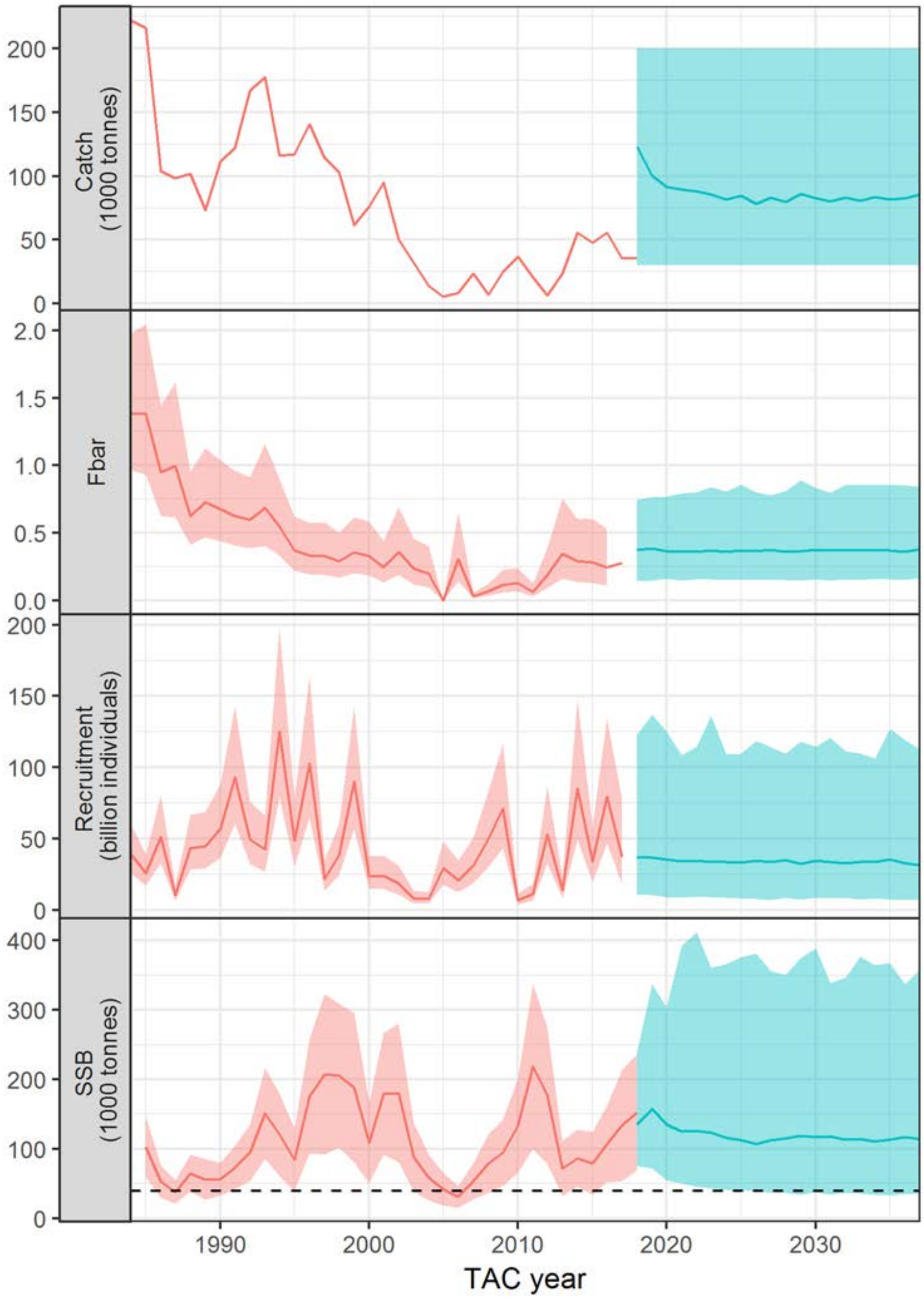


Figure 5.22. Sensitivity analysis: Summary result from the SESAM assessment of Norway pout (in red) and scenario values from scenario 29: (F_{cap} = 0.4, TACmin = 30 kt and TACmax = 200 kt and Eqsim SR) (in green). The lines show the median value and the shaded areas the 5th and 95th percentiles.

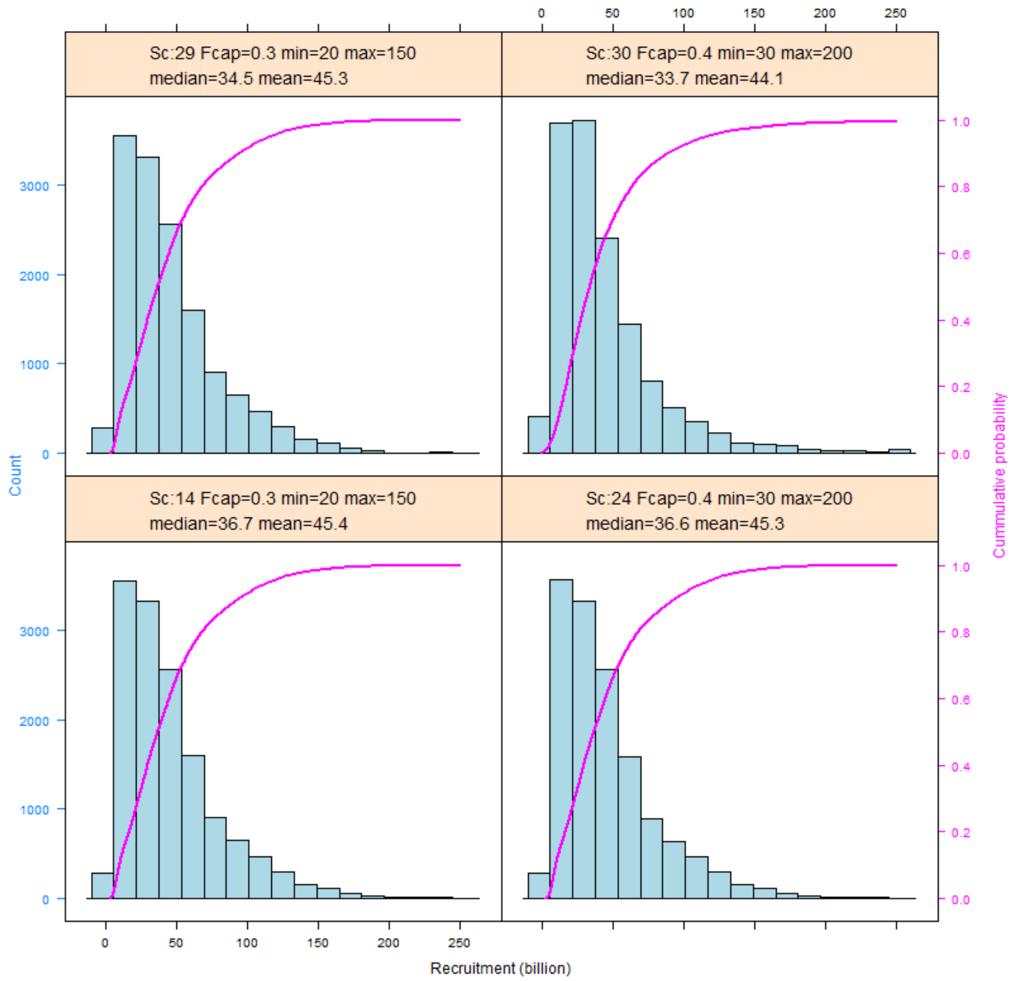


Figure 5.23. Sensitivity analysis: Long-term distribution of Recruitment from individual years and replicates, including the cumulative probability. Scenario 14 and 24 use the default SR, while scenario 29 and 30 use Eqsim SR. Recruitment larger than 250 billion have been truncated in the plot.

5.6 Summary of results

5.6.1 Default escapement strategy

The ICES TAC advice for Norway pout is based on an escapement strategy with no upper limit on TAC and no F_{cap}. This HCR is not precautionary (as tested here in scenarios that were intended to be unconstrained by $F_{historical}$), because the probability of quarter 4 SSB being below B_{lim} is around 10%. Even when $F_{historical}$ was set to 5.0 (an arbitrarily high value of F), this value was reached in 4.7% of the replicates (row atFhist in Table 5.2). Such high F would probably be capped if it was observed in the ICES advice, but the MSE used the high value. The results show that ICES should consider modifying the default escapement strategy either by setting an upper F (F_{cap}) or including conditions on TACmin/ TACmax as explored above, or mixing both approaches, in order to make the escapement strategy precautionary.

5.6.2 Request 1

Three HCR were identified as precautionary from the set of HCR considered, with combinations of lower bound of TAC (TACmin) in the range 20 000 tonnes to

40 000 tonnes and with an upper TAC (TACmax) in the range 150 000 tonnes to 200 000 tonnes. Table 5.1 shows the main performance statistics for these HCR that were found to be precautionary. In these HCR, TAC is set at TACmin in 20–25% of the years while TACmax is reached in 36–46% of the years. There is a less than 6% difference in median TAC from the three HCR, while the mean TAC differs by up to 19%. The variation in TAC from one year to the next is smallest for the lowest TACmax. It is assumed that the true F will not exceed a value of 0.89 ($F_{historical}$), which means that the full TAC will not be taken if F exceeds $F_{historical}$. $F_{historical}$ is reached in 9–16% of the years, which makes the results sensitive to the assumption that the fishery stops catching Norway pout when F exceeds $F_{historical}$.

Table 5.8. Summary statistics (2023–2037) for precautionary HCRs with application of TACmin and TACmax, but no Fcap (Fcap at 2.0 is considered as practically no Fcap).

F_{Cap}	TAC-min (t)	TAC-max (t)	P(SSB < B_{lim}) (%)	TAC median (t)	TAC mean (t)	TAC change (t)	At $F_{historical}$ (%)	At TAC-min (%)	At TAC-max (%)
2.0	20000	150000	3.74	130054	99597	44950	9.1	20.8	46.0
2.0	20000	200000	3.94	123299	118119	64608	16.4	21.7	36.4
2.0	30000	150000	4.86	128023	101179	41826	9.6	24.6	45.7

5.6.3 Request 2

More HCRs with combinations of TACmin and TACmax become precautionary when they are combined with an F_{cap} in the range 0.3 to 0.4 (Table 5.10). The probability of setting a TAC at TACmin is similar to the probability for HCRs without an F_{cap}, but the probability of reaching TACmax is considerably lower due to the application of F_{cap}. In these HCR, TACs become considerably lower when F_{cap} is applied. The absolute changes in TAC between years are smaller when F_{cap} is applied, partly because of the lower TAC in general. Applying F_{cap} makes the HCR robust to the assumption of an $F_{historical}$, as the probabilities of reaching $F_{historical}$ become significantly lower (max 3.1%) than for the HCR without an F_{cap} (max $F_{historical}$ at 16.4%).

Table 5.9. Summary statistics (2023–2037) for precautionary HCRs with application of Fcap, TAC-min and TACmax.

