WORKSHOP FOR THE REVIEW OF THE ASSESSMENT OF A NEW REBUILDING PLAN FOR WESTERN HORSE MACKEREL (WKWHMRP)

VOLUME 3 | ISSUE 37

ICES SCIENTIFIC REPORTS

RAPPORTS SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA

CIEM COUNSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

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

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

ISSN number: 2618-1371 I © 2021 International Council for the Exploration of the Sea

ICES Scientific Reports

Volume 3 | Issue 37

WORKSHOP FOR THE REVIEW OF THE ASSESSMENT OF A NEW REBUILDING PLAN FOR WESTERN HORSE MACKEREL (WKWHMRP)

Recommended format for purpose of citation:

ICES. 2021. Workshop for the review of the assessment of a new rebuilding plan for Western Horse Mackerel (WKWHMRP).
ICES Scientific Reports. 3:37. 232 pp. http://doi.org/10.17895/ices.pub.8023

Editors

Einar Hjörleifsson

Authors

Jose Beltran • Thomas Brunel • Andrew Campbell • Gonçalo Carvalho • Gersom Costas • Miguel Cunha • Ian Gatt • Michaël Gras • Einar Hjörleifsson • Pedro Jorge • Laurence Kell • Christoph Konrad • Ursula Krampe • Gwladys Lambert • Polina Levotin • Steve Mackinson • David Miller • Patrick Murphy • Sean O'Donoghue • Alessandro Orio • Rosana Ourens • Martin Pastoors • Carlos Gutiérrez Pedrajo • Sara Pipernos • Gaëlle Renard • Sonia Sánchez-Maroño • Ana Silva • Vanessa Trijoulet • Torcuato Teixeira Valoria • Luís Vicente • Ashley Wilson

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

Contents

i Executive summary

The Workshop for the review of a new rebuilding plan for Western Horse Mackerel (WKWHMRP) evaluated a rebuilding plan as proposed by PeLAC during three virtual meetings in February and March 2021. The review was based on a technical report submitted by PeLAC as well as additional analysis under the remit of WKWHMRP. Two independent reviewers concluded that all ToR were adequately covered and that the minimum requirements for simulation testing HCRs, as developed by WKGMSE process, were met.

The tool used was based adaption of the ICES standard software package EqSim to include alternative harvest rules with optional stability mechanisms, incorporate uncertainty in initial conditions and generate additional outputs for model validation and HCR performance. The starting conditions were based on the 2020 ICES stock synthesis assessment. Alternative more pessimistic 2014-2018 recruitment scenarios were explored. An assessment based on SAM forecast method was used as an alternative exploratory tool.

The simulation results indicate that the proposed plan offers the potential for rebuilding of the stock by 2028, with rebuilding considered to be achieved when the stock size has exceeded B_{pa} with a 50% probability for three consecutive years. Risk to B_{lim} falls below 5% by 2025. The alternative recruitment scenarios showed rebuilding potential although the time frame in the most pessimistic, although unlikely these scenarios indicated that the time frame may be longer than ten years.

ii Expert group information

Expert group name	The Workshop for the review of the assessment of a new rebuilding plan for Western Horse Mackerel (WKWHMRP)
Expert group cycle	Annual
Year cycle started	2021
Reporting year in cycle	1/1
Chair(s)	Einar Hjörleifsson, Iceland
Meeting venue(s) and dates	5 February 2021, online, 31 participants
	29 March 2021, online, 31 participants

1 Introduction

The Workshop for the review of a new rebuilding plan for Western Horse Mackerel (WKWHMRP) chaired by Einar Hjörleifsson, Iceland, and reviewed by Allen R. Kronlund, Canada, and Jonathan Deroba, USA, will be established and will meet online on 5th February 2021 to:

a) Review the approach (i.e., analytical methods, application and interpretation) described in the Pelagic Advisory Council (PelAC) report (listed below) for the evaluation of the proposed Harvest Control Rule (HCR) in the rebuilding plan. The review should consider:

  i) Whether the tools used (methods), and the model conditioning done (data), are appropriate for the stock;
  ii) Whether the minimum requirements for simulation testing HCRs, as developed by WKGMSE process, are met;
  iii) The appropriateness of the criteria used to draw conclusions on the performance of the HCRs

b) On the basis of (a), determine whether the HCR evaluations presented in the PelAC are sufficient to evaluate the proposed rebuilding plan against precautionary criteria and the rebuilding plan targets and measures detailed in Article 5 of the proposed rebuilding plan.

c) Depending on the outcome of (b), either:

  iv) analyze the results of the HCR evaluation, and develop conclusions on whether the proposed rebuilding plan can be considered precautionary and be used as the basis for ICES fishing opportunity advice for the stock; or
  v) Propose additional analyses or diagnostics that would allow for sufficient evaluation of the proposed plan

d) Should the proposed plan include elements that are in contradiction with ensuring that the stock is fished and maintained, also in the future, at levels which can produce MSY, comment specifically on such elements, and their consequences for ensuring MSY.

e) Deliver a report containing the key decisions and conclusions in relation to the TORs. The three independent external reviewers will also conduct additional scientific review.

The report detailing the evaluation and proposal of parameters for the HCR for the Western horse mackerel rebuilding plan is: Pastoors, M.A., Campbell, A., Trijoulet, V., Skagen, D., Gras, M., Lambert, G.I., Sparrevohn, C.R., and Mackinson, S. 2020. Report on Western Horse Mackerel Technical Focus Group On Harvest Control Rule Evaluations 2020. Report for Pelagic Advisory Council.

The proposed rebuilding plan is detailed in: ANNEX 1: PeLAC proposal for a rebuilding plan for Western horse mackerel. *Dated 28-Jul-2020.*

The ToR established were a result of the EU Commission request, see appendix.

Section 2 (Results) summarizes the main finings the WKWHMRP workshop with the Section 3 contain the reviewer's report. The participant list is given in Appendix 1 and the request from EU to ICES is given in Appendix 2. The two documents that were reviewed are given in Annex 3 and 4 with the Summary template for HCR modelling in Annex 5.

1.1 WKWHMRP Procedure

The review panel (chair and two external reviewers) was setup in January 2021. The independent reviewers delivered their report on 24 January 2021 and 8 February 2021. The ToR directed the reviewers and workshop participants to consider a specific report (PelAC report, see Annex 3) and proposed rebuilding plan. Based on the ToR and on additional discussion with ICES secretariat, it was the understanding of the review panel that it was not under the remit of WKWHMRP, as initially setup to undertake any major additional analysis. A virtual meeting (participant list in the appendix) was held on 5 February 2021 with the following agenda:

- Introduction of the PelAC report (Martin Pastoors) with questions and answers.
- Reviewers response.
- General questions and discussions among the WK-participants.
- Planning of WKWHMRP report writing.

Given the initial response of the reviewers in addition to the questions and discussion during the 5 February meeting it was concluded that additional work to support evaluation of the report and plan against the ToR is warranted. A list of items that needed further attention was established after the meeting and agreed upon at a virtual meeting held on 12 February 2021. The additional work relates not so much to the methodological approach used for the evaluation as a similar approach using same basic software has been recently used in, for example, the evaluation of Blue whiting (ICES, 2016). The additional work was delivered on 22. March (Working document, Annex 3) and a virtual meeting was held on 29. March. The conclusion, given the two documents was that the ToR's had been met. A condensed summary of the content of the two reviewed documents in given the following section.

2 Results

A summary of the simulation setup is provided in Annex 4.

2.1 Background

Annual catch advice for Western Horse Mackerel has been given on the basis of the ICES MSY approach since 2012, During this period the assessment model was changed (2017) and reference points revised (2017 and 2019). Despite the configuration of an assessment model (Stock Synthesis) offering greater flexibility than the previous (SAD) model, update assessments carried out since 2017 have often lead to a rescaling of the stock size, although to a reduced extent to that observed with the SAD model. Since the stock has been close to B_{lim} for most of this time, this has resulted in challenges for the provision of management advice consistent with recent estimates of stock development.

The development and adoption of a management plan for the provision of catch advice has been a long term goal of the Pelagic Advisory Council (PeLAC) since a plan was first proposed in 2008 and in 2020 a PeLAC technical focus group was established to identify and evaluate a number of potential harvest strategies. Given the proximity of the stock to B_{lim} and recognising that the risk to B_{lim} in the short term will remain above the precautionary 5% threshold in any evaluation, the focus group evaluated candidate harvest rules on the basis of proposing a recovery plan. The PeLAC submitted the evaluation to the European Commission for consideration who requested ICES review the evaluation. Further work was carried out by the same experts under the auspices of an ICES WKWHMRP.

2.2 Methods

The evaluation was carried out using two separate simulation tools. Both frameworks are of the 'short-cut' type whereby the future uncertainty in the assessment and short term forecast is incorporated via an emulator which is configured to generate future assessment errors consistent with those from the historic period (ICES, 2021). The simulation frameworks used are based on

1. An adaption of the ICES standard software package *EqSim*, used for the estimation of MSY reference points with the code updated to include alternative harvest rules with optional stability mechanisms, incorporate uncertainty in initial conditions and generate additional outputs for model validation and HCR performance.
2. The forecast module of the SAM assessment model although the results were not used to formally address the ToR..

Evaluations were performed for three different harvest rules:

- A constant F rule: a fixed F_{target} independent of the SSB
- An ICES-type rule: a fixed F_{target} when SSB is above $B_{trigger}$ with a linear decline in F to zero at the origin.
- A double breakpoint rule: a fixed F_{target} when SSB is above $B_{trigger}$ with a linear decline to 20% of F_{target} at B_{lim} . Below B_{lim} the target fishing mortality remains at 20% of F_{target}.

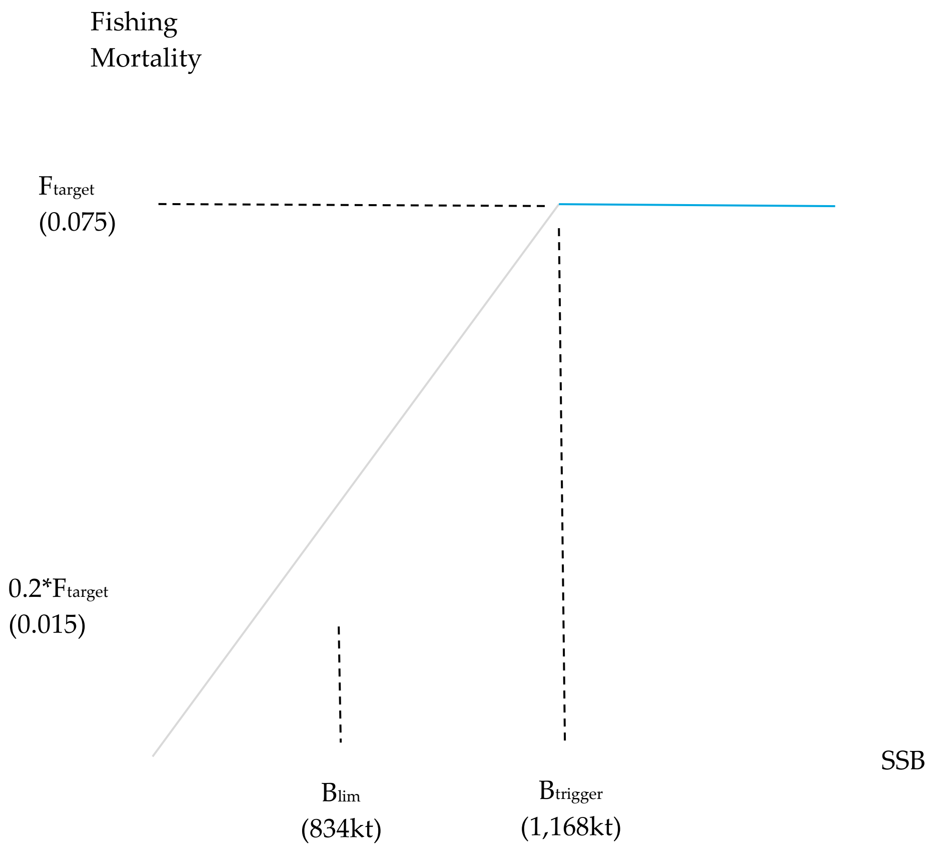
For all evaluations, $B_{trigger}$ was set at the corresponding MSY $B_{trigger}$ (B_{pa}) value. Six F_{target} values were simulated (0, 0.05, 0.075, 0.1, 0.125 and 0.15). Three variants of each rule to explore a range of TAC stabilisation measures were explored

- No stability mechanism
- A minimum TAC of 50kt
- A 20% limitation on the inter-annual variation in TAC, applied only when SSB is above $B_{trigger}$.

The simulations were conditioned on a range of operating models derived from the 2019 and 2020 SS3 assessments and also from an alternative SAM assessment. A range of additional runs were conducted to investigate the sensitivity of results with regard to assumptions on recent levels of recruitment, the parameterisation of assessment/advice error and reference point values.

Variability in starting numbers is incorporated via 1000 stock replicates derived from uncertainty estimates from the stock assessment. During the simulation period, uncertainty in weight at age is based on the variability observed in historic weight at age data and fishery selection from the stock assessment estimates. Maturity at age and natural morality for the simulation period are considered to be time invariant, as in the assessment. Future recruitment is modelled using the EqSim approach incorporating a segmented regression model with the breakpoint constrained at B_{lim}.

The main results are considered as those from the EqSim framework, conditioned on the most recent WG assessment i.e. the stock synthesis assessment from the 2020. The proposed management strategy is a double breakpoint rule with a 20% limitation on IAV when above $B_{trigger}$ with a target fishing mortality of 0,075 and is shown in figure 1.