F_{Cap}	TACmin (t)	TACmax (t)	P(SSB < B_{lim}) (%)	TAC Median (t)	TAC mean (t)	TAC change (t)	At $F_{historical}$ (%)	At TACmin (%)	At TACmax (%)
0.3	20000	150000	3.17	72265	77453	39497	0.7	19.1	14.3
0.4	20000	150000	3.55	89742	87686	42865	2.2	20.1	23.4
0.3	20000	200000	3.19	71968	81944	47072	0.9	19.2	6.7
0.4	20000	200000	3.61	88465	95345	54578	3.1	20.4	12.4
0.3	30000	150000	4.14	71706	79107	36744	1.1	22.9	14.2
0.4	30000	150000	4.55	89236	89391	40005	2.7	23.8	23.2
0.3	30000	200000	4.17	71367	83574	44260	1.4	23.1	6.6
0.4	30000	200000	4.67	88057	97107	51715	3.6	24.2	12.3

5.6.4 Comparison of results from Request 1 and 2

Comparison of the performance of the different HCR tested to answer Requests 1 and 2 in terms of catches and risks can be seen in Figure 5.24. The major differences on the TAC levels are due to F_{cap}, whereby the lower the F_{cap} level the lower the TAC. On the other hand the probability of falling below B_{lim} is determined by TACmin, being mediated by F_{cap}. The higher the TACmin the greater the risk and this is increased as F_{cap} increases. The application of a TACmin itself has a very limited effect on the median TAC.

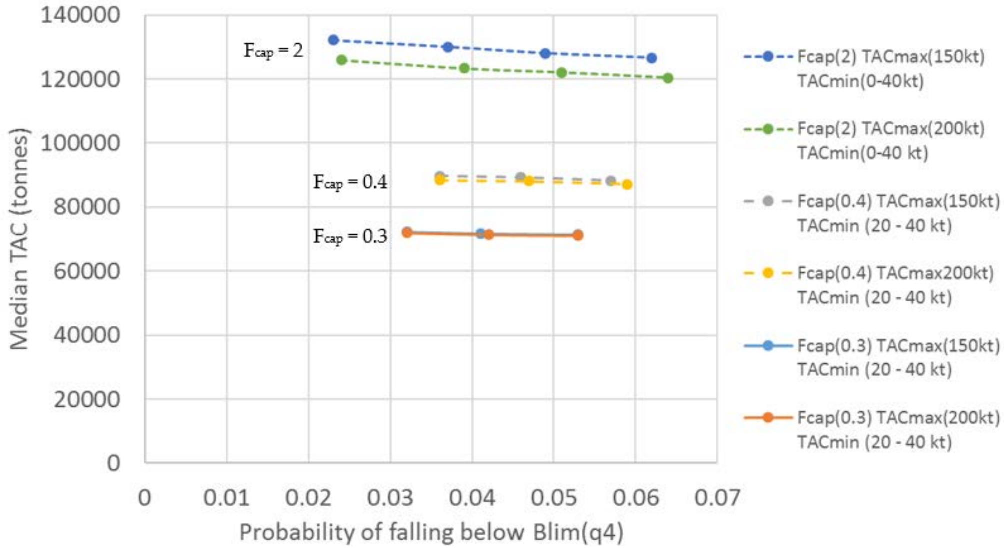


Figure 5.24 : TAC (Median TAC) versus probability of SSB below B_{lim} for different levels of F_{cap} (0.3, 0.4 and 2) and levels of TACmax (150 or 200 kt). Dots in lines from left to right refer to the increasing levels of TACmin (either from to 0 to 40kt for the F_{cap}(2) level or from 20 to 40kt for the other F_{cap} values).

5.6.5 Request 3

Request 1 and 2 provide HCR which includes a TACmin of up to 30 000 t. As an example of an HCR with a TACmin at 40 000 tonnes, the escapement SSB (the minimum SSB left after the TAC has been taken) was raised from B_{lim} (39 450 t) to 65 000 t. Three scenarios were found to be precautionary (Table 5.11) for HCR with combinations of F_{cap} in the range 0.3 to 0.4 and TACmax in the range 150 000 to 200 000 tonnes. TACmin is used for TAC in around 48% of the years, which gives a median TAC slightly above TACmin. Mean TAC for the three HCR is in the same order of size as for Request 2.

Table 5.10 Summary statistics (2023–2037) for precautionary HCRs with application of Fcap and TACmax in combination with a $B_{escapement}$ at 65 000 tonnes

F_{Cap}	TACmin (t)	TACmax (t)	P(SSB < B_{lim}) (%)	TAC median (t)	TAC mean (t)	TAC change (t)	At $F_{historical}$ (%)	At TACmin (%)	At TACmax (%)
0.3	40000	150000	4.87	47753	75843	34013	2.1	47.9	14.4
0.3	40000	200000	4.89	46278	80323	41546	2.4	48.3	6.6
0.4	40000	150000	4.95	46387	80923	37221	3.0	48.4	22.4

Consideration of other HCR including clauses for overriding the TACmin at low stock levels were not fully explored by WKNPOUT due to lack of time.

5.6.6 Sensitivity tests

The sensitivity of the choice of stock recruitment relationship was tested for two HCR (Table 5.12) using the Eqsim model as an alternative model. The performance statistics are robust to the choice of stock recruitment model with just a small difference in performance for the two approaches. The probability of SSB being below B_{lim} increases marginally with the use of the Eqsim model such that one of the two HCR gets a probability just slightly above 5%.

Table 5.11. Sensitivity test, HCR with application of the default and Equsim stock recruitment model.

F_{Cap}	TACmin (t)	TACmax (t)	P(SSB < B_{lim}) (%)	TAC Median (t)	TAC mean (t)	TAC change (t)	At $F_{historical}$ (%)	At TACmin (%)	At TACmax (%)
0.3	20000	150000	3.17	72265	77453	39497	0.7	19.1	14.3
0.3*	20000	150000	3.27	70236	76408	38972	0.7	18.7	13.7
0.4	30000	200000	4.67	88057	97107	51715	3.6	24.2	12.3
0.4*	30000	200000	5.95	83326	93837	49854	4.2	25.6	11.2

***applying the Eqsim model**

The MSE also tested the sensitivity to the uncertainty in the starting true states. By default, each of the HCR was evaluated using one thousand initial populations drawn from the SESAM assessment and projected for 20 years. These 1000 draws included uncertainty in the starting stock numbers, exploitation pattern, and recruitment. As a sensitivity test, the MSE was done again using one initial population, one exploitation pattern, and one set of past recruitment (the median estimate from SESAM) was projected 1000 times. Performance statistics like TAC and SSB for individual scenarios differ by generally less than 2–3% between methods using one initial population versus one thousand populations. The probability of SSB being below B_{lim} is typically 1–2 percentage points higher when the one thousand population approach is used. This means that some of the scenarios considered non-precautionary with all the uncertainties got a lower than 5% probability of SSB being below B_{lim} when the one populations approach was applied. We consider that the one thousand initial population corresponding to the uncertainty reported by SESAM better reflects the uncertainties in the assessment and this approach was chosen as the final method for MSE of this stock.

6 Effect of changing the TAC year

The special request asked ICES to evaluate whether the results of the MSE would be significantly changed if the TAC year were defined as 1 November to 31 October rather than a calendar year. The request is a bit confusing, because the current management set the TAC for the EU Member States fishing in EU waters from 1 November to 31 October, while Norway uses the calendar year (Jan–Dec). Furthermore, the advice is produced based on a forecast from 1 October to 30 September, however ICES considers that the forecast for quarter 4 to quarter 3 can be used directly for management purposes for the period 1 November–31 October (ICES advice on Norway pout, 2017).

In the response of June 2017 from ICES to the clients, it was stated that a TAC year following the calendar year will not be tested, as a proper MSE with a TAC year following the calendar year would also require a time shift in the assessment and forecast. There is no reviewed assessment available with such a time shift and therefore the shift in TAC period could not be tested. As no reply was received at ICES so far, it is understood that this was not a high priority in the request. Therefore, the answer provided to the request is based on directly testing a TAC set for the period 1 November to 31 October, based on an ICES assessment including catches for the quarter 1–3 of the terminal year.

The WK addressed briefly the effect of the different TAC periods between Norway and the EU, given the present timing of the ICES advice. The obvious implication of such a mismatch is that, if a strong reduction of the TAC were required, the implementation of such a reduction would take place a month later by EU fleets and about 3 months later by the Norwegian fleet. Notice that the fishing season for Norway pout in the Norwegian economic zone (NEZ) is from 1 April to 31 October and thus, fishing Norway pout in Norwegian waters is forbidden in November and December. In some years, the EU trades quotas of Norway pout against Norwegian quotas of other species. These quotas are caught by Norwegian fishermen in EU waters, mainly in q3 and q4. They can be caught until the end of the year even though it is at the expense of the TAC having finished for EU vessels by 1 November.

To assess what implications might exist, we have used the mean catches by quarters and by countries in the recent period 2013–2016 (Table 6.1) representing the status quo of the fishery. From the table, it follows that about 1/3 of EU catches in quarter 4 (assumed to be taken in October) and 100% of Norwegian catches in quarter 4, will sum up to 30.7% of the international catches in recent years, being taken before the new TAC becomes in force. This is about 18 700 t for a TAC set at about 61 000 t (the mean of the recent catches).

Table 6.1 Mean absolute and relative catches by countries and quarters in the period 2013–2016.

2013–2016	**Mean Catches**	**Mean Catches**	
Absolute Catches	Denmark	Norway	TOTAL
1	554	217	771
2	385	7,075	7,460
3	5,659	16,459	22,117
4	17,853	12,757	30,610
TOTAL	24,451	36,507	60,958
2013–2016			
Relative Catches	Denmark	Norway	TOTAL
1	0.9%	0.4%	1.3%
2	0.6%	11.6%	12.2%
3	9.3%	27.0%	36.3%
4	29.3%	20.9%	50.2%
TOTAL	40.1%	59.9%	100.0%

As the advice is based on a projection from 1 October to 30 September, it means that there will be a mismatch between the TAC assigned to this period and the catch made in the TAC period 1 November to 31 October, because part of the catch in the forecast period for the advice is always taken at the expenses of the previous TAC. The mismatch in the period (practically q4, q1, q2, q3) for the ICES advice TAC and the TAC period applied 1 November to 31 October will increase the risk to B_{lim} particularly in the cases when the TAC is reduced compared to the previous year. For the last three years, the sum of 1/3 of EU catches in q4 (assumed to be taken in October) and the 100% of Norwegian catches in q4 constitutes 30.7% of the annual catches. The significant percentage of catches in q4 indicates a problem which needs further investigation, if the present management system continues.

The present timing of assessment and TAC period was chosen as it allow the use of recruitment indices from the q3 surveys, and because the total catches in q3 have been low, such that the guesses on the total catch in q3 from observed catches in July–August of the terminal year is robust.

If, as result of the current MSE, a bounded harvest strategy based on a escapement strategy with a TACmin is adopted, a simple way to overcome this inconsistency would be equalling the TAC year to the catch forecast period used for the advice; both being applied over the period 1 October to 30 September. The issue of making the assessment and producing the advice in October will not be a problem, because for a harvest strategy with TACmin the fishery is never closed, and therefore the fishery can always start at TACmin and once known the advice for the new TAC, at the end of October, the TAC would be updated accordingly. In practice this would imply that the TACs would be enforced for the international catches over q4(Y)–q3(Y+1).

7 Discussion

The Norway pout assessment uses the newly developed SESAM model, which is a stochastic state space model with quarterly time steps. To actually mimic the SESAM model and forecast, it was necessary to develop new software before and during the WKNPOUT meeting. The end result is considered as a well-tested software, however the process would have been more optimal if suitable MSE software was fully tested before the WK.

The way to set up the Operating Model (OM) in terms of the starting inputs and the way to condition the simulations was debated since the beginning of WKNPOUT. Two major issues focused discussion throughout the process, a) the initial true states of the population and exploitation patterns and b) the maximum realizable fishing mortality the fishery can exert.

7.1 Choice of initial true states of the population, recruitment and exploitation patterns

Regarding the initial true states of the population and exploitation patterns, the WKNPOUT selected at the beginning to mimic the full uncertainty provided by the stochastic SESAM model (in ICES, 2017). This includes the uncertainty on initial stock numbers at age, exploitation pattern and recruitment values (for simplicity referred to as the 1000 initial population approach). The first approach was to generate 1000 replicates of the true states to be used as inputs for the 1000 simulations in the operating model of the MSE. An alternative was proposed during the WKNPOUT meeting because the preliminary results showed a high probability of SSB being below B_{lim}, even in cases with no fishery. This alternative approach omitted most of the uncertainty as the initial stock numbers, recruitment and the exploitation pattern were all equal to the median estimate from the last assessment (for simplicity referred to as the one initial population approach). This approach was tested further during the WKNPOUT, also after it was realised that the high probability of SSB being below B_{lim} in the initial method was due to a non-standard setting of the hockey stick recruitment model and not an effect of the initial population.

After the WKNPOUT meeting (once the hockey stick SR was set in the standard form), a sensitivity analysis showed that the 1000 initial population gave similar results as the one initial population approach, but that the 1000 initial population approach gave a slightly higher probability of SSB being below B_{lim}. Because the 1000 initial population approach was the initial model and because this approach best reflected the uncertainties in the SESAM assessment, the 1000 initial population approach was selected as the final and default method for MSE. This was however not a consensus decision, but mainly supported by the chairman for the WK and the scientists developing the MSE software.

The final choice of method was supported by the following observations:

a) The short term probability of SSB being below B_{lim} was not affected by the choice of the number of initial population. The long term probability of falling below B_{lim} (2023–2037) for any particular replicate seems also independent of the initial population number (Figure 7.1) which indicates that it is not the uncertainty in the initial stock numbers that makes the difference in the MSE performance.

b) Exploitation pattern has been highly variable in recent years. This is best reflected by the 1000 initial population approach, which by its draw of 1000 sets of exploitation pattern from F in the terminal year of the assessment avoid the assumption that exploitation pattern is exactly F in the terminal year of the assessment for all the replicates. Figure 7.2 shows an example of the uncertainty in the quarterly exploitation pattern used by the 1000 populations approach.

With the one initial population approach, future recruitments (for SSB above 72 101 t) are drawn from a pool of the 37 median recruitment estimated by the SESAM assessment. With the 1000 initial population approach, future recruitments are drawn from 1000 sets of recruitment drawn from the distribution of the 37 historical recruitments. This approach seems more appropriate to also reflect the uncertainties of the historical recruitment (see also section 4.5).

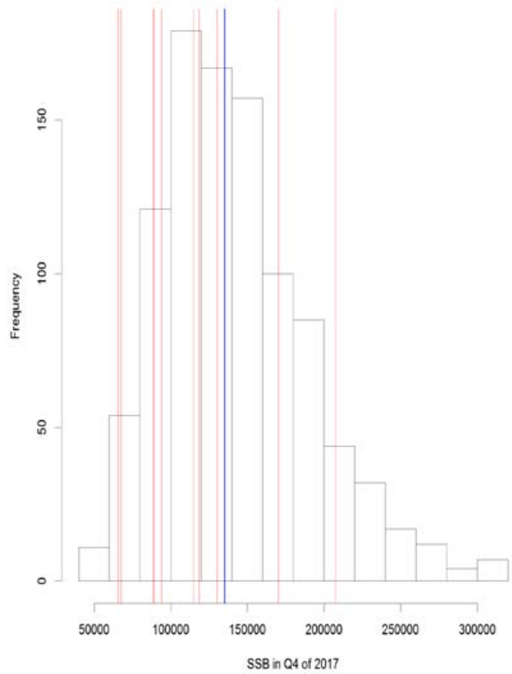


Figure 7.1. Distribution of SSB (Q4) in the first year of the simulation (default escapement, scenario 1). Blue line shows the median SSB. Red lines show SSB from the 10 iterations with the highest probability of SSB below B_{lim} in the period 2023–2037.

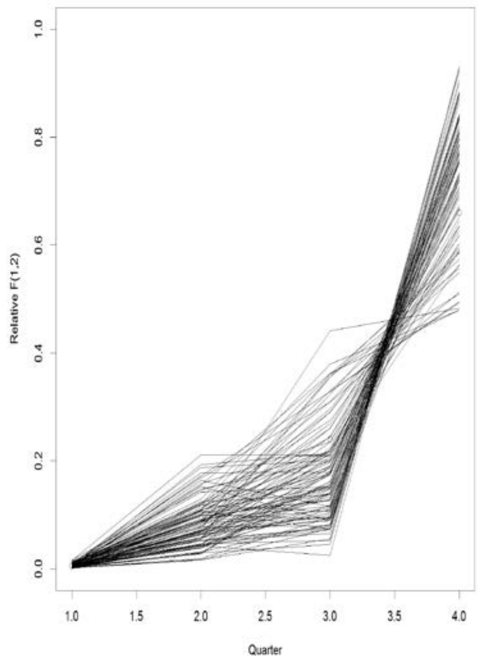


Figure 7.2. Relative F by quarter from 100 populations. The relative F is the sum of F at age 1 and 2 by quarter divided by the sum over all quarters.

7.2 Realism in the MSE regarding realised fishing mortality

There was a long discussion before and during the WKNPOUT meeting about the need for an upper value of true F ($F_{historical}$) used when the OM stock numbers are reduced by the TAC provided. One assumption could be that the TAC is always taken even though it would require a very high F. Another approach is the assumption of an "implementation error" where it is assumed that the fishery stops when F (and fishing effort) exceeds a maximum F observed in recent years. Previous MSE for this stock (ICES CM 2013, ACOM:66) has shown that the MSE results may be highly sensitive to the choice of $F_{historical}$, especially when the effect of a minimum TAC is evaluated.

It was argued that $F_{historical}$ should be quite a high value (e.g. 2) as an MSE should also test "worse case" scenarios where F might be higher than observed most recently. Others argued that the fishery has changed significantly (e.g. due to a more strict enforcement of by-catch regulations) over the last 20 years and that $F_{historical}$ should reflect the present state and not exceed the maximum values of assessment F (0.37, median value) during the last 20 years. WKNPOUT agreed on an $F_{historical}$ at 0.89, which is the 97.5th percentile of the highest estimate of F in the last 20 years (F in 2013 estimated by the 2017 assessment). Whether this value is the most appropriate is impossible to judge, however the following observations support an $F_{historical}$ which is not very high:

- Norway pout is a bottom resource, not a pelagic schooling resource, where CPUE declines by declining stock size. Given the low price (landed for production of fishmeal and oil), the economic outcome of the fishery will become too low to continue fishing at very low stock size.

- The stock area is wide and only part of the stock area is fished. "The Norway pout box" which is an area closed for Norway pout fishery will also provide some protection against a potential high F.
- The number of vessels and overall fishing capacity have been decreasing and the remaining fleets have often better alternatives (a better economy) than fishing for Norway pout, so that overall fishing effort affecting Norway pout has also decreased. A lower fishing effort for Norway pout is supported by the low F and the missing quota uptake in recent years.

To monitor the effect of $F_{historical}$, a performance statistic was created reporting the frequency the $F_{historical}$ was reached in the simulations (called atFhist). This statistic shows the frequency of times the TAC is not fully taken as it would require an F above $F_{historical}$. A few HCR with no F_{cap} are considered precautionary even though $F_{historical}$ is constraining the removed (fished) biomass in up to 18% of the years. It might be argued that in these cases the evaluation is too dependent on the value of $F_{historical}$ and that an e.g. F_{cap} should be applied.

On the other hand it might be unrealistic to assume that the catches in the future will be more than twice as big as catches in the last 20 years (see e.g. Figure 5.16). However, the WG preferred testing the HCR for the impacts the allowable level of catches would have on the development of the stock, provided they are taken (for realized F below some credible maximum $F_{historical}$), than testing the rules allowing many times the TACs are not taken (as a results of a low $F_{historical}$) (in a kind of implementation error).

7.3 Other points

All the HCR presented as precautionary in the long term (2023–2037) were also considered precautionary in the short term (2018–2022). The probability of SSB being below B_{lim} was less than 5% in all of the individual years of the period 2018–2022 for all the HCR.

The escapement strategy is aimed at achieving a high probability (95%) of having the minimum amount of spawning biomass required to produce MSY (B_{lim}) left to spawn the following year. This means that $B_{escapement}$ would be normally defined as the SSB at spawning time, which is in quarter 1 for Norway pout. However, for this stock ICES has chosen (Benchmark report) to use the beginning of Quarter 4 as the time to evaluate SSB against $B_{escapement}$, as this period aligns better with the TAC year (1 October–30 September). All the precautionary HCRs had a lower than 5% probability of SSB below 72 101 tonnes in quarter 1. The 72 101 tonnes is the lowest observed SSB in quarter 1 from the benchmark assessment (ICES, 2016) and this SSB is considered as the minimum value of SSB above which the probability of impaired recruitment is expected to be low. With a less than 5% probability of quarter 1 SSB below 72 101 tonnes, it is shown that the criteria of quarter 4 SSB above B_{lim} is sufficient to ensure a low risk of impaired recruitment.

8 Stakeholders opinions

This section contains stakeholder input to Management Strategy Evaluations (MSE) carried out during WKNPOUT. Although focus is on Norway pout most comments will apply more general to MSEs and how they are being carried out in ICES. The background for including this section in the report, is that the purpose and setup of MSE's was debated several times during the WKNPOUT meeting. A debate which the stakeholders would like to be reflected in the final report.

It is acknowledged that it is easier to come up with critic of such a complex process as a MSE, than it is to come up with alternative solutions that will work, both in theory and practice. The following section should be read in the light of this understanding.

General comments

Some of the participants at the WKNPOUT workshop argued that the overarching aim of an MSE is sometimes forgotten. The purpose of a MSE is to model and compare different strategies and how robust they are to input assumptions and data. The MSE should eventually enable managers to make informed decisions, which will finally lead to a better management of the fish resource, in this case Norway pout. However, the process as it is carried out now, leaves the impression that the MSEs are predominately used, not to evaluate and compare different strategies and scenarios, but mainly to estimate parameters (e.g. Ftarget, trigger points, minimum TACs), while still ensuring sustainability ($p(SSB > B_{lim}, 0.05)$). This means that the management strategy is evaluated without much consideration how a given fishery (or managers) response to the strategies evaluated. A type of error often referred to as implementation error. The responsibility for this is often the result of the formulation in the request. Basically, these two different approaches can be described as a qualitative approach, where different scenarios and strategies are compared, and a more quantitative approach where values are set such that the sustainability criteria are met.