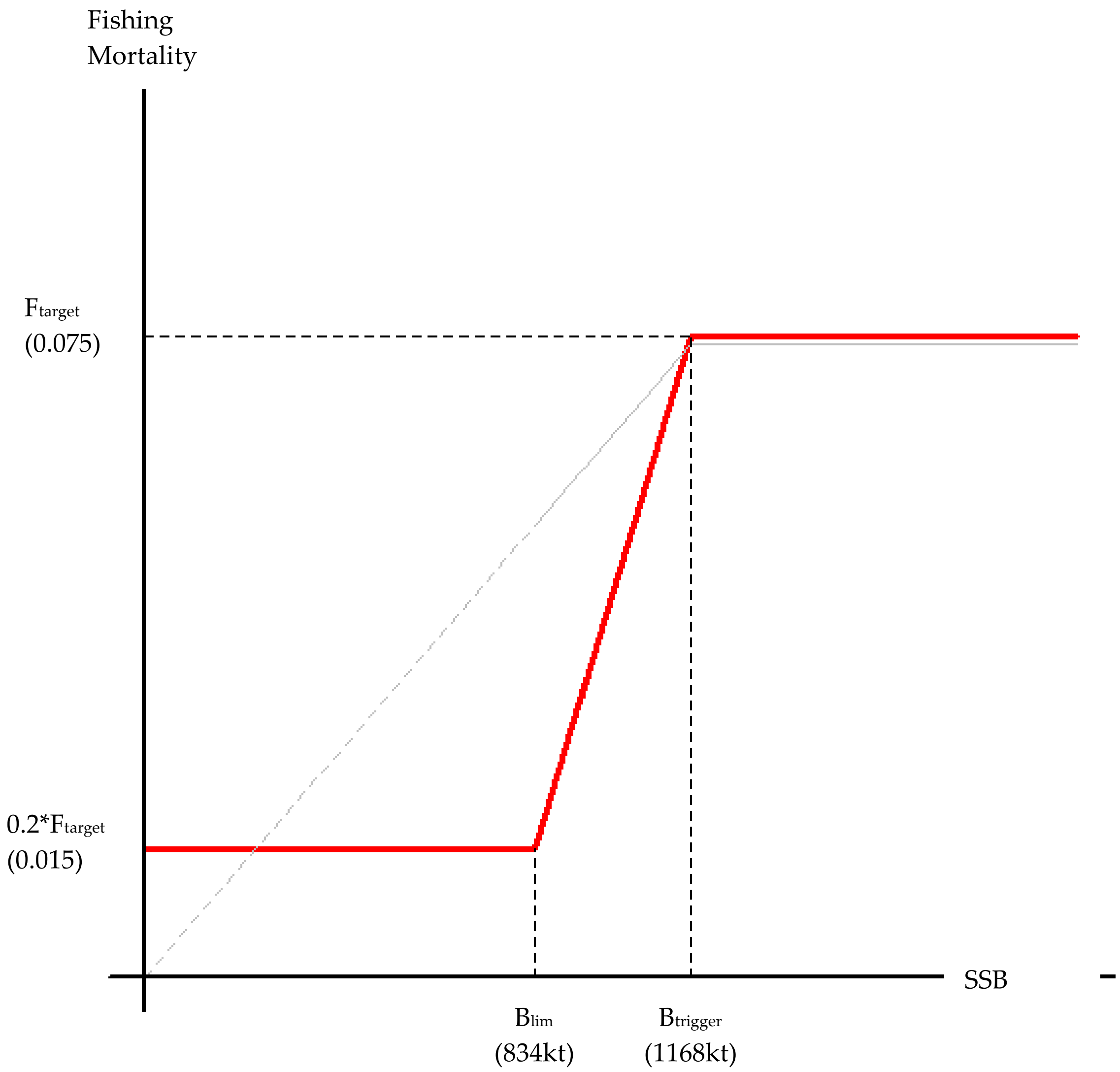


Figure 1: Schematic of the proposed double breakpoint harvest control rule for Western Horse Mackerel. When SSB is above the $B_{trigger}$ value (0.075), the fishing mortality to be applied is set as F_{target}. When SSB is below B_{lim} , a target fishing mortality of 20% of F_{target} is set. When SSB is below $B_{trigger}$ but above B_{lim} , the F applied is reduced linearly. The light grey line depicts the current ICES advice rule which stipulates a linear reduction in target F (from F_{MSY} (0.074)) to the origin. Note the steeper reduction associated with the proposed rule.

As the performance of the management strategy in the near term is of particular relevance in the context of evaluating recovery, annual values of performance metrics were considered rather than summarising by period. The performance indicators considered were

- SSB
- Realised fishing mortality
- Yield
- Simulated recruitment
- Probability that SSB<B_{lim}
- Recovery status, defined as the proportion of iterations that have remained above B_{pa} for 3 consecutive years, following a period of 1 year or more below B_{pa}.

2.3 Results and Conclusions

HCR performance

The results of the simulation for the proposed double breakpoint rule with a target fishing mortality of 0.075 are shown in figures 2 and 3.

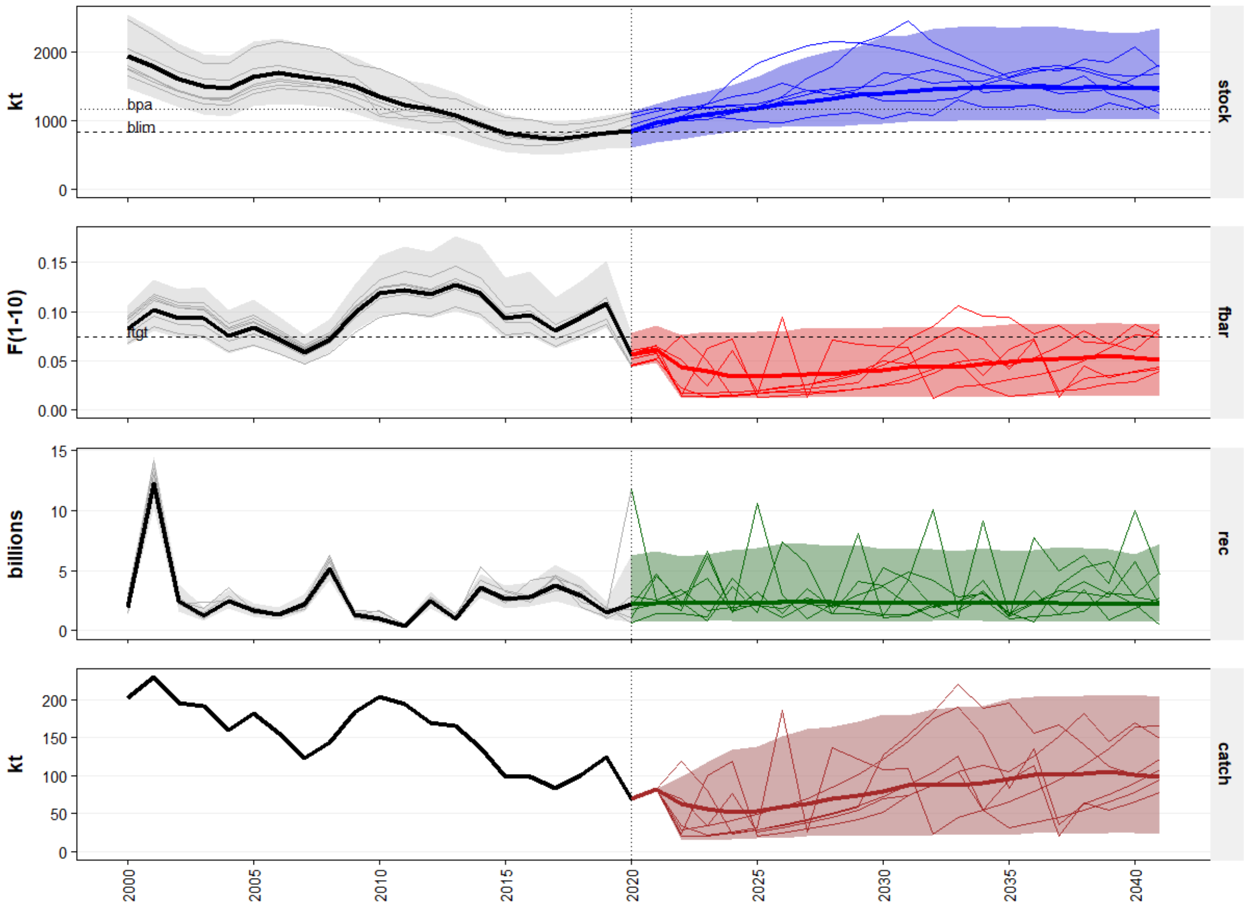


Figure 2 Historic stock development (pre 2019) and simulation output for the double breakpoint rule with F_{target} = 0.075 and a 20% TAC change limitation when above $B_{trigger}$. The top plot represents SSB (tonnes), the second fishing mortality, the third recruitment (age 0 in thousands) and the bottom yield (tonnes). The shaded area represents the 5th and 95th percentiles. The heavy line is the median value from the assessment and projection periods. The remaining lines depict five individual iterations.

A catch constraint applies in years 2019-2021 of the projection given estimates of total catch are available for these years. The first management year for the simulation is therefore 2022 during which the target fishing mortality is relatively low given the proximity of the stock to B_{lim} leading to median yields of the order of 50kt.

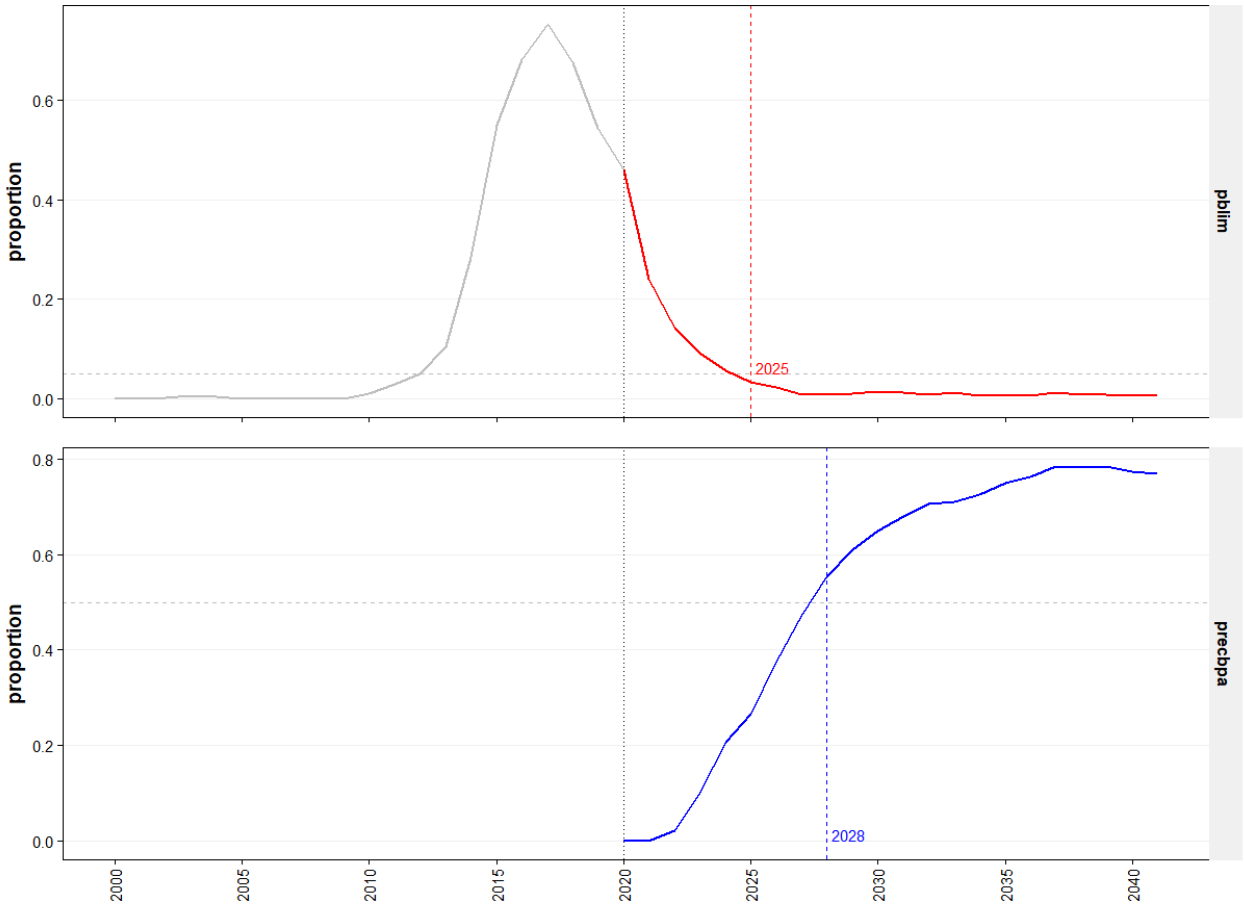


Figure 3. Annual probability of SSB<B_{lim} (top panel, 5% level indicated by horizontal dashed line) and proportion of replicates recovered to above B_{pa} (a recovered replicate is one which, having fallen below B_{pa} subsequently recovers are remains above B_{pa} for 3 consecutive years). Vertical blue line depicts the year when 50% of replicates have recovered to B_{pa}.

The simulation results indicate that the proposed plan offers the potential for rebuilding of the stock by 2027, with rebuilding considered to be achieved when the stock size has exceeded B_{pa} with a 50% probability for three consecutive years. Risk to B_{lim} falls below 5% by 2025

Robustness tests

The sensitivity to recent (stronger) recruitment was explored via three alternative scenarios where the 2014-2018 recruitments were considered to be 1) the equivalent of 50% of the assessment estimate 2) the geometric mean of the recruitment estimates over an extended period of generally low recruitment (2002 to 2013) or 3) the mean on the 5 lowest estimates from the full time series.

Under the all reduced recruitment scenarios the risk to B_{lim} increases in the initial years of the simulation (in contrast to the baseline scenario) as a greater proportion of the replicates are below B_{lim}. The cumulative effect of 5 years of reduced recruitment is a delay in the onset of the period of stock growth and reduced risk to B_{lim}, most pronounced for the most pessimistic recruitment scenario. . The rebuilding plan target could still be met with the proposed HCR, but the time period of rebuilding may extend past the ten-year rebuilding plan time frame.

Using the 2019 assessment as the basis for the operating model for the simulations leads to a more rapid rebuilding to Bpa than the 2020 assessment. For the F_{target} value of 0.075 with SS3, the rebuilding to Bpa would be in 2024 using the 2019 assessment and 2027 using the 2020 assessment. The main factor contributing to the delay is the lower estimates of recruitment associated with the 2020 assessment.

3 References

Cox, S. P., Benson, A. J., Doherty, B., and Johnson, S. 2018. Evaluation of potential rebuilding strategies for the Western Horse Mackerel (*Trachurus trachurus*) fishery.

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

ICES. 2016. Report of the Workshop on Blue Whiting (*Micromesistius poutassou*) Long Term Management Strategy Evaluation (WKBWMS), 30 August 2016, ICES HQ, Copenhagen, Denmark. ICES CM 2016/ACOM:53. 104 pp.

ICES. 2019. Workshop on Guidelines for Management Strategy Evaluations (WKGMSE2). ICES Scientific Reports. 1:33. 162 pp. http://doi.org/10.17895/ices.pub.5331.