During the WK one concrete example of where the two different MSE ideas crashes was the use of the parameter, called $F_{historical}$. $F_{historical}$ is the upper limit of fishing mortality that the stock can be fished with and is used as a parameter in the MSE simulation. What $F_{historical}$ should be set to, lead to some discussion. Some participants at the meeting formulated the argument that this value should be set such that the resulting parameters estimates (TACmax, TACmin) was estimated with high probability of being precautionary. A risk-aversive approach, which reasonably leads to that $F_{historical}$ should be set to the upper 97.5% CI for estimated F for a predefined period. Others argued that the upper CI might be risk-aversive, but it does not represent a realistic upper F that the fleet can fish with. Selecting a more "realistic" value should be done, only looking at the very latest years only, due to changes in the fleet, and further the $F_{historical}$ selected should also be the estimated F value and not the upper 97.5% CI. This approach might not be risk-aversive, but it is likely probably more realistic. Thus, the positions were between running a risk-aversive MSE where the estimated parameters were estimated with high probability of being precautionary, but also a MSE run which did not reflect the "real" world and hence a MSE run where the interpretation of evaluations parameters such as, the average inter-annual TAC variation, the average yield and the average fishing mortality, are difficult. The other approach would probably lead to a more realistic set of evaluation parameters which would facilitate the evaluation of the strategy, however the Fmax and Fmin estimated from this approach might not be set sufficient precautionary. There are pros and cons of both approaches which was discussed and acknowledge by the group, and hence it was decided to include this discussion.

Concrete input to Management Plan evaluation

Norway pout is distributed over a wide area, while the fisheries are constricted to a relatively small area due to bottom topography and area regulations. This was recognized by the WK but not directly implemented in the modelling. Indirectly, some of this is covered by the $F_{historical}$, discussed above, but the stakeholders think that future MSEs should include some availability estimates.

Stakeholders at the meeting identified the preferred options as combination of a max quota of 200 000 tonnes and a minimum quota of 30 000 tonnes. Although not needed in order to ensure precautionarity the stakeholders at the meeting supported combining the preferred combination with an F_{cap} in order to 1) get more stable TAC advice and 2) to increase years where fishing the advice TAC will result in F values that are below F observed in recent history.

9 References

ICES WKGMSE REPORT 2013, Report of the Workshop on Guidelines for Management Strategy Evaluations (WKGMSE), 21 - 23 January 2013

ICES. 2016a. Report of the Benchmark Workshop on Norway Pout (*Trisopterus esmarkii*) in Subarea 4 and Division 3.a (North Sea, Skagerrak, and Kattegat), 23–25 August 2016, Copenhagen, Denmark. ICES CM 2016/ACOM:35. 396 pp.

ICES. 2017. Report of the Working Group on Assessment of Demersal Stocks in the North Sea and Skagerrak (2017), 26 April–5 May 2017, ICES HQ. ICES CM 2017/ACOM:21. 1234 pp.

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Mathematica. 30 (1): 175–193.

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

Punt, A. E., Butterworth, D. S., Moor, C. L., De Oliveira, J. A. and Haddon, M. (2016), Management strategy evaluation: best practices. Fish Fish, 17: 303-334. doi:10.1111/faf.12104

10 List of presentations

General Introduction. A. Uriarte

Introduction to SESAM - Mollie Brooks

Methods of the MSE including the stochastic forecast - Mollie Brooks

simple HCRs. Daniel Howell

Other graphing options for presentation of results - Mollie Brooks

Details of the simulations with and without uncertainty in N, E, and R - Mollie Brooks

Reference points and stock-recruitment relation. Morten Vinther

Introduction to MSE runs and results. Morten Vinther

Several interim presentations during WKPOUT on advances of simulation results by Morten Vinther and Mollie Brooks (TACmin TACmax; No fishing; F_{cap} Run2; InitPop; day3_no1; day3_no2).

Methods again. Mollie Brooks (a presentation made for the a Webex meeting after the WKNPout).

Annex 1: List of participants

Name	Institute	Country (of institute)	Email
Andrés Uriarte (Chair)	AZTI-Tecnalia	Spain	auriarte@azti.es
Cecilie Kvamme	Institute of Marine Research	Norway	cecilie.kvamme@hi.no
Claus Reedtz Sparrevohn	Danish Pelagic Producers' Organisation	Denmark	crs@pelagisk.dk
Daniel Howell	Institute of Marine Research	Norway	daniel.howell@hi.no
Gjert Endre Dingsør	Norwegian Fishermen's Association	Norway	gjert@fiskebat.no
Helge Viksåland	Norwegian Fishermen's Association	Norway	helgeviks@gmail.com
Henrik S. Lund	Danish Fishermen's Association	Denmark	hl@dkfisk.dk
Mollie Elizabeth Brooks	DTU Aqua -National Institute of Aquatic Resources	Denmark	molbr@aqua.dtu.dk
Morten Vinther	DTU Aqua -National Institute of Aquatic Resources	Denmark	mv@aqua.dtu.dk
Søren Anker Pedersen	EUfishmeal	Denmark	sap@maring.org
Espen Johnsen (part time, on line)	Institute of Marine Research	Norway	espen.johnsen@hi.no
Sarah Millar	ICES Secretariat	Denmark	sarah-louise.millar@ices.dk
David Miller	ICES Secretariat	Denmark	david.miller@ices.dk

Annex 2: Agenda

The Initial tentative agenda was…

	Monday 26	Tuesday 27	Wednesday 28
9:00		Start: Structure of the report and proposal for the index of contents	Going through results of simulation HCR for TOR1, 2 & 3
9:30	Start; welcome(presentation of attendees); logistics in ICES Organising our Agenda of work introduction to the request and TORS	Going through results of simulation HCR for TOR1 Expanding simulations to cover TOR 2 and TOR 3	Going through results of simulation HCR for TOR1, 2 & 3
10:30	Overview of the progress of works achieved in advance and related presentations and discussions: Operated Model (Proyection Model & uncertainties covered) Management Procedure: Observation model and Harvest Control Rules HCR (range of params) + Inputs from stakeholders and rational flow of texting the HCRs to address the TOR Performance indicators	Writting of first texts and discussing any pending issues	Advancing texts Going through some sections of the report
13:00	Lunch	Lunch	Lunch
14:00	Others (while running the first testing of HCR) on TOR1	Going through results of simulation HCR for TOR1, 2 & 3	Wrapping up and definition of Next steps
15:00	sensitivity analysis to uncertain configurations of the OM Mismatch between TAC timing and advice timing?	Writting of first texts and discussing any pending issues	End of the Meeting
19:00	END of day	END of day	

The final agenda followed during the meeting was…

	Monday 26	Tuesday 27	Wednesday 28
9:00		Individual working	Individual working
9:30	Start; welcome(presentation of attendees); logistics in ICES Organising our Agenda of work introduction to the request and TORS	Going through results of simulation HCR for TOR1 & 2 Expanding simulations to better cover TOR 2	Going through results of simulation HCR for TOR1, 2
10:30	Overview of the progress of works achieved in advance and related presentations and discussions: Operated Model (Proyection Model & uncertainties covered) Management Procedure: Observation model and Harvest Control Rules HCR (range of params) + Inputs from stakeholders and rational flow of texting the HCRs to address the TOR Performance indicators	Start: Structure of the report and proposal for the index of contents Discussing the Operating Model for the Hockey stick relationship Discussing how to address TOR 3 and discussing any pending issues	Going through results of simulation HCR for TOR1, 2 Revisiting the index of contents including a a section for Stake holders Going through sections 1.2 interpretation of the request Discussing how to address TOR 3
13:00	Lunch	Lunch	Lunch
14:00	Others (while running the first testing of HCR) on TOR1	Going through results of simulation HCR for TOR1, 2	Wrapping up and definition of Next steps
15:00	sensitivity analysis to uncertain configurations of the OM	Mismatch between TAC timing and advice timing? and discussing any pending issues	End of the Meeting
19:00	END of day	END of day	

Annex 3: Annex to the ICES Response letter to EU on the request on Norway pout on 29 June 2017

Original request text:

The European Union and Norway jointly request ICES to advise on the management of Norway Pout in ICES Subarea IV (North Sea) and ICES Division IIIa (Skagerrak-Kattegat). The proposed management strategy is based on the ICES escapement strategy for Norway pout with the aim of achieving a high probability of having the minimum SSB required to produce MSY (Blim) surviving to the following year.

ICES is requested to evaluate:

1. *Whether a management strategy is precautionary if the TAC is constrained with a lower bound in the range of 20,000 tonnes to 40,000 tonnes and an upper bound in the range of 150,000 tonnes to 250,000 tonnes, or another range suggested by ICES.*
2. *Whether such a strategy would be precautionary if the TAC constraints referred to in paragraph 1 are overridden by a constraint on the maximum value of fishing mortality (Fcap), and whether the application of the Fcap would allow a precautionary strategy with a higher minimum TAC than if the Fcap was not applied.*
3. *Whether a provision to override the minimum value of the TAC when the stock is forecast to be below some threshold value would allow a precautionary strategy with a higher minimum TAC than if the escape-clause was not included, and whether such a provision would provide any additional benefit to the inclusion of an Fcap as referred to in paragraph 2.*

ICES is requested to indicate the results of the evaluation in a table that shows for the combination of parameter values selected for the evaluation:

- *The average inter-annual TAC variation*
- *The average yield*
- *The average fishing mortality*
- *The average escapement biomass*
- *The probability that the stock falls below Blim in the year following the fishing year over a 20 year period.*

ICES is additionally asked to indicate whether the results of the evaluation are significantly changed if the TAC year is defined as 1 November to 31 October rather than a calendar year.

Comments and proposed approach by ICES

- The proposed management strategy evaluation includes many options which, in combination, could result in a very large number of scenarios and corresponding evaluations resulting in many of these evaluations subsequently found to be unnecessary. In order to limit the number of options, competent stakeholders will be invited to actively participate in the process (ICES workshop), such that the number of options can be reduced concurrently in agreement with stakeholders. Prior to the suggested workshop, a list of possible options will be provided to Clients and relevant stakeholders for comments; this will narrow down the field of options for the actual workshop to focus on.

- We note that Fcap is an additional measure introduced by ICES to avoid overfishing within an escapement strategy where Bpa is used as target to avoid biomasses below B_{lim}. If the risk of falling below B_{lim} can be quantified directly, as it is in the currently applied Seasonal SAM assessment model (SESAM), Fcap is in theory not needed as an additional management measure to avoid overfishing in relation to B_{lim}. However, Fcap may act to stabilise long-term yield, and increase the biomass in the sea. This can also be obtained by using a maximum TAC. The effect of using this method or the previous method on average yield etc. will be discussed in detail in WKMSYREF5. Therefore, given the above, further discussions with Clients on point 2 of the request in relation to Fcap is considered useful.

Point 1 in the request:

Experts would like clarification as to whether the intention is to keep the lower bound fixed with the upper bound being something different in order to meet the precautionary criteria?

In case no prioritisation from clients is available, experts will need to limit the options space, which could lead to the following approach and interpretation as to the required analyses:

1) With use of the escapement strategy estimate SSB, yield and risk to B_{lim} as function of combinations of maximum TAC in the range 150–250 kt (step 50 kt) and minimum TAC in the range 20-40 kt (step 10 kt)
2) Identify combinations of the minimum and maximum TAC that provide < 5% probability (risk) of SSB below B_{lim}.
3) Select options for further analysis that represent the edges of the "area" of precautionary combinations. That is the options
    a. lowest minimum TAC and lowest maximum TAC
    b. lowest minimum TAC and highest maximum TAC
    c. highest minimum TAC and lowest maximum TAC
    d. highest minimum TAC and highest maximum TAC

Point 2 in the request:

1) For options a-d separately estimate SSB, yield and risk to B_{lim} as function of Fcap in the range 0.3–0.4 (step 0.05). This range of F covers the upper range of estimated yearly F's since 1995.
2) Quantify the effect of raising the minimum TAC on SSB, yield and risk to B_{lim}, for selected combinations of options a–d and Fcap values that produces the lowest risk to B_{lim}.

Point 3 in the request:

1) This request is not clear, but it is understood that a minimum TAC should not be applied in cases where a very low stock size is estimated. ICES assumes that

the stock threshold value referred to in request 3 is Total Stock Biomass (TSB) at the beginning of the TAC year. For now, ICES has no suggestion for such a value.

2) It is not possible to specify further which combinations of threshold values and options from request 1 and 2 that will be relevant to investigate.

At the benchmark in 2016, it was decided to replace the Seasonal XSA model with the SESAM model. The SESAM model and reference points are reviewed in a configuration with an assessment that does not follow the calendar year but includes quarter 4 of year y and quarters 1–3 of year y+1. The advice given by ICES, for the period 1 November to 31 October, follows this timing of the assessment. It is requested to make the MSE for an advice following the calendar year, and to evaluate if the results would have been significantly changed if the TAC year is defined as 1 November to 31 October. Such a request cannot be fulfilled, as it would require a new review of an assessment and reference points based on an assessment following the calendar year. It would, however, be possible to base the MSE on the current timing and make a qualitative evaluation of the effects of having a TAC year that follows the calendar year. This however, it is not consistent with the previous requests to the WGNSSK on Norway pout advice and management.

Suggested process from ICES

The ICES "Workshop to review the ICES advisory framework for short lived species, including detailed exploration of the use of escapement strategies and forecast methods" (WKMSYREF5), 5–11 September 2017 provides an opportunity to develop and review new approaches relevant to this request. The timing of WKMSYREF5 and other tasks makes it impossible to complete the requested MSE on Norway pout before the ICES TAC advice on Norway pout is released in early October.

The following plan is suggested:

- September 2017: Method development and review in WKMSYREF5.
- Autumn/winter 2017: Development of software, defining and running preliminary scenarios
- February 2018 (but not in the period 10-18 February): Physical 3-day workshop at ICES to answer the request.
- April-May 2018: Review of report of physical meeting by WGNSSK.
- May 2018: ADG and release of request.

Annex 4: Further evaluation of the ICES escapement strategy for Norway pout.

The ICES escapement strategy for Norway pout, with no F_{cap} or TACmax, is not considered precautionary (see section 5.6.1).

This annex presents evaluations of escapement strategies using:

- F_{cap} in the range 0.3 to 1.0
- $F_{historical}$ at 5.0 (practically no upper limit on F)
- TACmax at 5 million tonnes (practically the same as having no TACmax)
- No TACmin

Performance statistics are presented in Table 12.1. The probability of SSB below B_{lim} is less than 5%, for F_{cap} at 0.70 or lower (Figure 12.1). The highest TAC is obtained for F_{cap} around 0.75 (Figure 12.2). The probability of a zero TAC (row atTACmin in Table 12.1) increases from 12% at F_{cap} = 0.3 to 16% for F_{cap} = 0.70.

Table 12.1. Performance statistics for Escapement strategies with F_{cap}. Shaded scenarios have more than 5% probability of SSB in Quarter 4 being below B_{lim} (row "risk1.long.Q4" > 5%). $F_{historical}$ is set at 5.0.

	scen 51	scen 52	scen 53	scen 44	scen 55	scen 56	scen 57	scen 58	scen 59	scen 60
F_{cap}	0.30	0.40	0.50	0.60	0.65	0.70	0.75	0.80	0.90	1.00
TACmin	0	0	0	0	0	0	0	0	0	0
TACmax	5000000	5000000	5000000	5000000	5000000	5000000	5000000	5000000	5000000	5000000
F_{bar}.median	0.285	0.359	0.420	0.464	0.481	0.493	0.501	0.506	0.511	0.513
F_{bar}.mean	0.296	0.381	0.459	0.529	0.561	0.591	0.620	0.647	0.696	0.738
SSB.median	129571	119561	112752	107247	105331	103282	101749	100267	97762	95728
SSB.mean	158399	146646	137325	130073	127162	124466	122146	120214	117009	114542
risk3.short.Q4	1.9	2.3	2.9	3.8	4.2	4.8	5.0	5.3	6.3	6.8
risk1.long.Q4	1.74	2.41	3.09	4.03	4.49	4.87	5.27	5.63	6.39	6.98
risk3.long.Q4	2.5	3.6	4.3	5.0	5.2	5.9	6.1	6.9	7.6	8.5
risk1.long.Q1	1.52	2.03	2.54	3.21	3.53	3.78	4.09	4.45	4.90	5.47
atFhist	0.0	0.0	0.0	0.1	0.1	0.2	0.3	0.4	0.6	0.9
TAC.median	72586	89291	100258	106754	108711	109479	109538	109404	108796	107904
TAC.mean	84033	102487	117329	128911	133978	138244	142081	145609	151660	156512
TRC.median	72586	89291	100258	106754	108711	109479	109538	109404	108796	107904
TAC.change	60130	78017	94713	109981	117281	123917	130079	135644	146109	155145
atTACmin*	12.1	13.1	13.8	14.8	15.3	15.5	15.9	16.0	16.3	16.6
atTACmax	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
YearToRecover.Q4	1.23	1.24	1.21	1.21	1.20	1.20	1.20	1.20	1.20	1.19
YearToRecover.Q1	1.63	1.67	1.70	1.73	1.73	1.75	1.75	1.75	1.75	1.75

***TAC at zero tonnes**

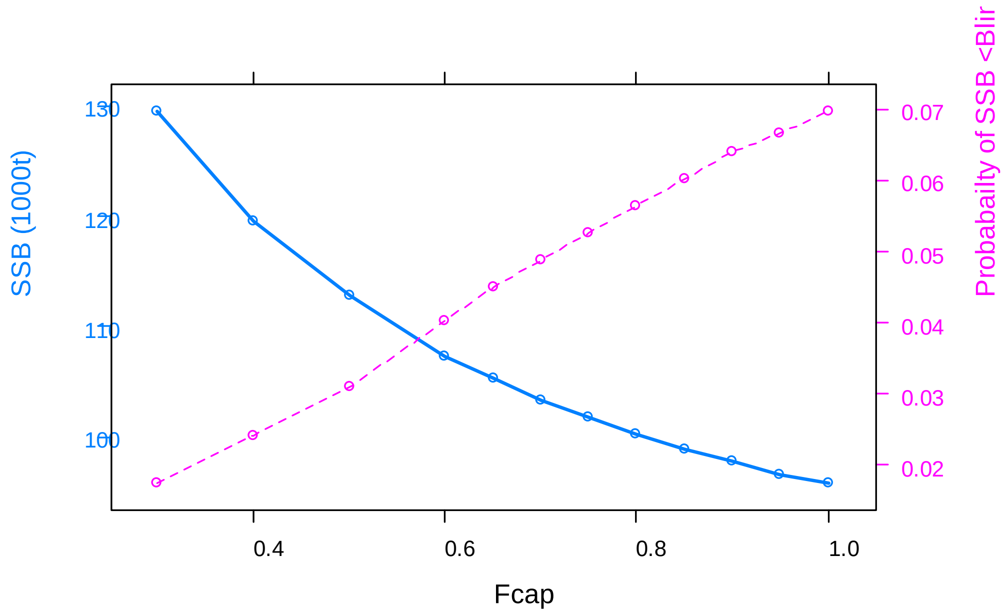


Figure 12.1. Median SSB in quarter 4 and probability of SSB below B_{lim} as function of F_{cap}.

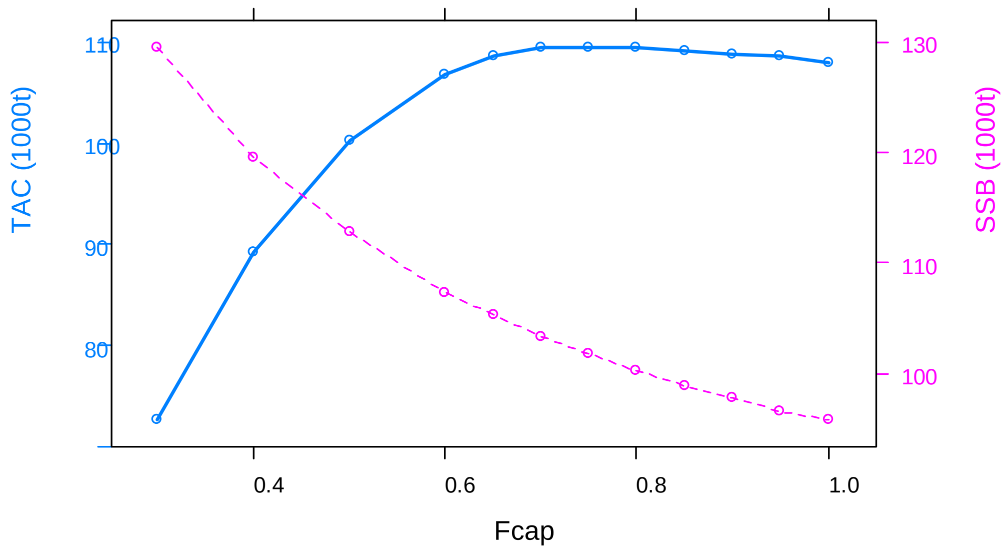


Figure 12.2. Median TAC and median SSB in quarter 4 as function of F_{cap}.

Annex 5: Reviewers comments

Review of management strategy evaluation for Norway Pout (WKNPOUT, 26–28 February 2018)

By: Manuela Azevedo (IPMA, Lisbon), 24 April 2018

This review examines the technical aspects of the MSE simulation testing carried out to evaluate several HCR for a management strategy for Norway Pout in Subarea 4 and Division 3.a(report of the Workshop for management strategy evaluation for Norway Pout (WKNPOUT) that took place 26–28 February 2018 at ICES Headquarters: downloaded from the sharepoint, version dated 06 April 2018).

The draft report is easy to read, the methods are described in a comprehensive way and the results are well documented. The number of possible scenarios to address requests 1 to 3 could be enormous and experts prepared in advance of the workshop a list of options that were discussed with stakeholders attending the meeting and to provide guidance on the workshop work. Thirty four HCRs were simulation tested with an MSE 'short-cut' approach (i.e.where the stock assessment is emulated by adding observation error to the 'true' population). Request 3 was addressed by simulation testing four scenarios with TACmin (40 000 tonnes), combinations of TACmax (150 000 and 200 000 tonnes)and F_{cap} (0.3 and 0.4) and evaluated based on a higher $SSB_{escapement}$, set at B_{pa} = 65 000 tonnes. Having reviewed the report I have no major issues regarding the technical aspects and I agree with the workshop conclusions. I provide comments below focusing on some particular aspects of the work.