ICES. 2021. The third Workshop on Guidelines for Management Strategy Evaluations (WKGMSE3). ICES Scientific Reports. 2:116. 112 pp. http://doi.org/10.17895/ices.pub.7627

Annex 1: List of participants

Name	Institute	Country (of Institute)	Email
Alessandro Orio	SLU	Sweden	alessandro.orio@slu.se
Ana Silva	Federação das Pescas dos Açores	Portugal	dcosta@federacaopescasacores.pt
Andrew Campbell	Marine Institute	Ireland	andrew.campbell@marine.ie
Ashley Wilson	The Pew Trusts	UK	awilson@pewtrusts.org
Carlos Gutiérrez Pedrajo	Federación de Cofradias de Pescadores de Cantabria	Spain	federacion@fecopesca.es
Christoph Konrad	JRC	Italy	Christoph.KONRAD@ec.europa.eu
David Miller	ICES	Denmark	david.miller@ices.dk
Einar Hjörleifsson (*Chair*)	MFRI	Iceland	einar.hjorleifsson@hafogvatn.is
Gaëlle Renard	OP PECHEURS D'AQUITAINE	France	gaelle.renard@pecheursdaquitaine.eu
Gersom Costas	Instituto Español de Oceanografía	Spain	gersom.costas@ieo.es
Gonçalo Carvalho	Marine Conservation and Cooperation Association	Portugal	gcarvalho@sciaena.org
Gwladys Lambert	Cefas	UK	gwladys.lambert@cefas.co.uk
Ian Gatt	Scottish Pelagic Fishermen's Association	UK	ian.gatt@scottishpelagic.co.uk
Jose Beltran	OPP-Lugo (member of PeLAC)	Spain	josemfbeltran@gmail.com
Laurence Kell	SEA++	UK	laurie@seaplusplus.co.uk
Luís Vicente	ADAPI	Portugal	adapi.pescas@mail.telepac.pt
Martin Pastoors	Pelagic Freezer-Trawler Association	Netherlands	mpastoors@pelagicfish.eu
Michaël Gras	JRC	Italy	Michael.GRAS@ec.europa.eu
Miguel Cunha	ADAPI	Portugal	antonio.cunha@testacunhas.pt
Patrick Murphy	IS&WFPO	Ireland	Carmel@IrishSouthAndWest.ie
Pedro Jorge	ADAPI	Portugal	pedrojorge@frip.pt

Name	Institute	Country (of Institute)	Email
Polina Levotin	Imperial College London	UK	polina.levontin02@imperial.ac.uk
Rosana Ourens	Cefas	UK	rosana.ourens@cefas.co.uk
Sara Pipernos	The Ocean Foundation	USA	spipernos@oceanfdn.org
Sean O'Donoghue	Killybegs Fishermens (PeLAC)	Ireland	sean@kfo.ie
Sonia Sánchez-Maroño	AZTI	Spain	ssanchez@azti.es
Steve Mackinson	Scottish Pelagic Fishermen's Association	UK	steve.mackinson@scottishpelagic.co.uk
Thomas Brunel	Wageningen University & Research	Netherlands	thomas.brunel@wur.nl
Torcuato Teixeira Valoria	Organización OPP 31 PESCAGALICIA	Spain	torcuato@r.gal
Ursula Krampe	EU Commission DGMARE	Belgium	Ursula.Krampe@ec.europa.eu
Vanessa Trijoulet	DTU	Denmark	vtri@aqua.dtu.dk

Annex 2: EU request

Request from	**European Commission, DG MARE, Unit C1**
Contact within organisation	**MARE-SCIENTIFIC-ADVICE@ec.europa.eu**
Content contact person	**Cannelle Beauchesne; cannelle.beauchesne@ec.europa.eu**
Request announced	**September 2020**
Request received	
Answer deadline client	
Request code (client)	**20-07 Western horse mackerel**
Request code (ICES)	
Request	**EU request to ICES on the assessment of a new rebuilding plan for Western Horse Mackerel (Trachurus trachurus) in ICES Subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a–c, and 7.e–k**

Request to ICES:

ICES is requested to evaluate the proposal for a rebuilding plan for Western Horse Mackerel as prepared by the Pelagic Advisory Council in July 2020:

- In particular, ICES is requested to assess whether this plan is seen as precautionary on the short term as well as on the long term.
- ICES is furthermore requested to assess whether the plan is consistent with the objectives to ensure stock recovery and bringing the biomass above sustainable levels within the indicated timeframes, and whether it is consistent with the maximum sustainable yield objectives of the CFP.
- Should the proposed plan include elements that are in contradiction with ensuring that the stock is fished and maintained, also in the future, at levels which can produce MSY, ICES is requested to comment specifically on such elements, and their consequences for ensuring MSY.

Planning ICES	
Request (budget) accepted	
ICES contact person	
WG(s) involved	
Preparation timing	
Review group	
Advice drafting group	
ACOM Webex	
Release date	

ANNEX I

PELAC proposal for a rebuilding plan for Western horse mackerel

28 July 2020

Background

1. The development of a robust and scientifically evaluated management plan for Western horse mackerel (WHOM) has been a long-term objective of the Pelagic Advisory Council (PELAC). We achieved this very early on in our existence in 2007, but unfortunately the agreed management plan was no longer considered precautionary by ICES in 2013. Efforts to achieve this objective again have been ongoing since 2015, but have encountered a number of challenges.

2. In 2015, the Marine Institute (MI), together with Cefas, sought to update the agreed management plan of 2007, which was no longer considered precautionary. There was a change in perception of the stock related to perceived changes in the egg survey, which changed the perception of the assessment as well. The MI and Cefas conducted two evaluations, and found that even with no fishing, the risk of falling below B_{lim} was more than 5%. While the SSB appeared to increase, the uncertainties were still so high that it increased slowly. The uncertainty in the assessment was therefore too large to conduct a meaningful Management Strategy Evaluation (Campbell et al. 2015).

3. In 2017, the inclusion of new data sources during a benchmark meeting resulted in a new assessment approach (ICES 2017). From that assessment, new reference points were estimated.

4. In 2018, ICES issued an advice for a considerable increase in TAC, close to MSY $B_{trigger}$, due to re-scaling of the assessment (ICES 2018).

5. An external expert (Landmark Fisheries) carried out an analysis to look at possible HCRs for potential management plans for WHOM. In most of the scenarios, the stock was expected to increase (Cox et al. 2018). The outcomes were presented at WGWIDE in 2018. The conclusion was that while the approach was welcomed, it did not take into account the right types of uncertainty in the starting conditions, which then lead to an overly optimistic evaluation (ICES 2018).

6. In 2019, an inter-benchmark meeting led to a revision of the reference points, indicating the stock was just above B_{lim} (ICES 2019). A collaboration between scientists working on different rebuilding methods for herring stocks subject to a zero catch advice (Celtic Sea, Western Baltic spring spawning and 6a herring) was set up, to explore whether these techniques could be applied to WHOM (PELAC 2019). While formally the stock is not in the rebuilding phase, it could potentially happen at any moment because of revisions in the assessment. The overall stock biomass levels are considered low, the assessment is volatile and the uncertainties are great. Therefore, the PELAC considers the development of a rebuilding plan more appropriate than a management strategy.

7. In February 2020, an ICES workshop on rebuilding plans (WKREBUILD) was held, which followed from the lack of criteria within ICES to evaluate rebuilding plans, making it difficult to get them through the ICES system. The work on Western horse mackerel was presented as a case study. A recommendation to ICES followed from the workshop regarding the estimation rebuilding timeframe, as opposed to the current ACOM rule that requires rebuilding to take place within one year (ICES 2020).

8. A dedicated group of experts on Western horse mackerel and management strategy evaluation assembled under the remit of the Pelagic Advisory Council in the first half of 2020, to develop the data and models to evaluate potential recovery measures for the Western horse mackerel stock. The group carried out a full evaluation of potential recovery measures based on two stock assessments (stock synthesis as used by ICES WGWIDE and an exploratory SAM assessment) and using two evaluation techniques (EqSim simulator and SAM HCR forecast). On the basis of the evaluations, the PELAC selected a specific set of recovery measures that are embedded in this rebuilding plan. The evaluation process is fully documented and will be submitted for scientific review by ICES.

Rebuilding Plan Western horse mackerel

ICES Subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a–c, and 7.e–k in the Northeast Atlantic

Objective

The purpose of the Western horse mackerel rebuilding plan is to ensure stock recovery to safe biomass levels and a long-term stock exploitation that is consistent with the precautionary approach and with achieving the objective of maximum sustainable yield (MSY).

Criteria and definitions

Article 1 - Subject matter

This rebuilding plans pertains to the Western horse mackerel stock.

Article 2 - Geographical definitions of stock

ICES Subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a–c, and 7.e–k in the Northeast Atlantic.

In certain times of the year, for the purposes of the scientific assessment, the divisions between the Western and North Sea horse mackerel stocks change. ICES division 4a and the Western part of division 3a are considered to be part of the North Sea stock in quarters one and two, but are part of the Western stock in quarters three and four (ICES 1989).

Article 3 Definitions

1. "Rebuilding plan timeframe": the timeframe for achieving the rebuilding plan target is a maximum of ten years, although all attempts will be made to realise that target within five years.

2. "Rebuilding plan target": when the spawning stock biomass is greater than B_{pa} for a minimum of three consecutive years.

Biological reference points used in this recovery plan are defined in the introduction to the 2018 ICES advice (ICES (2018). 1.2 Advice basis, ICES. ices.pub.4503).

Article 4 Reference points

1. The applicable biomass reference points for the Western horse mackerel stock shall be as follows: B_{lim} = 834 480 tonnes.

   MSY $B_{trigger}$ = B_{pa} = 1 168 272 tonnes.

   It should be noted in case of this rebuilding plan the value of MSY Btrigger is identical to Bpa and should be read as one wherever mentioned in the text. Should this relationship change in the future the plan is no longer valid.

2. The maximum fishing mortality associated with Maximum Sustainable Yield (F_{msy}) for the Western horse mackerel stock shall be as follows: F_{msy}= 0.074.

These values are based on the 2019 inter-benchmark report (ICES 2019).

Article 5 Rebuilding plan targets and measures

1. The rebuilding plan will be considered to be achieved when the spawning stock biomass is greater than B_{pa} for a minimum of three consecutive years.
2. The timeframe for achieving the rebuilding plan target is a maximum of ten years although all attempts will be made to realise that target within five years.
3. The TAC setting mechanism during the rebuilding plan shall be as follows:
   a. When the stock (SSB) is estimated to be below Blim in the assessment year, the TAC will be fixed with a fishing mortality equivalent to 20% of Fmsy = 0.015.
   b. When the stock (SSB) is estimated to be between Blim and Bpa in the assessment year, the TAC will be fixed with a fishing mortality equivalent to:

      0.015 + (SSB-Blim)/(Bpa-Blim) * (Fmsy-0.015).
   c. When the stock (SSB) is estimated to be above B_{pa} in the assessment year, the TAC shall be fixed with a fishing mortality equal to Fmsy (0.074), subject to the constraint that the change in TAC compared to the current (assessment) year does not exceed 20%.

Article 6 End of the rebuilding plan

The rebuilding plan may be superseded by a long-term strategy for the stock when, according to ICES, the spawning stock biomass is assessed to have been above B_{pa} for three consecutive years.

Should any other underlying assumption, or the definitions of the stocks in Article 2, of the rebuilding plan change based on new scientific knowledge this rebuilding plan will be deemed no longer to be applicable.

Article 7 Evaluation and implementation

This rebuilding plan will be submitted to the European Commission by the Pelagic AC with a request that the Commission forward it to ICES for scientific review of the management strategy evaluation of this rebuilding plan. The Pelagic AC requests that the rebuilding plan option, if deemed precautionary by ICES, be included in the short-term forecast options table for the following year and thereafter in the ICES advice.

References

Campbell, A., De Oliveira, J. A. A., Kelly, C., and Roel, B. 2015. Report on Western Horse mackerel Management Strategy Evaluations (MSE). Report for Pelagic Fishing Industry.

ICES, 2017. Report of the Benchmark workshop on widely distributed stocks (WKWIDE) 30 January–3 February 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:36.

ICES 2018. Horse mackerel (*Trachurus trachurus*) in Subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a–c, and 7.e–k (the Northeast Atlantic). In Report of the ICES Advisory Committee, 2018. 28 September 2018:http://www.ices.dk/sites/pub/Publication%20Reports/Advice/2018/2018/hom.27.2a4a5b6a7a-ce-k8.pdf 1.2 Advice basis, ICES. ices.pub.4503.

Cox, S. P., Benson, A. J., Doherty, B., and Johnson, S. 2018. Evaluation of potential rebuilding strategies for the Western Horse Mackerel (Trachurus trachurus) fishery.

ICES. 2018. Report of the Working Group on Widely Distributed Stocks (WGWIDE) 28 August - 3 September2018, Torshavn, Faroe Islands. ICES C.M. 2018 / ACOM:23.

ICES. 2019. Interbenchmark Protocol on Reference points for Western horse mackerel (*Trachurus trachurus*) in subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a-c,e-k (the Northeast Atlantic) (IBPWHM). ICES Scientific Reports. 84 pp.

PELAC 2019 Minutes to the PELAC WHOM Focus Group 20 June 2019 Schiphol : https://www.pelagic-ac.org/media/afbeeldingen/Minutes%20WHOM%2020.06.2019_FINAL.pdf

ICES. 2020. Workshop on guidelines and methods for the evaluation of rebuilding plans (WKREBUILD). 2:55. 79 pp. . Volume 2, Issue 55. 79 pp pp.

ICES. 1989. Report of the Working Group on the assessment of pelagic stocks in Divisions VIIIc and IXa and horse mackerel. 10-19 May 1989, ICES Headquarters, Copenhagen. ICES CM 1989/Assess:19: 143pp.

ICES. 2019. Horse mackerel (*Trachurus trachurus*) in Subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a–c, and 7.e–k (the Northeast Atlantic). In Report of the ICES Advisory Committee, 2019. 1 October 2019. http://ices.dk/sites/pub/Publication%20Reports/Advice/2019/2019/hom.27.2a4a5b6a7a-ce-k8.pdf

Ref. Ares(2020)4284784 - 17/08/2020

WESTERN HORSE MACKEREL TECHNICAL FOCUS GROUP ON HARVEST CONTROL RULE EVALUATIONS 2020

M.A. Pastoors, A. Campbell, V. Trijoulet, D. Skagen, M. Gras, G.I. Lambert, C.R. Sparrevohn, S. Mackinson

14/8/2020

Contents

The Western horse mackerel technical Focus Group consisted of the following members:

Martin Pastoors, PFA (acting chair)

Andrew Campbell, MI, Ireland

Vanessa Trijoulet, DTU-AQUA, Denmark

Dankert Skagen, Fisheries Science Consultant, Norway

Michael Gras, MI, Ireland

Gwladys Lambert, CEFAS, England

Claus Sparrevohn, DPPO, Denmark

Steve Mackinson, SPFA, Scotland

The group met during the period June 2019 – August 2020 to collate information, carry out analyses and report findings that are embedded in the current report.

Executive summary

This report has brought together many different topics that are related to the western horse mackerel stock in an attempt to develop a potential rebuilding plan for the stock. Even though western horse mackerel was not classified by ICES as in need of rebuilding in their latest advice (ICES, 2019a), the general perception within the fishing industries has been that the stock has been in a poor state but that there have been some positive signals in recent recruitment. Using the new recruitments to improve the stock status requires a careful management approach. The PELAC has been a proponent of developing management plans for all stocks in their remit. In this case, the PELAC has termed the approach a rebuilding plan because of the current stock status of the stock.