- **Operating model**

The OM (simulation of the dynamics of the natural and fishery systems) is based on the seasonal (quarter of a year) and age (age groups 0–3+) structured population dynamics used in the Seasonal Stochastic Assessment Model (SESAM), following the Norway pout benchmark in 2016.The parameters uncertainty from the last stock assessment (starting population in the 3rd quarter of 2017and exploitation pattern) are used to randomly draw 1000 replicates of the 'true' population in the end of 2017 TAC year (start of the 2018 TAC year), which are projected over 20 years. The exploitation pattern is kept constant in each replicate simulation. The quarterly natural mortality rates, stock and catch weights-at-age and proportion mature-at-age are kept constant in each replicate throughout the 20 years of simulations. The OM conditioning, based on the best available scientific knowledge of the fleet behaviour and biological dynamics of the stock, is considered adequate for an MSE. I have, however, the following comments regarding the sensitivity test to recruitment:

Future recruitment (number-at-age 0) is assumed to occur in quarter 3 (Q3), dependent on the SSB from Q1. The S-R relationship is of type 5 (no apparent S-R signal and no evidence of impairment R, WKNPOUT benchmark 2016) and as such the approach adopted by the workshop to simulate future R was to assume a hockey-stick S-R relationship with a forced break-point at the lowest estimated SSB Q1, of 72 101 tonnes in 2005, such that: for SSB Q1 > 72 101 tonnes future R is simulated as a random sample from the SESAM recruitment time-series estimates with uncertainty; for SSB Q1 ≤ 72 101 tonnes future recruitment is governed by the steep line of the hockey-stick S-R relationship (breakpoint at SSB = 72 101 tonnes and R = 43 861 million; line slope of 0.6083325) and assuming lognormal truncated errors (truncation based on the distribution of historical standard normal deviations from the mean of log R). While the approach to simulate future R is considered adequate it would be helpful to plot examples of R from a few (2–3) simulated populations superimposed to the shaded area in the R plots (summary results plots: figures 5.1, 5.16, 5.18) to show how much jagged and variable R can be in individual populations during the simulation period and how they match the behaviour of the historical recruitment. Norway pout is a short-lived species and SSB is largely determined by incoming recruitment (20% mature-at-age 1, 100% mature-at-age 2) making its population dynamics very dependent on R variation. Despite no evidence of impairment R, the frequency of strong year classes have decreased since 1999. The median R in the 20 years simulation period (~43 861) is above the mean R since 1999 (33 683). Sensitivity tests on the influence of simulated R were carried out assuming in the OM a S-R relationship fitted to three types of S-R functional relationships (Ricker, B&Holt, hockey-stick) weighted by the 'Buckland' method (Section 4.5.2). What was the estimated weight for each S-R type? What was the estimated value of the hockey-stick break-point? The results from this sensitivity test show minor changes in the performance statistics but this could be due to the fact that the hockey-stick S-R was the dominant S-R type in the 'Buckland' fit, with break-point similar to the break-point assumed in the base-run simulations. If this is the case then the sensitivity test to recruitment does not add robustness to the proposed HCRs.

- **Implementation model**

The 'true' realized F when the OM stock numbers are reduced by the TAC provided was constrained by an upper value of 0.89, which is the 97.5 percentile of the stock assessment Fbar in the last 20 years. The rationale behind this assumption is the perception that fishing is not economically viable at low stock size, that only part of the stock area is exploited and that fishing capacity and fishing effort has decreased in recent years, which seems sensible reasons to me to support an Fhistorical = 0.89 in the implementation error approach. Results show the percent of years where the real catch is predicted to be lower than the TAC, allowing to monitor the effect of Fhistorical in the simulations.

- **Management strategy**

The MSE is a 'short-cut' approach since an assessment cycle with SESAM is not included in the management procedure/strategy component. Computing time for the MSE can be large and many scenarios were tested for Norway pout (~ 2 hours/HCR, running in parallel across cores!) so this might be the reason not to perform a 'full-feedback' MSE. The last assessment indicates a retrospective bias with overestimation of SSB(ICES WGNSSK Report, 2017, Figure 12.3.8). Although the retro point estimates of SSB are inside the upper confidence interval of SSB from the last assessment, it would be useful to explore the likely effect of this retrospective pattern on the computed risks to SSB (probability of SSB Q4 falling below B_{lim}) to add robustness to the MSE (e.g. running a full-feedback MSE for one of the scenarios showing risk to SSB close to 5%).

- **Performance statistics**

Two types of risk were computed to evaluate if the HCRs are precautionary. Risk type 3.long.Q4 (*maximum* probability that SSB-Q4 is below B_{lim} = 39 450 tonnes, where maximum is taken across 2023–2037) is always higher than risk type 1.long.Q4 (*average* probability that SSB-Q4 is below B_{lim}, where the average is taken across 2023–2037). For example, risk3.long.Q4 is above 5% for scenarios 4,5, 6 (request 1), 18,19,21 to 24 (request 2) and 41 to 43 (request 3) while risk1.long.Q4 is below or just below 5% in these scenarios. Since the simulations still show variable probabilities in the long-term (e.g Figure 5.2) and to support the assumptions made that it is caused by an insufficient number of replicates in the MSE (1000), performance statistics could be presented for one of the scenarios simulated with a higher number of iterations. Nevertheless, according to ICES (WKGMSE, 2013) risk 1 can be considered an adequate measure of risk level for short-lived species, as SSB in any one year is almost independent of the previous years. This can be considered the case of Norway pout with highly variable recruitment rapidly influencing SSB. Moreover, results are also presented for risk1 at spawning time (Q1) showing < 5% probability of SSB-Q1 below B_{lim}-Q1 (72 101 tonnes) for the scenarios that are precautionary under the criteria $SSB_{escapement}$ in Q4 above 39 450 tonnes, giving confidence to the workshop conclusions.

Minor comments to be considered for the final version of the report

Text:

- Sec 3.1.2, 2nd paragraph, last sentence: According to the assessment result, this Blim corresponds to ~~3947t~~ 39450 tonnes (Bloss = SSB value in quarter 4, 2005); 7th paragraph, last sentence: In all simulations .... below Blim of ~~3947~~t 39450 tonnes (Bloss = SSB value in quarter 4, 2005), ...assessed.
- 8th paragraph, 2nd sentence: i suggest to add 'SSB in Q4 of 2005 (Bloss) is now estimated somewhat lower than in the benchmark assessment *(30742 tonnes in the 2017 stock assessment vs 39447 tonnes in the benchmark assessment)*, while ...
- Sec 4.2.2.5: it would help to show the variance

Tables:

- Table 4.6: shows the median exploitation pattern but what about the variance?
- Table 4.7, risk3.short.Q4: the maximum risk in one of the years 2018-2022 ~~2018-2012~~ is used (ICES...);
- Table 5.1, legend: ..., while scenarios 41-44 ~~31-34~~ use $B_{escapement}$ at B_{pa}.
- Table 5.9, values of TAC change (tonnes): according to Table 5.3 should be44950, 64608,41826 (as per Table 5.3)
- Table 5.10, value of 'At Fhistorical (%)' for scenario Fcap=0.3, TACmin=30kt, TACmax=150kt, should be 1.1 (as per Table 5.4)

Figures:

- Sec 5.5.2: I suggest to add a plot showing the 'Buckland' methodwith the estimated weight and fit for each S-R type;
- Plots showing the probability of SSB-Q4 being below Blim by year and scenario (e.g Fig 5.2): suggest to add an horizontal line at 0.05;
- Fig. 5.10: panel for scenario 18 is missing.
- Fig 5.16: add to figure legend – (the shaded area in the catches scenario is defined by TACmin and TACmax).
- Figure 5.17, legend: scenario 44 is not precautionary so Fig legend should be 'Request 3, ($SSB_{escapement}$ =65kt): Long-term distribution of TAC for individual years and replicates, including the cumulative probability, by ~~precautionary~~ scenario~~s~~.'
- Figure 7.2, legend: Relative F by quarter from 1000 ~~100~~ populations.

Review of management strategy evaluation for Norway Pout (WKNPOUT, 26-28 February 2018)

By: Martin Dorn (NMFS, Seattle), 1 May 2018

This is an external expert review of a management strategy evaluation (MSE) of Norway pout conducted by WKNPOUT during a 26–28 February 2018 meeting. A draft document was provided to me on 6 April, which afforded sufficient time to review the document. The default harvest policy for Norway pout is an escapement policy designed to achieve a high probability that the spawning biomass after fishing is above a pre-specified level. Catch recommendation are set for the coming year so that there is a 95% probability that the spawning stock will be above the Bl_{im} threshold. Usually the harvest policies are implemented by applying some target harvest rate to the stock assessment estimate of abundance. Here instead application of the harvest policy requires evaluation of the extreme tail of a distribution representing the uncertainty in the assessment. As such it is a difficult harvest control rule to test in a simulation framework. Most assessments contain a mix of estimated and assumed parameters, and therefore will understate the true uncertainty of the assessment.

The goal of the MSE was to evaluate a request to consider modifications to the default escape policy for Norway pout. Managers were interested in the feasibility of implementing upper and lower bounds on the TAC, and imposing a limit on the fishing mortality Fcap. When an upper limit is invoked on an escapement policy it is because a larger harvest is possible that would still achieve the 95% probability of being above B_{lim}. The contrary situation applies when catch is set to the lower limit. In this situation, a lower catch would be required to stay above 95% probability of being above B_{lim} in that year, so setting the catch to the lower bound would imply that probability of being above B_{lim} would be somewhat less than 95% in that year. In some years, the stock will have a lower than 95% probability of being above B_{lim} even without fishing.

The MSE evaluates these candidate harvest control rules by simulating the population dynamics of Norway pout, conditioned on the stock assessment model results using SESAM, a seasonal implementation of SAM, starting with the 2018 TAC year and projecting the stock forward for twenty years. Rather than conducting a full stock assessment in each year using SESAM, a stock assessment emulator is used to mimic the stock assessment. This is a much quicker method than doing a full-feedback MSE, but may not have captured the full range of uncertainty associated with conducting an annual stock assessment and then applying a harvest control rule based on that assessment.

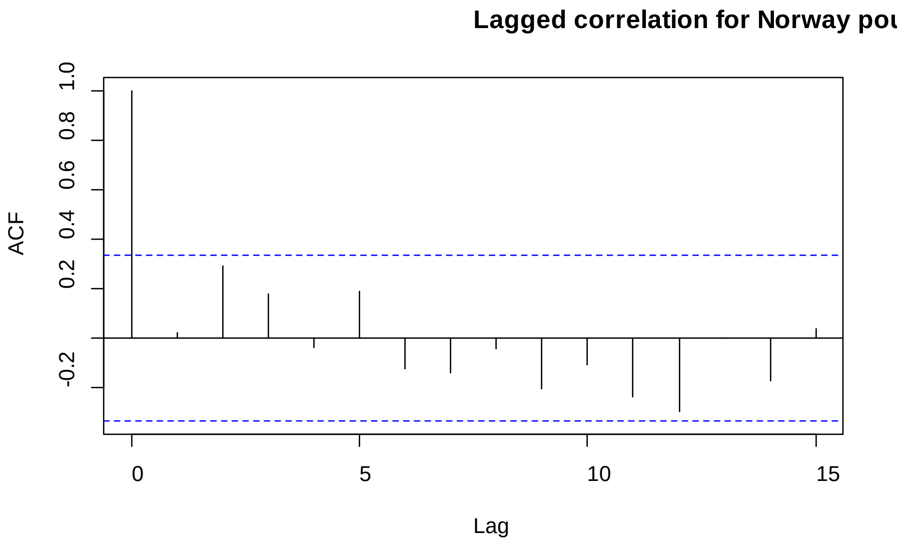
Overall my review found that the MSE was carefully done, with the methods clearly described, and results reported in appropriate figures and graphs. There was good discussion of the issues that were addressed. Below I comment on the operating model and the experimental design, and identify several of the report conclusions that I believe are not fully supported by the work presented.