Substantial progress has been made over the past few years on horse mackerel stock ID (Farrell et al., 2020). The full genome sequencing of horse mackerel from samples taken all the way from the Skagerrak to the Mediterranean and North Africa, has yielded a suitable panel of SNP markers that can be used to differentiate between the different horse mackerel stocks. The strongest differentiation between populations was between the northern and southern populations, with the boundary being in the middle of Portugal. The North Sea population is clearly distinct from the Western population and it should be possible to tell the difference from mixed samples with a high probability (>93%). This would also allow screening of catches in 7d and 7e on the contribution of western and North Sea populations. The separation between the northern and southern populations could mean that the current division between western and southern horse mackerel is not adequate, as the northern part of 9a is currently included in the southern population. A similar split in the middle of Portugal has also been observed for boarfish (Farrell et al., 2016) and could indicate a biogeographical feature.

Length compositions of the catches are an important element of the assessment approach for western horse mackerel, because Stock Synthesis uses length composition in combination with age-length key to estimate the age compositions within the model. Part of a rebuilding plan for western horse mackerel could be to evaluate differences in length compositions in the catches in certain areas and to take specific measures to protect incoming recruitment. Therefore, we planned to carry out an analysis of length compositions by area and season. However, we found that such data is not currently available for all years. Length data for western horse mackerel is currently not included in the ICES InterCatch database. Instead, length data has been processed on a year by year basis in non-standardized Excel spreadsheets. A time series of length compositions by area and season can therefore only be derived by manually working through the spreadsheets and extracting the required information. This was not feasible as part of the project to develop and evaluate a rebuilding plan for western horse mackerel. We recommend to WGWIDE that the full time series of catch at length by country is recreated from the Excel spreadsheets and input into InterCatch to allow for future interrogations of the data and an underpinning of the input data to the stock assessment.

In order to understand how a stock would respond to recovery measures, it is useful to consider the age composition in the spawning stock which illustrates how recruitment in the previous years contributed to the present spawning stock. To this end, an SSB per recruit analysis has been carried out. As one should expect for a relatively long-lived species with low mortality, the spawning stock is currently rather old. At F =0.075, the mean age is about 9 years, 80% is older than 5 years and 20% older than 12 years. So, an improved recruitment will take some time to materialize as increased SSB.

The current stock assessment method for western horse mackerel is Stock Synthesis 3, as agreed in the WKWIDE benchmark of 2017 (ICES, 2017b). Reference point were also set at WKWIDE 2017 but have subsequently been updated in the IBPWHM 2019 (ICES, 2019b). In addition, an exploratory SAM assessment has been carried out as part of IBPWHM 2019. This was done in order to get a second view on stock trends but also to be able to run the SAM HCR forecast as part of the development of a potential rebuilding plan. The exploratory SAM assessment (https://www.stockassessment.org/setStock.php?stock=WHOM2018) was initiated with the same input data as was used for the Stock Synthesis assessment of WGWIDE 2018 (ICES, 2018) with the exception of the length frequency data, which was not used. The PELACUS survey data was therefore only used as an index of biomass within SAM. The process of fine-tuning the assessment lead to the binding of the observation variances for certain variables and to the application of a fixed selectivity pattern (correlation coefficient ϱ=1 in the F random process (https://github.com/martinpastoors/wgwide/blob/master/R/HOM%20optimization_SAM.R). A comparison of Fbar and SSB between the SS3 assessments of WG2018 and 2019 with the SAM assessment (WG18SAM, WG19SAM), shows that the general trends are the same but that there are some deviations in certain periods (e.g. the SSB in the late 1980s is estimated substantially higher in SAM compared to SS3). The Stock Synthesis results are in general a bit smoother compared to SAM.

In order to be able to use the SAM assessment as an alternative assessment in the rebuilding plan evaluation, we needed to estimate reference point for this assessment. In doing so, we aimed to follow the same procedure as during IBPWHM 2019 (ICES, 2019b). However, one of the elements of the reference point estimation, triggered a more in-depth study: the role of assessment uncertainty parameter Fcv and Fphi. There has been little standardization in how Fcv and Fphi have been calculated in different benchmarks where reference points were estimated. Fcv is expected to capture the assessment error in the advisory year and Fphi is the autocorrelation in assessment error in the advisory year (ICES, 2014a). We documented the method for generating the input data for the calculations and explored the sensitivity of Fcv and Fphi to the assessment that was used (both for western horse mackerel and for Atlantic mackerel). We found that there can be a high dependence of Fphi on the assessment that is used to compare against the Fset (the fishing mortalities that are back-calculated from the observed catches and the annual forecasts). When the assessment that is used has values that are all higher or lower than the Fset values, then Fphi will be close to zero. To our knowledge, this behaviour of Fphi was unknown so far. We also found that the number of years that is used for calculating Fcv and Fphi may have an impact on the values. In the recommendations from WKMSYREF3 it is stated that 10 years (or more) should be taken. A further study should be undertaken to assessment the impacts of using different time periods for estimating Fcv and Fphi.

During the IBPWHM 2019, reference points were estimated for western horse mackerel based on the 2018 WGWIDE assessment and using default values for Fcv and Fphi (0.212 and 0.423) and using a segmented regression through Blim (segregBlim). In order to calculate reference points for the exploratory SAM assessment and to explore the sensitivity to the assessment year, reference points were calculated on the basis of the 2018 or 2019 assessments for SS and SAM. The reference points for the SAM assessment are based on the 2018 assessment. Bpa and Blim are lower than the values for the SS assessment, while the Fmsy is higher. The calculated reference points were not sensitive to the assessmentyear that was used for the calculation for both the SS and SAM assessments.

Note that the calculated value for FMSY_final for the 2018 SS WGWIDE option (0.079) differs slightly from the value in IBPWHM 2019 (0.074). While a full explanation for this difference could not be arrived at, it is expected that this could have to do with the random seed and the instability of some of the calculations.

RP	WG18	WG18SAM	WG19	WG19SAM
Blim	834480	611814	885341	612635
Flim	0.1107	0.1612	0.1049	0.1756
Fpa	0.07909	0.1152	0.07493	0.1254
MSYBtrigger	1168272	856540	1239478	857689
FMSY	0.09102	0.1262	0.08665	0.1353
FP05	0.08398	0.1255	0.07826	0.1402
FMSY_final	0.07909	0.1152	0.07493	0.1254

HCR evaluations

The HCR analyses represent two different assessment methods (SS3 and SAM) and two different HCR evaluation tools (EqSim and SAM HCR). Both HCR evaluation tools are of the type 'short-cut' with appropriate conditioning of the uncertainties in the assessment based on historical CV and autocorrelation in line with the recommendations from WKMSYREF3 and WKMSYREF4 . The evaluations followed the guidelines from WKGMSE2 (ICES, 2019c) and WKREBUILD (ICES, 2020).

Three different types of harvest control rules were evaluated:

- Constant F strategy: fixed Ftarget independent of biomass level
- ICES Advice Rule: breakpoint at Btrigger and straight decline in F to zero below Btrigger.
- Double Breakpoint rule: breakpoint at Btrigger and straight decline in F to 20% of Ftarget at Blim. Below Blim continued fishing at F = 0.2 * Ftarget.

For each of the HCRs, a number of different target fishing mortalities were explored (0.0, 0.05, 0.075, 0.1, 0.125, 0.15). No evaluation of different Btrigger values was carried out, so that all evaluations used MSY Btrigger as the trigger point. All HCRs where evaluated with three variants:

- Without any additional constraints
- With a minimum TAC of 50 kT
- With a maximum 20% inter-annual variation (IAV) in TAC, but only when the stock is above Btrigger)

Two simulation tools were used: the EqSim simulator and the SAM HCR forecast. The EqSim simulator is a further worked up version of the SimpSIM approach that was used for the blue whiting MSE in 2016 (ICES, 2016). The code was further developed by Andrew Campbell and Martin Pastoors to improve standardization, documentation and visualization of results. EqSim makes use of an Operating Model (OM) and a Management Procedure (MP). The SAM HCR forecast is a simple stochastic forecast with HCR to evaluate management for fish stocks that need rebuilding in the short-term. The stochastic forecasts start from what we believe is the current level of the stock, i.e. the assessment estimates currently used for tactical management advice, with consideration of the uncertainty in these estimates. Rebuilding is evaluated forward for a specified number of years and for different target fishing mortality values.

The EqSim with SS3 results indicate that the constant F strategy is the least cautious rule and the double breakpoint rule is the most cautious rule. Under the F strategy rule with a Ftarget of 0.075, rebuilding to Bpa is only just being achieved (probability just above 50%) by 2025, while in the double breakpoint rule this is expected to be achieved in 2024 with substantially higher probabilities of remaining above Bpa. The first year of rebuilding to Bpa in the double breakpoint rule with target fishing mortalities up to 0.1 is the same as the first year of rebuilding under the zero fishing scenarios.

Similar results have been obtained with the EqSim with SAM evaluations although the levels of SSB are slightly higher and risk to Blim is slightly lower. According to these evaluations, rebuilding to Bpa could be obtained by 2022 in all scenarios.

The SAM HCR with SAM evaluations have only been carried out for the ICES Advice Rule scenario, as this was intended more as a contrasting study rather than a full analysis of HCR evaluation. Again, we find similar patterns in simulated stock trends, but SSB is estimated higher in the SAM evaluation than in the EqSim evaluations and risk to Blim stays below the 0.05 threshold in SAM HCR for all target fishing mortalities that have been explored.

Given that the EqSim with SS3 evaluation is closest to the ICES advisory practice, this was used as the basis for the preferred rebuilding plan by the PELAC. The PELAC preferred options are:

- Target fishing mortality at Fmsy = 0.074 (approximated by 0.075 in the simulations)
- Blim at ICES Blim (834 480 t)
- Btrigger at ICES MSY Btrigger (1 168 272 t)
- Double breakpoint rule with 20% constraint on IAV above Btrigger
- Minimum F when stock is below Blim at 20% of Fmsy = 0.015

The selected rebuilding plan has a 50% probability of rebuilding to Blim by 2021 (similar to zero catch option) and a 50% probability of rebuilding to Bpa/MSY Btrigger by 2024 (similar to the zero-catch option). Furthermore, the probability of being below Blim remains well below 5% for the duration of the simulation.

In this scenario, the average catch in the years 2021-2025 is expected to be lower than recent catches. However, after rebuilding, catches should be able to be maintained around 100 000 tonnes.

1 Introduction

1.1 Challenge

The Western Horse mackerel Focus Group of the Pelagic Advisory Council (PELAC) has been set up in 2015 already to a develop a PELAC proposal for a rebuilding plan or management plan for Western Horse mackerel. After several iterations (see below), the Focus Group initiated a technical working group to develop an operational evaluation tools for management plan evaluation and to evaluate potential Harvest Control Rules, so that PELAC could come to a recommended procedure. Such a recommended procedure, including the evaluation that was carried out, would need to be submitted for review to ICES to establish whether the evaluation procedure is in line with scientific standards and that the results of the HCR are in conformity with the precautionary approach and the MSY approach.

1.2 What happened before

An overview is presented of the attempt to develop a management plan for Western horse mackerel in the ICES area. After an initial egg-survey based management rule had been agreed and evaluated in 2008 (ICES, 2008), the management plan was called into question in 2011 which lead to the statement by ICES in 2013 that the plan was no longer precautionary (ICES, 2013a). In the years 2014-2015, CEFAS and the Marine Institute were commissioned by the Pelagic Regional Advisory Committee to evaluate potential new management plans (Campbell et al., 2015). The SAD assessment that was used to assess the stock in those years, and that underpinned the MSEs for Western horse mackerel, was so uncertain, that the results were that in the case of no-fishing, the stock was expected to increase, but the uncertainty in the stock was also increasing, to the effect that the probability of being below B_{lim} was larger than 5% for the next 40 years to come. Apparently, the framing of those MSEs could not resolve to a meaningful and acceptable management plan.

A second iteration occurred after the stock had been benchmarked in 2017 and was using the Stock synthesis model for the assessment (ICES, 2017). Using the methods described by Cox et al. (Cox and Kronlund, 2008), a proof-of-concept full-feedback MSE[1] was commissioned with Landmark Fisheries Research, Canada (Cox et al., 2018). The evaluations were directed at different fishing strategies, including strategies where fishing would continue when the biomass would be below B_{lim}. The results of the analysis demonstrated a clear recovery potential of the stock under different fishing scenarios, mostly dependent on the recruitment assumptions and the target fishing mortality. However, the starting conditions of the simulated populations did not include uncertainty, and therefore the behaviour of the MSE may have been estimated too positively.

For a final iteration of the management plan evaluation, it was anticipated to use the guidelines from WKGMSE2 (ICES, 2019c) and WKREBUILD (ICES, 2020) to plan for the next step in the development of the management plan. This work is embedded in the current report.

[1] A full-feedback MSE means that the assessment (and forecast) are run within the Management Strategy Evaluation (MSE) framework for each year and for each iteration.

1.3 Approach

The approach during the Focus Group on Western Horse mackerel was to convene a number of physical meetings to identify the main issues and to plan regular updates. In June 2019, a technical subgroup was set up to further carry out the technical analyses that were required. This subgroup was closely affiliated with the ICES WKREBUILD workshop that was going to take place in February 2020.

The first technical subgroup meeting was held on 20-21 June 2019. After presenting the state of affairs during WKREBUILD 2020, a series of online meetings was held during May and June 2020 to finalize the evaluation tools and to carry out the studies and evaluations. Specific focus was paid to the following topics:

- Stock ID (through the genetic work coordinated by Edward Farrell, UCD)
- Analysis of length compositions of catches (Gwladys Lambert, Martin Pastoors)
- Analysis of SSB per recruit (Dankert Skagen)
- Stock assessment (with focus on exploratory SAM assessment; Vanessa Trijoulet and Martin Pastoors)
- Reference points and calculation of Fcv and Fphi (Martin Pastoors)
- Development of HCR evaluation tools
    - EqSim (Andrew Campbell, Martin Pastoors)
    - SAM HCR (Vanessa Trijoulet)
- Application of HCR tools to evaluate different potential rebuilding plan (Andrew Campbell, Vanessa Trijoulet, Martin Pastoors)
- Presentation of results to the PELAC western horse mackerel focus group (Martin Pastoors, Andrew Campbell)

2 Horse mackerel stock ID

Recently, a study has been completed on the population structure of the Atlantic horse mackerel (*Trachurus trachurus*) as revealed by whole-genome sequencing (Farrell et al., 2020). The executive summary of that report is repeated below:

"The Atlantic horse mackerel, Trachurus (Linnaeus, 1758) is a species of jack mackerel distributed in the East Atlantic, from Norway to west Africa and the Mediterranean Sea. It is a pelagic shoaling species found on the continental shelf and it is one of the most widely distributed species in shelf waters in the northeast Atlantic, where it is targeted in pelagic fisheries. In the northeast Atlantic region, the species is assessed and managed as three stocks: the Western, the North Sea and the Southern. Despite the commercial importance of the horse mackerel, the accuracy of alignment of these stock divisions with biological units is still uncertain.

The aims of this study were to identify informative genetic markers for the stock identification of horse mackerel and to estimate the extent of genetic differentiation among populations distributed across the distribution range of the species. For this we used modern sequencing techniques that allowed us to assess genetic variants in the entire genome. We discovered that while the populations differ in a small fraction of their DNA (< 1.5%), such genetic differences are significant as they likely represent natural selection and might be involved in local adaptation. We validated a small fraction of these highly differentiated genetic variants by a SNP assay and demonstrated that they can be used as informative molecular markers for the genetic identification of the main stock divisions of the Atlantic horse mackerel.

The results, based on the analysed samples, indicated that the North Sea horse mackerel are a separate and distinct population. The samples from the Western stock, west of Ireland and the *northern Spanish shelf, and the northern part of the Southern stock, northern Portugal, appear to form a genetically close group. There was significant genetic differentiation between the northern Portuguese samples and those collected in Southern Portuguese waters, with those in the south representing a separate population. The North African and Alboran Sea samples were distinct from each other and from all other samples.*

These results indicate that a further large-scale analysis of samples, with a greater temporal and spatial coverage, with the newly identified molecular markers is required to test and reassess the current stock delineations."

The main conclusions of the genetic work can be summarized as follows:

- A suitable panel of SNP markers can be identified to carry out routine population assignments of mixed samples.
- Main differentiation between populations is between northern and southern populations, with the boundary being in the middle of Portugal. Although more work needs to be done on this finding, this could imply that the current division between western and southern horse mackerel is not adequate, as the northern part of 9a is currently included in the southern population.
- The North Sea population is clearly distinct from the Western population and it should be possible to tell the difference from mixed samples with a high probability (>93%?). This allows screening of catches in 7d and 7e on the contribution of western and North Sea populations.

3 Length compositions of catches

A short study was initiated to analyse the length composition of catches by country, area, year and quarter. Length compositions could be informative on selectivity in different areas and fisheries and could therefore also be used to generate specific management measures as part of a rebuilding plan.

In the current SS assessment framework, length compositions are used as the key metric for catches in combination with age-length keys to generate age compositions dynamically. So, while it might be expected that the length information is readily available, this turned out to be not the case. The length data that is submitted by country, is not submitted in a standardized format and not included in the InterCatch database. Historical length data by country has been processed on an annual basis using ad hoc Excel spreadsheets and cannot be easily extracted. Therefore, no real progress has been made on this topic.

Recommendation:

- The Western Horse Mackerel Focus Group recommends to WGWIDE that the full time series of catch at length by country is recreated from the Excel spreadsheets and converted into InterCatch to allow for future interrogations of the data and an underpinning of the input data to the stock assessment.

4 Contribution of recruitment to SSB

Dankert W. Skagen, June 2020

For the understanding of how a stock responds to recovery measures, it is useful to consider the age composition in the spawning stock, to illustrate how recruitment in the previous years contribute to the present spawning stock. When we calculate SSB per recruit, we do this by calculating the sequence of numbers at age as they are reduced by mortality, starting with one recruit. Then we multiply numbers at each age with weight and maturity at that age to get biomass per recruit of the spawners at each age. The sum of these over all ages is the total SSB per recruit, which is normally what is presented, but the age profile of the SSB per recruit can also be interesting in itself. For example, when we consider a rebuilding strategy, it gives us an indication of how fast SSB can be expected to improve when recruitment improves. The age distribution in the spawning stock of course depends on the fishing mortality level, as does the total SSB per recruit.

The actual SSB at some age is the SSB per recruit at that age, multiplied with the number of recruits born in that cohort. Accordingly, the total SSB in any year is a weighted sum of previous recruitments. The products of cohort recruitment times SSB per recruit at age, summed over all ages. In an equilibrium where all weighting factors are constant, SSB is proportional to the mean recruitment, since it is the sum of SSB per recruit at age, raised by the recruitment.

This simple relation also gives us an easy direct means of calculating how the variation in recruitment carries over to variation in SSB. In probability theory, there is a very simple formula for variance of a weighted sum of independent components. Here the components are annual recruitment, with a presumably known variance, and the weightings are the SSB per recruit at age. Although this only covers the effect of one source of variation in SSB, the recruitment variation is a major source so a direct calculation of the variance, without elaborate bootstrap procedures, can be useful as a proxy in the early phase of management plan developments, and also for understanding the effect of variable recruitment.

Below is a set of age distributions in the SSB per recruit for Western horse mackerel (Figure 2). The data on weights, maturities, natural mortality and selection were those used as input to the short-term prediction by WGWIDE in 2019.

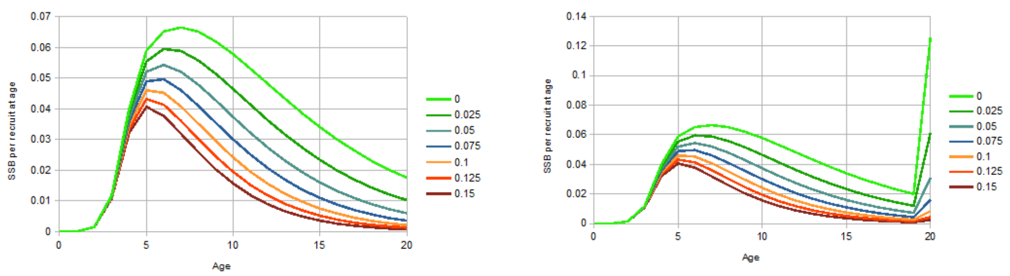


Figure 1 *SSB at age for a range of fishing mortalities (F1-10) With (right) and without (left) regarding age 20 as a plus group.*

Figure 3 shows SSB per recruit as function of F1-10, with the same input data, and in addition the 95 % confidence interval assuming a CV on recruitment of 0.6. which is slightly lower than the CV of the recruitments 1983 – 2018 according to the WGWIDE assessment in 2019, excluding the strong 2001 year class. In the same figure, the mean age in the SSB as function of the F1-10 is also shown.

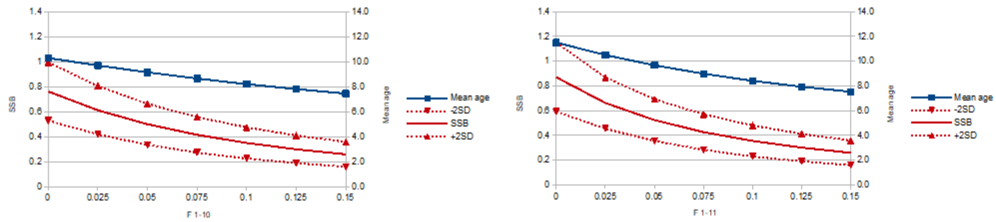


Figure 2 *Mean age (blue) and SSB (Mean ±2SD) for a range of fishing mortalities (1-10). Using only age up to 20 (left, without a plusgroup) and using all ages (right, with a plusgroup at 20). The SDs are the effect of recruitment variation, assuming a CV of 0.6*

As one should expect for a relatively long-lived species with low mortality, the spawning stock is rather old. At F =0.075, the mean age is about 9 years, 80% is older than 5 years and 20% older than 12 years. So, an improved recruitment will take some time to materialize as increased SSB. The results also indicate that with a low F, the plus group still does matter. Finally, the historical variation in recruitment translates into a confidence interval for long term equilibrium SSB that for F=0.075 ranges from approximately 700 to 1400 when the mean recruitment is 2500.

5 Stock assessment of Western horse mackerel

5.1 Stock synthesis assessment

WGWIDE 2019: The SS model with new length and age data from the commercial fleet, and the 2018 information from the 2 surveys available, is presented as the final assessment model. Stock numbers-at-age and fishing mortality-at-age are given in Tables 7.3.1.1 and 7.3.1.2, and a stock-summary is provided in Table 7.3.1.3 and illustrated in Figure 7.2.11.2. SSB peaked in 1988 following the very strong 1982 year class. Subsequently SSB slowly declined till 2003 and then recovered again following the moderate-to-strong year class of 2001 (a third of the size of the 1982 year class). Year classes following 2001 have been weak: 2010 2011, and 2013 recruitments in particular have been estimated as the lowest values in the time-series together with the 1983. The 2008 year class has been estimated to be fairly strong. Recruitment estimates for 2014-2018 are the highest observed since 2008 and are higher than the geometric mean estimated over the years 1983-2018. SSB in 2017 is estimated as the lowest in the time-series. Fishing mortality was increasing after 2007 as a result of increasing catches and decreasing biomass as the 2001 year class was reduced. Since 2012 F has then been decreasing, dropping to low values in 2015-2018 due to lower catches and a reduced proportion of the adult population in the exploited stock.

5.2 SAM assessment

IBPWHM 2019: Since the benchmark in 2017 (ICES, 2017b), the Western horse mackerel assessment has been carried out using the Stock Synthesis method. This method allows for the incorporation of length frequency information and the dynamic estimation of growth. The Stock Synthesis assessment of western horse mackerel utilizes the length distributions of the commercial catch and from the samples obtained during the PELACUS survey, while the other information is provided as biomass (total catch, egg survey) or age specific data (recruitment index). The SS assessments that have been carried out since the benchmark in 2017 have generally shown narrow confidence intervals, yet the annual revisions in estimated stock size and fishing mortality between subsequent assessments has been substantial. These retrospective revisions are not well understood. In addition, there has been some concern about the complex nature of the input data to the Stock Synthesis method and the ability to adequately quality control the input data and model performance.

As part of the Interbenchmark of Western horse mackerel, it was agreed to explore the possibility of an alternative assessment approach to Stock Synthesis. The intention was to test methods that are more familiar to members of the WGWIDE assessment group. It was decided to use the SAM model as the alternative approach because it is already being used for mackerel and blue whiting and because it will allow for an evaluation of harvest control rules in a similar manner as is currently being applied for Western Baltic Spring Spawning herring.

The exploratory SAM assessment (https://www.stockassessment.org/setStock.php?stock=WHOM2018) was initiated with the same input data as was used for the Stock Synthesis assessment of WGWIDE 2018 (ICES, 2018) with the exception of the length frequency data, which was not used. The PELACUS survey data was therefore only used as an index of biomass within SAM. When using the default SAM configuration, the assessment output displayed a strong retrospective pattern and very large uncertainty in both F and SSB. A process of fine-tuning the assessment lead to the binding of the observation variances for certain variables and the application of a fixed selectivity pattern (correlation coefficient ϱ=1 in the F random process, that was originally allowed to change by year (https://github.com/martinpastoors/wgwide/blob/master/R/HOM%20optimization_SAM.R). The only aged-structured observation available for this stock is for the commercial catch. As a result, the model has a tendency to over-fit these observations, notably for the older ages. This induced important variations in fishing selectivity over time that seemed inconsistent and led to very large retrospective patterns in both SSB and F. Fixing the fishing selectivity over time resulted in a significant improvement in these retrospective patterns for only a slightly larger AIC (1217.453 vs. 1212.974 with variable relative fishing mortality). The final exploratory assessment from this exercise was selected on the basis of the trade-off between a low AIC and reduced retrospective pattern.

A comparison of Fbar and SSB between the SS3 assessments of WG2018 and 2019 with the SAM assessment (WG18SAM, WG19SAM).

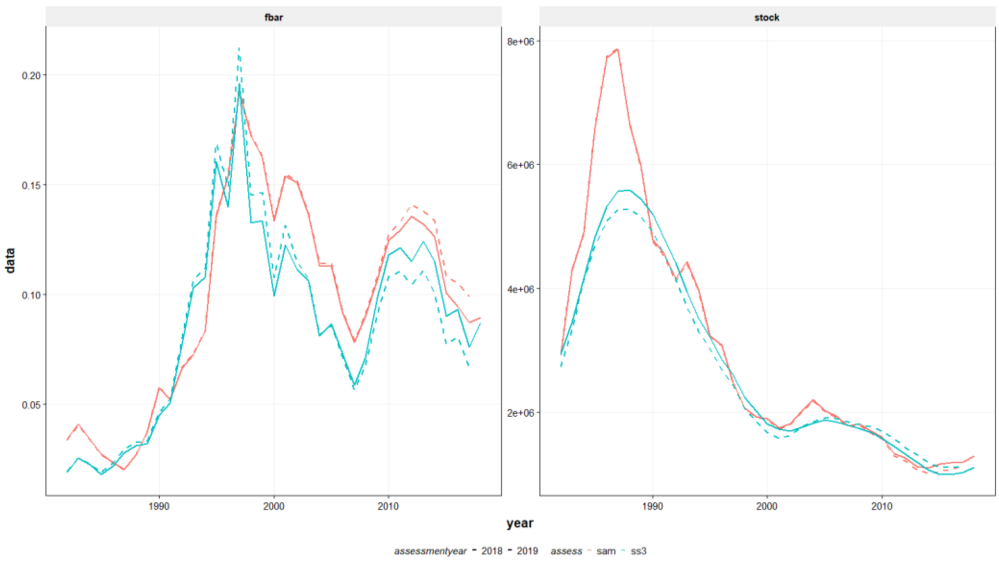


Figure 3 Time trends for Fbar and SSB for the SS3 (red) and SAM (blue) assessments for WG2018 and 2019.

6 Fcv and Fphi uncertainty parameters

The standard approach in ICES for estimating biological reference points is based on the EqSim software conditioned on the most recent assessment. Uncertainties in the assessment are included through two parameters: Fcv and Fphi, where Fcv is expected to capture the assessment error in the advisory year and Fphi is the autocorrelation in assessment error in the advisory year (ICES, 2014a). Methods for deriving Fcv and Fphi are loosely described in the WKMSYREF3 report (ICES, 2014a, p. 11):

"The estimated realised catch and F (Fyr) for the previous 10 years (or more) are taken from the most recent assessment. The annual ICES advice sheets issued in y-1 are consulted to estimate the F_{ya} that would have been advised to obtain the estimated catch. Where the appropriate catch is not available in the catch option table linear interpolation is used to estimate the F_{ya}. The deviation in year y d_y is calculated as $\log_e(F_{yr}/F_{ya})$, the standard deviation σ_m of the log deviations gives the marginal distribution. The conditional standard deviation σ_c is calculated as $\sigma_m \sqrt{(1-\varphi^2)}$, where φ is the autocorrelation of the AR(1) process. Then σ_c [and] φ are input parameters for Eqsim."