- **Comments on operating model**

I support drawing samples from the assessment results to initiate the simulation (rather than using just the point estimates). Based on my experience, this is usually how the operating models are conditioned on stock assessment results.

However I think that the approach for simulating recruitment could have been done more cleanly. For nearly all scenarios (except the sensitivity runs using other SR relationships) recruitments when spawning biomass was above the breakpoint in the hockey stick model were obtained by resampling from estimated recruitments, and then estimation error was added in. Below the breakpoint, recruitments were simulated from a log normal distribution with a standard deviation consistent with historical recruitment variability. It is unusual to add estimation error when resampling from recruitment (I have never seen this before). Simulating recruitment using a different process when above and below the breakpoint is probably not a good idea, and may cause artifacts in the results. A better approach would have been to simulate recruitment from a log normal distribution above and below the breakpoint.

There is also a concern that that long term patterns in recruitment such as autocorrelation or decadal variation were not evaluated (the assumption was that each annual recruitment was an independent draw). To evaluate this assumption, I calculated the ACF for the historical Norway pout recruitment time series (below). Although none of the lagged correlations would be considered significant (extending past the dotted line), there is a suggestion of negative correlations at decadal scale that may cause problems in keeping the consistently above B_{lim}. Nevertheless the evidence for this is relatively weak, so I concluded that the independence assumption is appropriate.



Assuming the exploitation pattern stays constant throughout the simulation may understate uncertainty, though it was unclear to me what impact this might have on results.

I have some concern about a shortcut approach that was used in implementing the stochastic forecast. The stochastic forecast tunes the fishing mortality multiplier until there is a 5% probability of the stock being below B_{lim} across all the simulation replicates, and the TAC is computed as the median of the replicates. As the document notes, the correct way to do this would be to adjust the TAC and iteratively solve for the fishing mortality multiplier within each replicate until the requisite 5% probability is attained. There is no evaluation of whether the shortcut method a good approximation, and the document simply ignores away the potential problems. This is especially important because of many of the alternatives are very close to the 5% probability threshold.

- **Comments on experimental design**

There was a thorough evaluation of various permutations of harvest control rules to address the questions posed in the request from managers. The number of simulations appeared to be adequate to generate robust results. There was a good range of sensitivity analyses conducted.

One concern is that uncertainty in the estimation of the B_{lim} reference point was not considered. A decision was made to use previously established B_{lim} reference point, even though a stock recruitment analysis indicated considerable support for a segmented regression suggesting a much higher B_{lim} (~190 000 t) compared to the current value (72 101 t). While continuity in the provision of scientific recommendations is important, it was not clear to me what criteria used arrive at the decision to retain the current B_{lim} value. Finally, the performance of the difference harvest policy could have be evaluated with this much higher B_{lim} as a sensitivity test, but I strongly suspect that that none of the harvest policies evaluated would be considered precautionary under the higher B_{lim}.

- **Comments on conclusions**

The conclusion that the default harvest strategy is not precautionary is based on an analysis with a cap on fishing mortality of $F_{historical} = 5.0$. All other scenarios used a cap on the fishing mortality of $F_{historical} = 0.89$, since the fishing fleet would be unlikely to have the capacity or motivation to impose a fishing mortality above a certain level. This seems a reasonable and realistic assumption to make when evaluating an escapement harvest policy, but for some reason this constraint was not used in the analysis of the default harvest policy. Therefore I do not support the document's conclusion that the default harvest policy is not precautionary. This concern could easily be addressed by rerunning the default harvest policy with a $F_{historical} = 0.89$.

Some of the conclusions of the document are dependent on the definition of risk that is used. A decision made to use an ICES type 1 risk definition (average over years of the probability spawning biomass is below B_{lim}), rather than the type 3 definition recommended by the ICES WKGMSE Report, 2013 (maximum probability that spawning biomass is below B_{lim} over a number of years). Although these two calculation should give the same results once the simulation equilibrates, they gave different answers for period simulated in the MSE for a number of scenarios. For example, none of the request 1 scenarios that included minimum TAC would be considered precautionary when a type 3 definition of risk is used.

Annex 6: Reply to reviewers on the management strategy evaluation for Norway Pout (WKNPOUT, 26–28 February 2018)

By Mollie Elizabeth Brooks, Morten Vinther and Andrés Uriarte

Replies to María Manuela Azebedo's comments.

1) While the approach to simulate future R is considered adequate it would be helpful to plot examples of R from a few (2–3) simulated populations superimposed to the shaded area in the R plots (summary results plots: figures 5.1, 5.16, 5.18) to show how much jagged and variable R can be in individual populations during the simulation period and how they match the behaviour of the historical recruitment.

ANSWER: The suggestion was not followed as the report was in the final edition.

2) Despite no evidence of impairment R, the frequency of strong year classes have decreased since 1999. The median R in the 20 years simulation period (~43 861) is above the mean R since 1999 (33 683). Sensitivity tests on the influence of simulated R were carried out assuming in the OM a S-R relationship fitted to three types of S-R functional relationships (Ricker, B&Holt, hockey-stick) weighted by the 'Buckland' method (Section 4.5.2). What was the estimated weight for each S-R type? What was the estimated value of the hockey-stick break-point? The results from this sensitivity test show minor changes in the performance statistics but this could be due to the fact that the hockey-stick S-R was the dominant S-R type in the 'Buckland' fit, with break-point similar to the break-point assumed in the base-run simulations. If this is the case then the sensitivity test to recruitment does not add robustness to the proposed HCRs.

ANSWER: weighting factors appear in section 5.5.2 in page 56. "The Eqsim fit of SR estimates the highest weight for segmented regression (i.e. hockey stick 53%), some weight on the Ricker SR (38%) and low weight on the Beverton-Holt SR (9%) (Figure 5.21). Notice that the hockey stick fitting resulted in an inflection point around 190 000 t, well above the current B_{lim} 72 101 tonnes (B_{loss} in quarter 1 of 2005, as adopted in the benchmark)." Therefore the reason for the results from this sensitivity test to show minor changes in the performance statistics is not due to an artificial forcing of the inflection point of the segmented regression. The explanation is probably due measuring risk as the probability of falling below current adopted B_{loss} (current B_{lim} = 39 450 tonnes) instead of assessing this relative to the actual fitted inflection point around 190 000 t.

3) The MSE is a 'short-cut' approach since an assessment cycle with SESAM is not included in the management procedure/strategy component. Computing time for the MSE can be large and many scenarios were tested for Norway pout (~ 2 hours/HCR, running in parallel across cores!) so this might be the reason not to perform a 'full-feedback' MSE. The last assessment indicates a retrospective bias with overestimation of SSB(ICES WGNSSK Report 2017, Figure 12.3.8). Although the retro point estimates of SSB are inside the upper confidence interval of SSB from the last assessment, it would be useful to explore the likely effect of this retrospective pattern on the computed risks to SSB (probability of SSB Q4 falling below B_{lim}) to add robustness to the MSE (e.g.

running a full-feedback MSE for one of the scenarios showing risk to SSB close to 5%).

ANSWER: The reviewer is right, there seems to be retrospective bias in the estimate of SSB and Recruitment from the SESAM. This was not discussed this in the report, but we should also have looked at the retrospective pattern in the WGNSSK rep. It would be nice to have a full-feedback MSE (including the SESAM assessment in the loop) for one of the scenarios, but at the current stage of reporting this was considered unfeasible, so it has been left for future.

Performance statistics

4) Two types of risk were computed to evaluate if the HCRs are precautionary. Risk type 3.long.Q4 (*maximum* probability that SSB-Q4 is below B_{lim} = 39 450 tonnes, where maximum is taken across 2023–2037) is always higher than risk type 1.long.Q4 (*average* probability that SSB-Q4 is below B_{lim}, where the average is taken across 2023–2037). For example, risk3.long.Q4 is above 5% for scenarios 4,5, 6 (request 1), 18, 19, 21 to 24 (request 2) and 41 to 43 (request 3) while risk1.long.Q4 is below or just below 5% in these scenarios. Since the simulations still show variable probabilities in the long-term (e.g Figure 5.2) and to support the assumptions made that it is caused by an insufficient number of replicates in the MSE (1000), performance statistics could be presented for one of the scenarios simulated with a higher number of iterations. Nevertheless, according to ICES (WKGMSE, 2013) risk 1 can be considered an adequate measure of risk level for short-lived species, as SSB in any one year is almost independent of the previous years. This can be considered the case of Norway pout with highly variable recruitment rapidly influencing SSB. Moreover, results are also presented for risk1 at spawning time (Q1) showing < 5% probability of SSB-Q1 below B_{lim}-Q1 (72 101 tonnes) for the scenarios that are precautionary under the criteria $SSB_{escapement}$ in Q4 above 39 450 tonnes, giving confidence to the workshop conclusions.

ANSWER: The reviewer is right in the remarks, in general Ris1 < Risk3 (except at stationary conditions). The actual inequality might be due to insufficient number of replicates. But in general, the comments of the reviewer support the approach followed by WKNPOUT of adopting Risk 1 in the long term as the best estimate of Risk. The WKNPOUT agreed to run an additional case with 10 000 iterations to check the performance statistics whether the two measurements of risks converge towards the same values. See it in reply number 13. The results support the decision made by the group of adopting Risk 1 in the long term as the best estimate of Risk 3.

5) The Minor comments for the final version of the report we almost all considered for the clean version of the report

Replies to Martin Dorn's comments.

6) The MSE evaluates these candidate harvest control rules by simulating the population dynamics of Norway pout, conditioned on the stock assessment model results using SESAM, a seasonal implementation of SAM, starting with the 2018 TAC year and projecting the stock forward for twenty years. Rather than conducting a full stock assessment in each year using SESAM, a stock assessment emulator is used to mimic the stock assessment. This is a much quicker method than doing a full-feedback MSE, but may not have captured the full range of uncertainty associated with conducting an annual stock assessment and then applying a harvest control rule based on that assessment.

ANSWER: the reviewer is right pointing out that the procedure of the stock assessment emulator might not encapsulate the full uncertainty associated to a full assessment. But for practical reason of speed of the MSE process this simplification was adopted. See also the answer to point 3 (above). In any case the emulator assures used the actual Variance Covariance matrix from the assessment in 2017. As said in page 19, Section 4.3.1,: "the assessment emulator randomly draws an estimated state (log N and log E) from a multivariate normal distribution which has the true state as the mean and a variance-covariance matrix as estimated by SESAM (Figure 4.2) (ICES, 2017)."

7) the approach for simulating recruitment could have been done more cleanly. For nearly all scenarios (except the sensitivity runs using other SR relationships) recruitments when spawning biomass was above the breakpoint in the hockey stick model were obtained by resampling from estimated recruitments, and then estimation error was added in. Below the breakpoint, recruitments were simulated from a log normal distribution with a standard deviation consistent with historical recruitment variability. It is unusual to add estimation error when resampling from recruitment (I have never seen this before).

ANSWER: More correctly, for each of the 1000 replicates the historical recruitment was initially drawn from the historical values with estimation error, but no estimation error was added after bootstrapping. Every initial 1000 states contains an initial stock numbers at age, exploitation pattern and a past recruitment series from which recruitment will be bootstrapped. Therefore the uncertainty affects the initial state and past series of recruitments. This approach is being implemented in MSE procedures when the initial populations comes out from Bayesian assessments where the initial true states can be drawn from the posterior distributions of current and past states (see for instance the anchovy MSE STECF 2014).