The role of Fcv and Fphi in the process of estimating reference points is that they are used to calculate Fp05 which is used as the precautionary buffer on Fmsy, because Fp05 is the value whereby a (less than) 5% annual probability exists that SSB will be below Blim in the long term If the directly estimated Fmsy is larger than Fp05, then Fmsy needs to be reduced to Fp05.

When applying this approach to the western horse mackerel data, we found that there were important sensitivities in calculating the parameters Fcv and Fphi. This initial finding let us to carry out a broader review of the behaviour of Fcv and Fphi for a number of widely distributed pelagic stocks where reference points were recently estimated (western horse mackerel and Atlantic mackerel). The results will be summarized in a working document to ACOM in September 2020. While there has in general been ample attention during benchmark workshops to the estimation of reference point – albeit they are often carried out AFTER the benchmark instead of DURING the benchmark – we found that the documentation of the selection of data and the method to calculate the Fcv and Fphi has been mostly lacking. In most cases it is not clear how many years have been used, nor how the values for the interpolated fishing mortalities have been generated.

Western horse mackerel

Fset and SSBset were calculated from the historical assessment data. Realized catch by year was taken from the most recent advice document. Catch1fcy and Catch2fcy are the two catch options that bracket the actual realized catch in the forecast year and F1fcy and F2fcy are the associated fishing mortalities. Fset is the interpolated fishing mortality that matches the realized catch in a particular forecast.

In the case of horse mackerel, this procedure could not be followed for estimating the SSBset, because only one value of SSB in the forecast year is presented in the forecast tables.

tacyear	catchrealized	catch1fcy	catch2fcy	f1fcy	f2fcy	ssb1fcy	ssb2fcy	fset	ssbset
2011	193268	186433	201312	0.1048	0.1135	-	-	0.108797	1911900
2012	166579	155125	174007	0.0944	0.1064	-	-	0.101679	1879742
2013	165258	155633	170000	0.1638	0.18	-	-	0.174653	1568380
2014	136360	129640	144621	0.1541	0.1734	-	-	0.162757	749334
2015	98419	85820	99304	0.1053	0.1229	-	-	0.121745	601099
2016	98811	98544	99710	0.0997	0.1009	-	-	0.099975	718285
2017	82961	82526	84289	0.1105	0.113	-	-	0.111117	511789
2018	101682	99129	108515	0.081	0.089	-	-	0.083176	818082

The calculation of cv and phi for fishing mortality and SSB is shown below (figure 4). Fassess and SSBassess are taken from the WGWIDE 2019 assessment. The explanations below are only given for fishing mortality, but the same procedures apply to SSB.

The F deviation in year y d_y is calculated as ln(Fassess/Fset). The standard deviation σ_m (=lnSTD) of the log deviations gives the marginal distribution. The autocorrelation in the log deviations φ (=Fphi) is calculated by correlating the deviations 2011-2017 with the deviations 2012-2018 (this is the autocorrelation of the AR(1) process). The conditional standard deviation σ_c (=Fcv) is calculated as $\sigma_m \sqrt{(1-\varphi^2)}$.

In the case of western horse mackerel, Fcv is estimated at 0.2193 and Fphi at the very low value of 0.0212. This can be explained by the almost complete lack of overlap between Fassess and Fset because the most recent assessment estimates a substantially lower fishing mortality than was assumed in the forecasts. The F correlation plot below therefore shows a close to flat line. During IBPWHM 2019, reference points have been calculated using Fcv = 0.212 and Fphi = 0.423 (the default EqSim values) and thus substantially different from the calculated values.

Note that SSBcv and SSBphi have been calculated in the same way, but they are not currently used in the EqSim approach for estimating reference points.

A simulation study on the impact of different values of Fcv and Fphi on the Fmsy for western horse mackerel is shown below (figure 5). Fcv is on the horizontal axis, while the coloured lines indicate the values of Fphi. The five panels demonstrate the five steps in arriving at the final Fmsy.

- Estimate Fmsy without constraints
- Calculate Fpa (has been done previously).
- If Fmsy is larger than Fpa, set Fmsy_interim to Fpa
- Calculate Fp05 with Eqsim using Fcv, Fphi and Blim
- The final Fmsy is the minimum of Fp05 and Fmsy_interim.

The simulation study demonstrates that a larger Fcv leads to a lower Fp05 and also that a larger Fphi leads to the Fp05 being more sensitive to the impact of Fcv. Therefore, the estimated values of Fcv and Fphi can have an important impact on the Fmsy that is calculated in EqSim.

Horse mac Fishing mortality

Year	Fassess	Fset	ln(Fass)	ln(Fset)	Ln(Fassess/Fset)
2009					
2010					
2011	0.1212	0.1088	-2.1103	-2.2183	0.1080
2012	0.1151	0.1017	-2.1622	-2.2859	0.1237
2013	0.1242	0.1747	-2.0855	-1.7450	-0.3405
2014	0.1151	0.1628	-2.1622	-1.8155	-0.3467
2015	0.0901	0.1217	-2.4074	-2.1058	-0.3016
2016	0.0930	0.1000	-2.3755	-2.3028	-0.0727
2017	0.0759	0.1111	-2.5778	-2.1972	-0.3807
2018	0.0871	0.0832	-2.4409	-2.4868	0.0459

lnSTD	0.2194
Fcv	0.2193
Fphi	0.0212

Horse mackerel SSB

Year	SSBassess	SSBset	ln(SSBass)	ln(SSBset)	Ln(SSBassess/SSBset)
2009					
2010					
2011	1211620	1911900	14.0075	14.4636	-0.4561
2012	1164950	1879742	13.9682	14.4466	-0.4785
2013	1087630	1568380	13.8995	14.2656	-0.3660
2014	955525	749334	13.7700	13.5269	0.2431
2015	838866	601099	13.6398	13.3065	0.3333
2016	786772	718285	13.5757	13.4846	0.0911
2017	761613	511789	13.5432	13.1457	0.3975
2018	811685	818082	13.6069	13.6147	-0.0079

lnSTD	0.3586
SSBcv	0.2927
SSBphi	0.5776

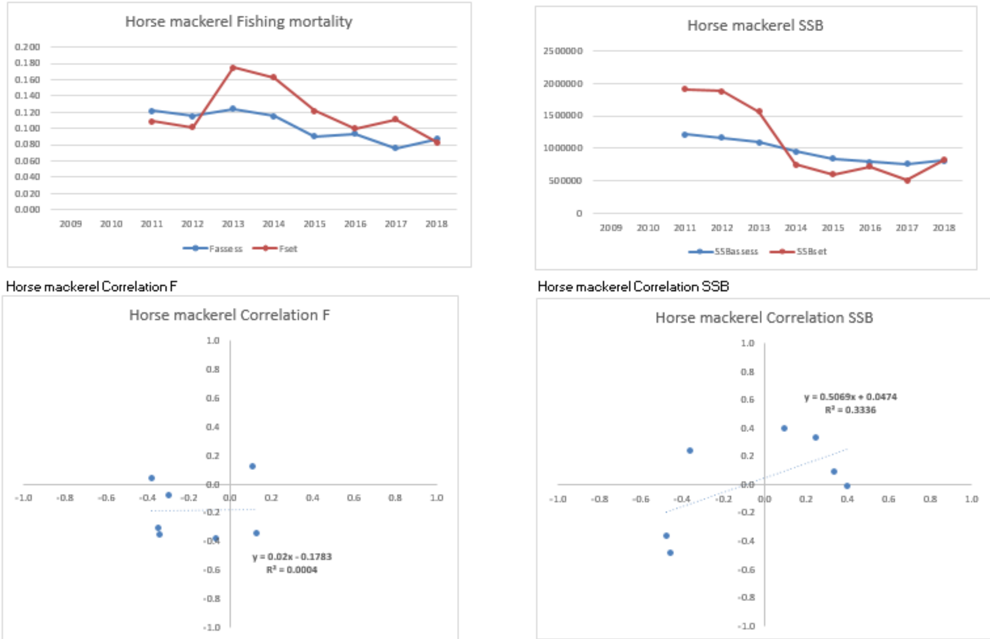


Figure 4 Calculation of Fcv, Fphi, SSBcv and SSBphi for western horse mackerel

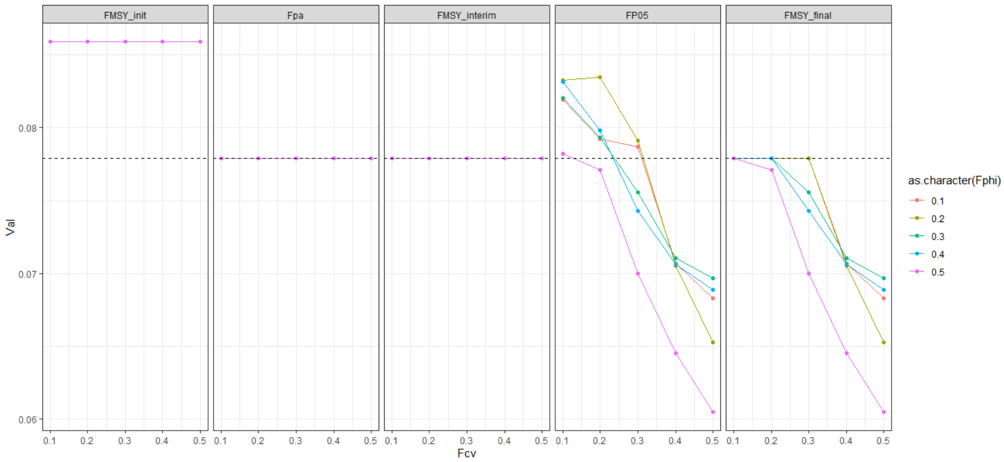


Figure 5 *Simulated values of the impact of Fcv and Fphi on the reference points for western horse mackerel.*

Atlantic mackerel

Following the same procedure as outlined above, we obtained the following values for Fset and SSBset for Atlantic mackerel.

tacyear	catchrealized	catch1fcy	catch2fcy	f1fcy	f2fcy	ssb1fcy	ssb2fcy	fset	ssbset
2009	737969	707000	831000	0.25	0.3	2891000	2842000	0.262488	2878762
2010	875515	726000	996000	0.29	0.42	2397000	2293000	0.361989	2339409
2011	946661	884093	959773	0.31	0.34	2697368	2668541	0.334802	2673535
2012	892353	742000	927000	0.26	0.34	2710000	2638000	0.325018	2651484
2013	931732	930000	1116000	0.41	0.51	2390000	2310000	0.410931	2389255
2014	1393000	1300000	1400000	0.291	0.318	4594000	4573000	0.31611	4574470
2015	1208990	1054000	1396000	0.26	0.36	4344000	4276000	0.305319	4313183
2016	1094066	960009	1235608	0.28	0.38	3766022	3712034	0.328642	3739761
2017	1155944	1067828	1281394	0.28	0.35	4398536	4358095	0.308882	4381850
2018	1026437	977765	1122906	0.405	0.48	3043254	3013235	0.430151	3033187

In the case of mackerel, we were particularly interested in the effect of the assessment year on the calculation of Fcv and Fphi because of the substantial change in perception between the 2018 and the 2019 assessments. Therefore, we calculated Fcv and Fphi for each assessment year separately.

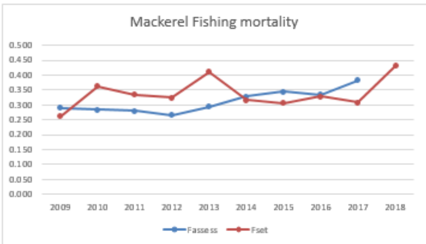
Similar to the observations for Western horse mackerel, the impact of the final assessment year is noticeable here. Due to the revision of the assessment in 2019, there is almost no overlap between the fishing mortalities from the assessment and those derived from the historical forecasts. This impacts on the estimated Fphi (0.3080 using the 2018 assessment, 0.0076 using the 2019 assessment).

MACKEREL 2018

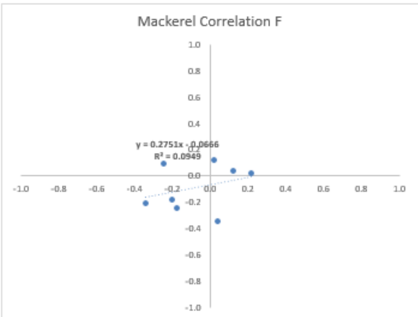
Mackerel Fishing mortality

Year	Fassess	Fset	ln(Fass)	ln(Fset)	Ln(Fassess/Fset
2009	0.290	0.2625	-1.2389	-1.3376	0.0986
2010	0.283	0.3620	-1.2626	-1.0161	-0.2464
2011	0.280	0.3348	-1.2744	-1.0942	-0.1801
2012	0.265	0.3250	-1.3291	-1.1239	-0.2052
2013	0.293	0.4109	-1.2282	-0.8893	-0.3389
2014	0.329	0.3161	-1.1117	-1.1517	0.0400
2015	0.345	0.3053	-1.0647	-1.1864	0.1217
2016	0.33507	0.3286	-1.0934	-1.1128	0.0194
2017	0.38298	0.3089	-0.9598	-1.1748	0.2150
2018		0.4302			

lnSTD	0.1929
Fcv	0.1835
Fphi	0.3080



Mackerel Correlation F

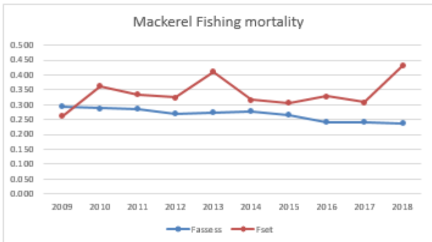


MACKEREL 2019

Mackerel Fishing mortality

Year	Fassess	Fset	ln(Fass)	ln(Fset)	Ln(Fassess/Fset
2009	0.294	0.2625	-1.2242	-1.3376	0.1134
2010	0.288	0.3620	-1.2448	-1.0161	-0.2287
2011	0.286	0.3348	-1.2518	-1.0942	-0.1575
2012	0.270	0.3250	-1.3093	-1.1239	-0.1855
2013	0.273	0.4109	-1.2983	-0.8893	-0.4090
2014	0.278	0.3161	-1.2801	-1.1517	-0.1285
2015	0.265	0.3053	-1.3280	-1.1864	-0.1416
2016	0.241	0.3286	-1.4230	-1.1128	-0.3102
2017	0.241	0.3089	-1.4230	-1.1748	-0.2482
2018	0.238	0.4302	-1.4355	-0.8436	-0.5919

lnSTD	0.1865
Fcv	0.1865
Fphi	0.0076



Mackerel Correlation F

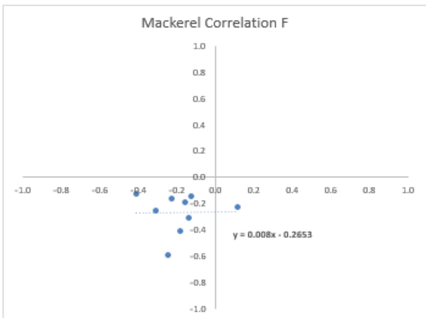


Figure 6 Comparison of Fcv and Fphi for Mackerel based on the assessments of 2018 and 2019.

Conclusions

While an elaborate procedure has been outlined to derive reference points for category 1 and 2 stocks in ICES (ICES, 2017a) based on the work of MSYREF workshops (ICES, 2013b; ICES, 2014a; ICES, 2014b; ICES, 2015), we conclude from our studies on western horse mackerel and Atlantic mackerel that insufficient attention has been given to the method of estimating forecast uncertainty and the impact of that uncertainty on the estimated reference points (notably Fmsy). Here we started with a method for documenting how the Fset is being derived from the historical data, so that at least the estimates of Fcv and Fphi are transparent and can be recreated.

We also note that there can be a high dependence of Fphi on the assessment that is used to compare against the Fset. When the assessment that is used has values that are all higher or lower than the Fset values, then Fphi will be close to zero. To our knowledge, this behaviour of Fphi was unknown so far.

Finally, we note that the number of years that is used for calculating Fcv and Fphi may have an impact on the values. In the recommendations from WKMSYREF3 it is stated that 10 years (or more) should be taken. A further study should be undertaken to assessment the impacts of using different time periods for estimating Fcv and Fphi.

7 Estimation of reference points for SS and SAM assessments

During the IBPWHM 2019, reference points were estimated for western horse mackerel based on the 2018 WGWIDE assessment and using default values for Fcv and Fphi (0.212 and 0.423) and using a segmented regression through Blim (segregBlim). In order to calculate reference points for the exploratory SAM assessment and to explore the sensitivity to the assessment year, reference points were calculated on the basis of the 2018 or 2019 assessments for SS and SAM.

The reference points for the SAM assessment are based on the 2018 assessment. Bpa and Blim are lower than the values for the SS assessment, while the Fmsy is higher. These values will be used in the subsequent evaluations (section 8)

The changes due the assessment year were minor for both the SS and SAM assessments.

RP	WG18	WG18SAM	WG19	WG19SAM
Blim	834480	611814	885341	612635
Flim	0.1107	0.1612	0.1049	0.1756
Fpa	0.07909	0.1152	0.07493	0.1254
MSYBtrigger	1168272	856540	1239478	857689
FMSY	0.09102	0.1262	0.08665	0.1353
FP05	0.08398	0.1255	0.07826	0.1402
FMSY_final	0.07909	0.1152	0.07493	0.1254

8 HCR evaluations

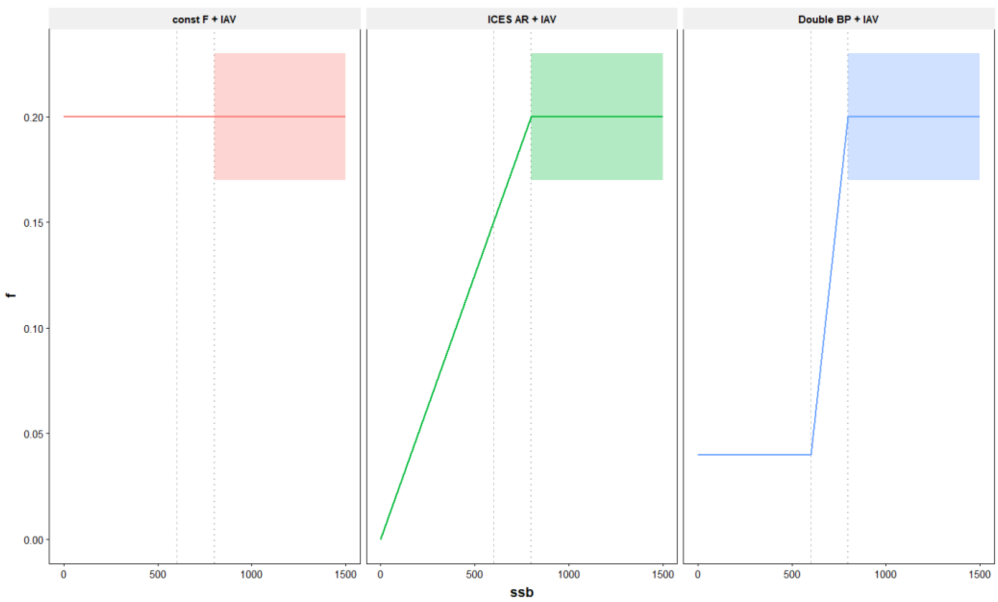
8.1 Type of HCRs evaluated

Three different types of harvest control rules were evaluated:

- Constant F strategy: fixed Ftarget independent of biomass level
- ICES Advice Rule: breakpoint at Btrigger and straight decline in F to zero below Btrigger.
- Double Breakpoint rule: breakpoint at Btrigger and straight decline in F to 20% of Ftarget at Blim. Below Blim continued fishing at F = 0.2 * Ftarget.

For each of the HCRs, a number of different target fishing mortalities were explored (0.0, 0.05, 0.075, 0.1, 0.125, 0.15). No evaluation of different Btrigger values was carried out, so that all evaluations used MSY Btrigger as the trigger point. All HCRs where evaluated with three variants:

- Without any additional constraints
- With a minimum TAC of 50 kT
- With a maximum 20% inter-annual variation (IAV) in TAC, but only when the stock is above Btrigger)



8.2 HCR evaluation tools

The base assessments ("Operating model") of the evaluations were either the WGWIDE 2019 SS3 assessment (ICES, 2019d) or the exploratory SAM assessment that was carried out as part of the IBPWHM 2019 (ICES, 2019b).

As input to the SS3 simulations, 1000 iterations were generated from respective assessments. For SS3 this was done by generating 10000 iterations and then resampling 1000 of them so as to end up with the same starting conditions as in the stock assessment itself.

The 1000 SAM iterations were generated by using the SAM simulate function, based on the IBPWHM 2019 exploratory SAM assessment; these were then converted to FLSAM objects which were again converted to 1000 FLStock objects[2]

The SRR model was the constrained segmented regression (SegRegBlim), similar to the IBPWHM 2019, while leaving out the exceptionally strong 1982 year class.

Two simulation tools were used: the EqSim simulator and the SAM HCR forecast

The EqSim simulator is a further worked up version of the SimpSIM approach that was used for the blue whiting MSE in 2016 (ICES, 2016). The code was further developed by Andrew Campbell and Martin Pastoors to improve standardization, documentation and visualization of results. Some key improvements where:

- the development of standardized codes for Operating Models (OM) a Management Procedures (MP), including new types of HCR elements.
- the development of standardized codes for statistical outputs and visualization thereof.

The SAM HCR forecast is a simple stochastic forecast with HCR to evaluate management for fish stocks that need rebuilding in the short-term. This method enables the investigation of several management strategies without the need of intensive computer power, while still accounting for different sources of uncertainty. The stochastic forecasts start from what we believe is the current level of the stock, i.e. the assessment estimates currently used for tactical management advice, with consideration of the uncertainty in these estimates. Rebuilding is evaluated forward for a specified number of years (here: 23 years) and for different target fishing mortality values (Ftarget)

The method was developed as an extension to the stockassessment R package for the SAM model (Nielsen and Berg, 2014; Berg and Nielsen, 2016) and applied to western horse mackerel[3].

We applied two different assessments to two different evaluation tools as follows:

	WGWIDE19 SS3	WGWIDE19 SAM
EqSim simulator	Yes	Yes
SAM HCR forecast	No	Yes

For each evaluation, we scanned over different F target values: 0, 0.05, 0.075, 0.10, 0.125, 0.15.

Each simulation was run over 23 year, split into the following periods:

[2] https://github.com/ices-eg/wk_WKREBUILD/blob/master/EqSimWHM/Scripts/HOM%20SAM%20simulator.r

Note: running the code required running it in batches of around 200 iterations due to unexplained errors arising when running for larger batches. This issue has not been solved, except by running it in multiple batches.

[3] https://github.com/vtrijoulet/SAM/tree/master2

- Current period (CU): 2018-2020
- Short term (ST): 2021-2025
- Medium term (MT): 2026-2030
- Long term (LT): 2031-2040

8.3 EqSim simulator tool

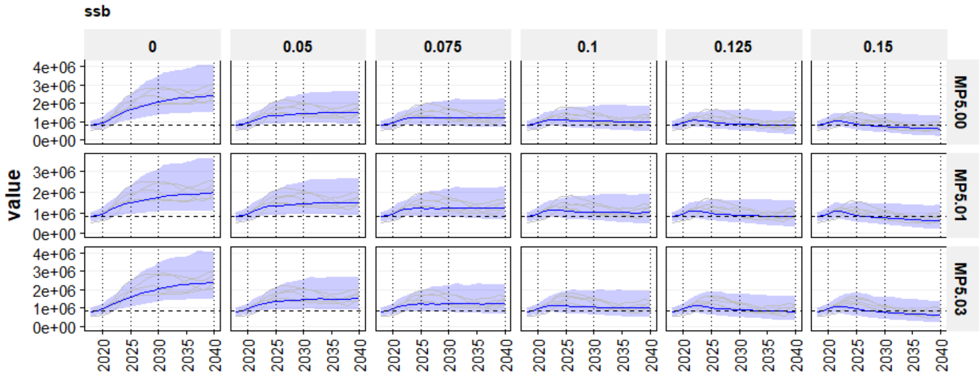
8.3.1 Eqsim applied to SS3 assessment

The SS3 assessment was run with OM2.2:

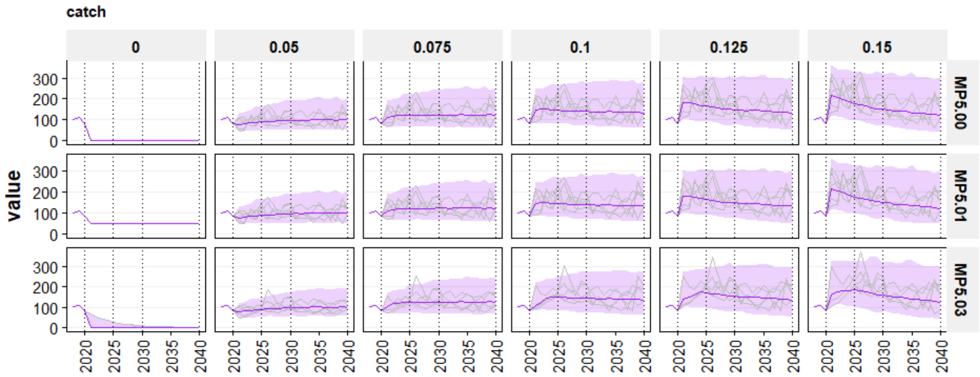
```
#WGWIDE2019 Update assessment, IBPWHM reference points, stochastic bio and selection
OM2.2 <- list("code" = "OM2.2",
            "desc" = "WGWIDE19",
            "IM" = NA,
            "SRR" = "SRR.WG19.SegReg_Blim.exterm", "RecAR" = TRUE, maxRecRes = c(3,-3),
            "BioYrs" = c(2008,2017), "BioConst" = FALSE,
            "SelYrs" = c(2008,2017), "SelConst" = FALSE,
            "Obs" = NA,
            refPts = list("Fpa" = 0.074, "Flim" = 0.103, "Fmsy" = 0.074, "Bpa" = 1168272,
                          "Blim" = 834480, "MSYBtrigger" = 1168272, "Bloss" = 761613),
            "pBlim" = 0.05)
```

8.3.1.1 Constant F strategy

- MP5.00 constant F;
- MP5.01 constant F with minimum TAC of 50kT;
- MP5.03 constant F with 20% IAV on TAC constraint above Btrigger.



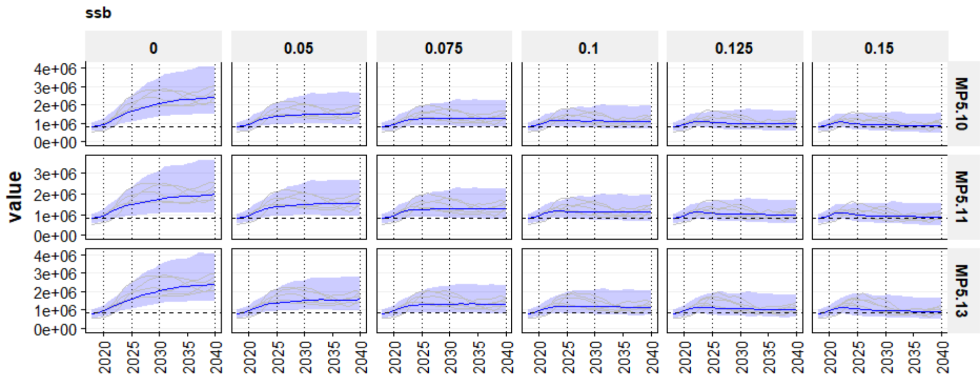
•

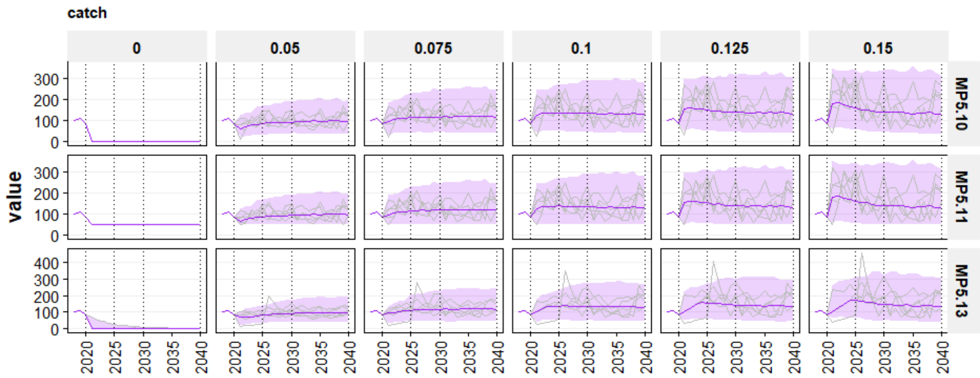


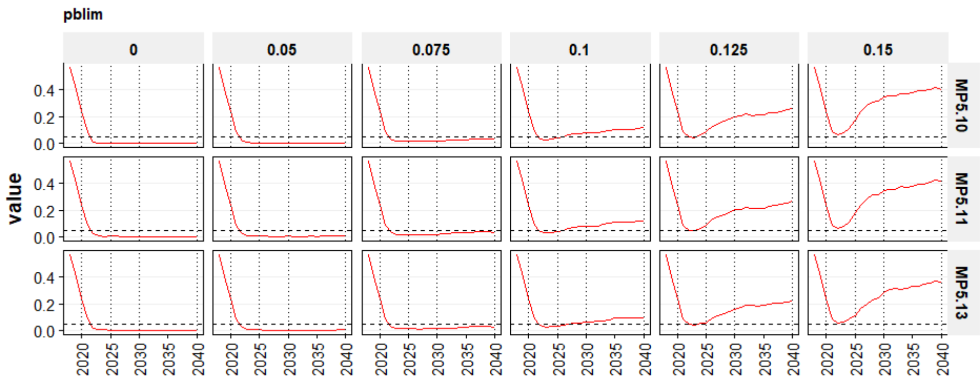
8.3.1.2 ICES Advice Rule

Scenarios 5.1, 5.11 and 5.13 (ICES advice rule variants)