Scientific, Technical and Economic Committee for Fisheries (STECF) - Evaluation/scoping of Management plans - Data analysis for support of the impact assessment for the management plan of Bay of Biscay anchovy (COM(2009)399 final). (STECF-14-05). 2014. Publications Office of the European Union, Luxembourg, EUR 26611 EN, JRC 89792, 128 pp.

8) Simulating recruitment using a different process when above and below the breakpoint is probably not a good idea, and may cause artefacts in the results. A better approach would have been to simulate recruitment from a log normal distribution above and below the breakpoint.

ANSWER: it might be seen inconsistent to use resampling (for SSB above B_{lim}) and a "left-leg" hockey stick model with log-normal distributed errors for SSB below B_{lim}. We had however a good reason for that for Recruitment approach (historical recruitment is not log-normal distributed, and we have no estimates of recruitment for SSB below B_{lim}). Therefore the group kept with the approach followed.

9) There is also a concern that that long term patterns in recruitment such as autocorrelation or decadal variation were not evaluated (the assumption was that each annual recruitment was an independent draw). To evaluate this assumption, I calculated the ACF for the historical Norway pout recruitment time series (below). Although none of the lagged correlations would be considered significant (extending past the dotted line), there is a suggestion of negative correlations at decadal scale that may cause problems in keeping the consistently above B_{lim}. Nevertheless the evidence for this is relatively weak, so I concluded that the independence assumption is appropriate.

ANSWER: the team agreed with the reviewer's comments. Given the weak autocorrelation this was considered not a major issue for the validity of results.

10) I have some concern about a shortcut approach that was used in implementing the stochastic forecast. The stochastic forecast tunes the fishing mortality multiplier until there is a 5% probability of the stock being below B_{lim} across all the simulation replicates, and the TAC is computed as the median of the replicates. As the document notes, the correct way to do this would be to adjust the TAC and iteratively solve for the fishing mortality multiplier within each replicate until the requisite 5% probability is attained. There is no evaluation of whether the shortcut method a good approximation, and the document simply ignores away the potential problems. This is especially important because of many of the alternatives are very close to the 5% probability threshold.

ANSWER: The reviewer was correct in his comments and it was admitted during WKNPOUT meeting. The WK actually explored how well the ad hoc forecast procedure performs with the HCRs within the OM. We thought it would be convenient to validate or not the procedure as it is being currently implemented. However, we admit this is an issue for the WG to improve in future.

11) One concern is that uncertainty in the estimation of the B_{lim} reference point was not considered. A decision was made to use previously established B_{lim} reference point, even though a stock recruitment analysis indicated considerable support for a segmented regression suggesting a much higher B_{lim} (~190 000 t) compared to the current value (72 101 t). While continuity in the provision of scientific recommendations is important, it was not clear to me what criteria used arrive at the decision to retain the current B_{lim} value. Finally, the performance of the difference harvest policy could have be evaluated with this much higher B_{lim} as a sensitivity test, but I strongly suspect that that none of the harvest policies evaluated would be considered precautionary under the higher B_{lim}.

ANSWER: There was no consensus that this was not a type 5 stock as agreed during the last Benchmark in 2016 (defined as "Stocks with no evidence that recruitment has been impaired or with no clear relation between stock and recruitment (no apparent S–R signal).”). Justification for not changing B_{lim} was given in Section 3.1.2 in pages 8 and 9. The approach of the group was to admit that there might be a segmented regression with a higher inflection point than current B_{lim} (as fitted with with Eqsim SR), and testing through a sensitivity analysis the actual risk of falling below current B_{lim}= B_{loss} (below past ranges of biomasses). The sensitivity analysis shows that stablishing the inflection point as best fitted to the data with Eqsim SR (i.e. around 190 000 t), results in negligible changes to the risk on current B_{lim} (at least for the cases with F_{cap} around 0.3 and 0.5, as shown in Section 5.5.2 (pages 56 & 57)). Obviously, if B_{lim} would be set at the inflection point of the fitted segmented regression (i.e. around 190 000 t) then the group considers (as the reviewer) that very few (if any) of the tested HCRs would be precautionary. Probably only HCRs with TACmin = 0 could be precautionary, because in those, the forecast will target the new B_{lim} reference point without being overridden.

12) The conclusion that the default harvest strategy is not precautionary is based on an analysis with a cap on fishing mortality of $F_{historical}$ = 5.0. All other scenarios used a cap on the fishing mortality of $F_{historical}$ = 0.89, since the fishing fleet would be unlikely to have the capacity or motivation to impose a fishing mortality above a certain level. This seems a reasonable and realistic assumption to make when evaluating an escapement harvest policy, but for some reason this constraint was not used in the analysis of the default harvest policy. Therefore I do not support the document’s conclusion that the default harvest policy is not precautionary. This concern could easily be addressed by rerunning the default harvest policy with a $F_{historical}$ = 0.89.

ANSWER: A new run for the default harvest policy with a $F_{historical}$ = 0.89 has been done and its results regarding performance statistics can be seen in the table below. ”Scen d” shows that with an assumption of $F_{historical}$ at 0.89, the default escapement strategy becomes precautionary (risk1.long.Q4 = 2.4%). This conclusion is however very dependent on the value of $F_{historical}$ as this value is reached in 35.9% of the cases (atF_{hist} = 35.9%). ICES will probably not accept such high sensitivity to a value that we can’t measure or control, and may prefer going for an HCR with an F_{cap}, as shown for instance in “Scen a” (and see Annex 4 to the report).

Table A1. Scenarios without TACmin and TACmax. (Shaded scenarios have more than 5% probability of SSB in Quarter 4 being below B_{lim})

	Scen a: Bounded by F_{cap} = 0.70	Scen b: Bounded by F_{cap} = 0.75	Scen c: Practically No bounds	Scen d: Bounded by $F_{historical}$ = 0.89
$F_{historical}$	5	5	5	0.89
F_{cap}	0.7	0.75	5	5
TACmin	0	0	0	0
TACmax	5000000	5000000	5000000	5000000
F_{bar}.median	0.493	0.501	0.494	0.558
F_{bar}.mean	0.591	0.62	0.952	0.519
SSB.median	103282	101749	89310	107051
SSB.mean	124466	122146	104769	125191
risk3.short.Q4	4.8	5	9.2	2.6
risk1.long.Q4	4.87	5.27	9.2	2.4
risk3.long.Q4	5.9	6.1	10.4	3.4
risk1.long.Q1	3.78	4.09	7.3	2
atF_{hist}	0.2	0.3	4.7	35.9
TAC.median	109479	109538	98696	120894
TAC.mean	138244	142081	181387	206749
TRC.median	109479	109538	98696	109125
TAC.change	123917	130079	208235	207012
TAC.changeRel	13007	13747.4	22494.9	16205.8
atTACmin	0	0	0	0
atTACmax	0	0	0	0
YearToRecover.Q4	1.2	1.2	1.2	1.2
YearToRecover.Q1	1.75	1.75	1.79	1.67

13) Some of the conclusions of the document are dependent on the definition of risk that is used. A decision made to use an ICES type 1 risk definition (average over years of the probability spawning biomass is below B_{lim}), rather than the type 3 definition recommended by the ICES WKGMSE REPORT, 2013 (maximum probability that spawning biomass is below B_{lim} over a number of years). Although these two calculations should give the same results once the simulation equilibrates, they gave different answers for period simulated in the MSE for a number of scenarios. For example, none of the request 1 scenarios that included minimum TAC would be considered precautionary when a type 3 definition of risk is used.

ANSWER: Both risks are reported and will be included in the draft advice to ICES. As mentioned in the WKGMSE 2013 (ICES CM 2013 / ACOM 39), In the stationary situation, Prob 3 = Prob 1 and only Prob 1 should be computed. In summary for the short term Risk 3 is taken as the reference whilst for the long term Risk 1 is taken as the best estimation of Prob 3 (as pointed out by WKGMSE 2013). See also point 4 above.

The WKNPOUT agreed to run an additional case with 10 000 iterations to check the performance statistics on Risk 1 and Risk 3, whether the two measurements of risks converge towards the same values in the long term. The result of this exercise are shown in the table and figure placed below. By running 10 000 replicates on scenario 21, risk3.long.Q4 get far closer to the estimation risk1.long.Q4 than when the comparison was based only on a 1000 replicates. And from the estimates of risks in the case of a 1000 replicates the one closer to the final values with 10 000 replicates is that of risk1.long.Q4. Therefore, results support the decision made by the group of adopting Risk 1 in the long term as the best estimate of Risk 3.

Table A2. Performance statistics for request 2, scenario 21 with 1000 and 10 000 replicates

	1000 replicates	10000 replicates
F_{cap}	0.3	0.3
TACmin	30000	20000
TACmax	150000	150000
F_{bar}.median	0.298	0.296
F_{bar}.mean	0.327	0.327
SSB.median	128784	130875
SSB.mean	160648	163653
risk3.short.Q4	3.3	4.0
risk1.long.Q4	4.1	4.3
risk3.long.Q4	5.5	4.6
risk1.long.Q1	3.9	4.1
atF_{hist}	1.1	0.7
TAC.median	71706	72161
TAC.mean	79107	79897
TRC.median	71692	72090
TAC.change	36744	36778
TAC.changeRel	0.6	0.6
atTACmin	22.9	22.5
atTACmax	14.2	14.7

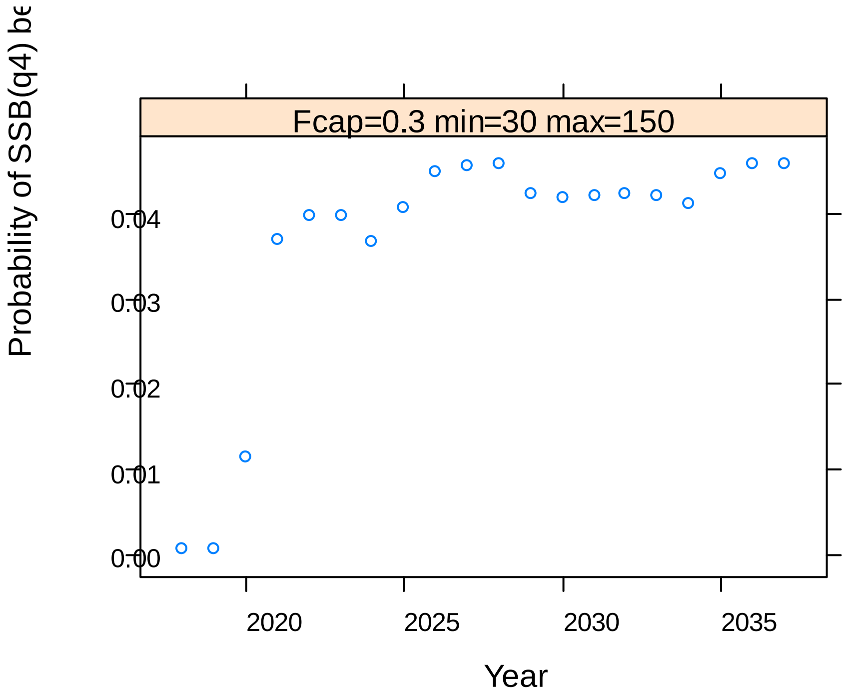


Figure A1. Probability of SSB in Quarter 4 is below B_{lim} by year for scenario 21 with 10 000 replicates.