- MP5.10 ICES AR
- MP5.11 ICES AR, min TAC = 50kt
- MP5.13 ICES AR, 20% IAV, only above Btrigger





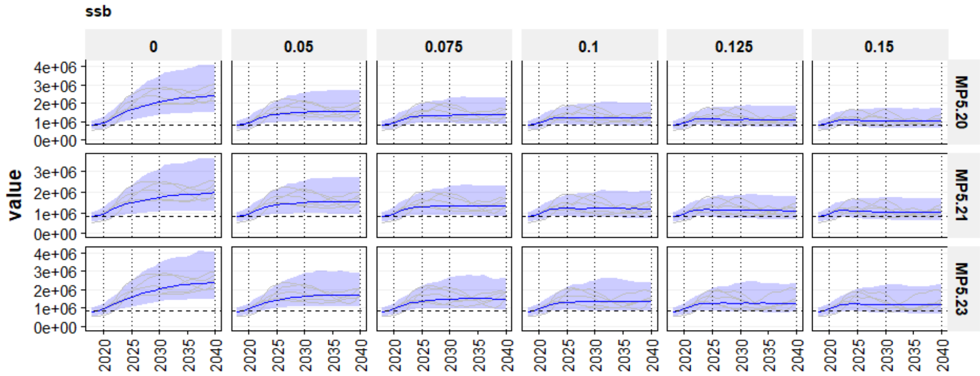


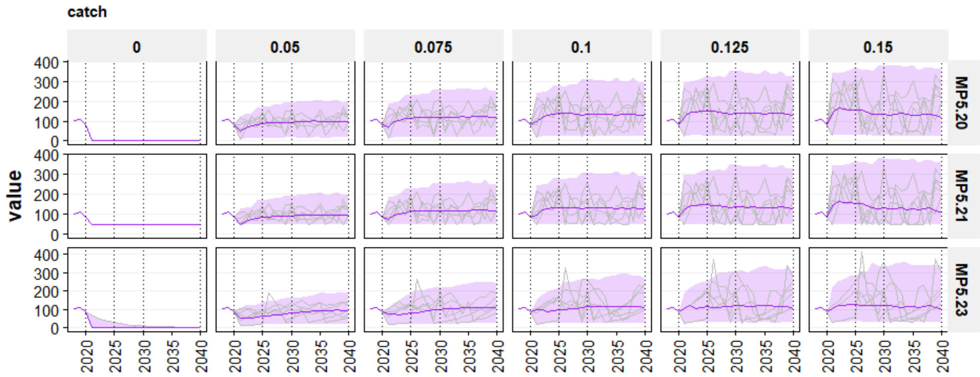
8.3.1.3 Double Breakpoint Rule

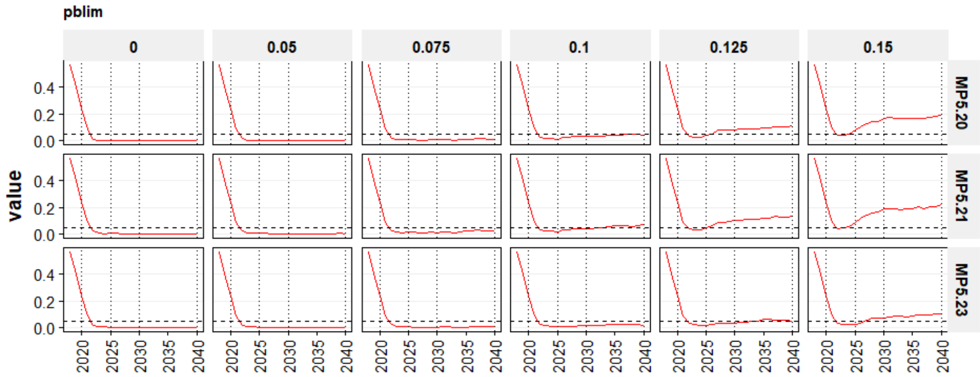
This HCR is similar to the blue whiting HCR that was evaluated in 2016 (ICES, 2016).

- MP5.20 Double BP
- MP5.11 Double BP with minimum TAC of 50kT
- MP5.13 Double BP with 20% IAV constraint above Btrigger.

Minimum F in the Double breakpoint rule is 20% of Ftarget.





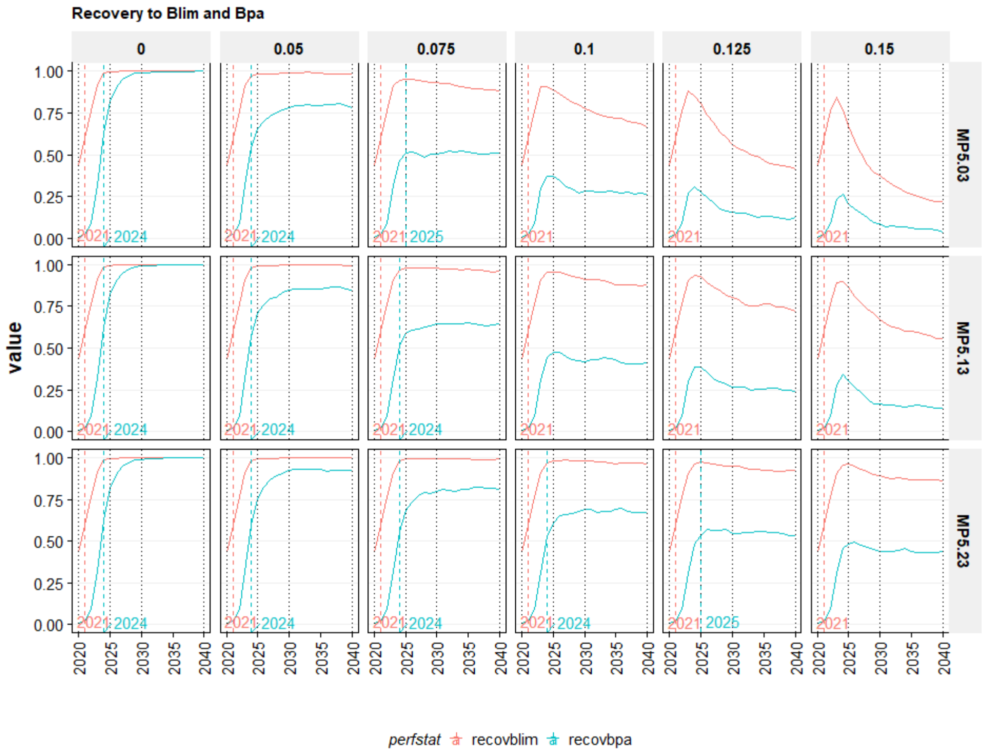


8.3.1.4 First year of achieving rebuilding with 20% IAV constraint scenarios

The first year of achieving rebuilding to Blim and Bpa was calculated as the first year where the probability of being above Blim or Bpa was larger than 50%. The analysis was carried out for the following scenarios:

- MP5.03 constant F with 20% IAV on TAC constraint above Btrigger.
- MP5.13 ICES AR, 20% IAV, only above Btrigger
- MP5.13 Double BP with 20% IAV constraint above Btrigger.

Results indicate that the constant F strategy is the least cautious rule and the double breakpoint rule is the most cautious rule. Under the F strategy rule with a Ftarget of 0.075, rebuilding to Bpa is expected to be achieved is only just being achieved (probability just above 50%) by 2025, while in the double breakpoint rule this is expected to be achieved in 2024 with substantially higher probabilities of remaining above Bpa. The first year of rebuilding to Bpa in the double breakpoint rule with target fishing mortalities up to 0.1 is the same as the first year of rebuilding under the zero fishing scenarios.



8.3.2 Eqsim applied to SAM assessment

The SS3 assessment was run with OM2.2:

```
#WGWIDE2019 SAM assessment, IBPWHM method for reference points, stochastic bio and selection
OM2.3 <- list("code" = "OM2.3",
              "desc" = "WGWIDE19_sam",
              "IM" = NA,
              "SRR" = "SRR.WG19.SegReg_Blim.exterm", "RecAR" = TRUE, maxRecRes = c(3,-3),
              "BioYrs" = c(2008,2017), "BioConst" = FALSE,
              "SelYrs" = c(2008,2017), "SelConst" = FALSE,
              "Obs" = NA,
              refPts = list("Fpa" = 0.115, "Flim" = 0.161, "Fmsy" = 0.115, "Bpa" = 856540,
                            "Blim" = 611814, "MSYBtrigger" = 856540, "Bloss" = 604476),
              "pBlim" = 0.05)
```

Note that the biomass reference points have been estimated separately for the SAM assessment, and are a bit lower than for the SS assessment (see section 7).

8.3.2.1 Constant F rule with SAM assessment

Results for the constant F rule are not presented because it was clear that this option would not be selected by the PELAC for the potential rebuilding plan.

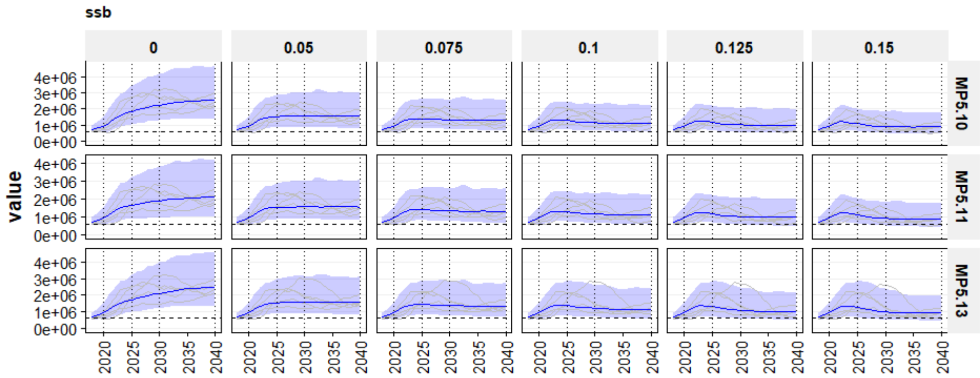
8.3.2.2 ICES Advice Rule with SAM assessment

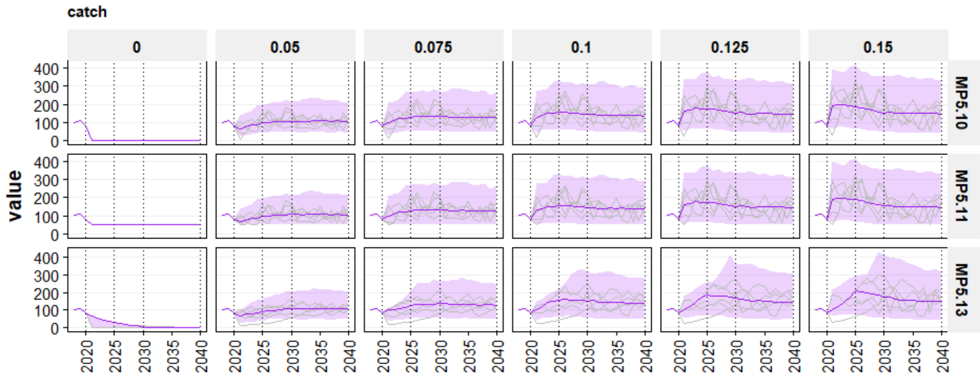
Scenarios 5.10, 5.11 and 5.13 (ICES advice rule variants)

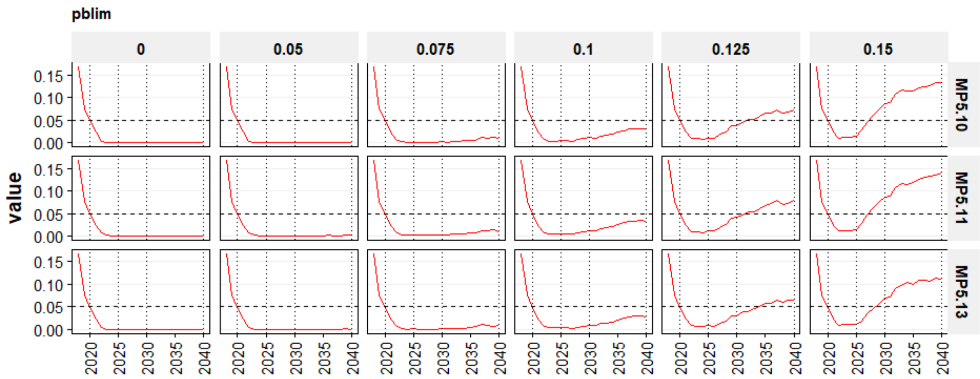
- MP5.10 ICES AR;

- MP5.11 ICES AR with minimum TAC of 50kT;
- MP5.13 ICES AR with 20% IAV constraint above Btrigger.

While the probability of being below Blim decreases in the beginning of the simulation period, for all F targets, the probability of being below Blim start to increase again after 2025 when target fishing mortalities are too high (e.g. > 0.075).







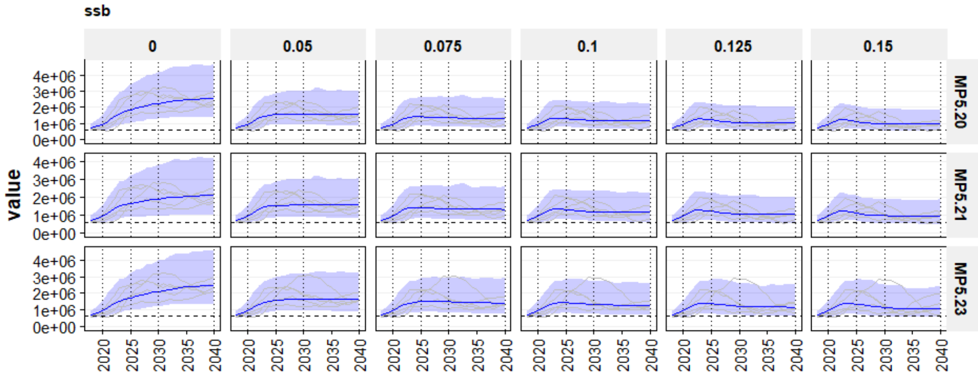
8.3.2.3 Double Breakpoint Rule with SAM assessment

This HCR is similar to the blue whiting HCR that was evaluated in 2016 (ICES, 2016).

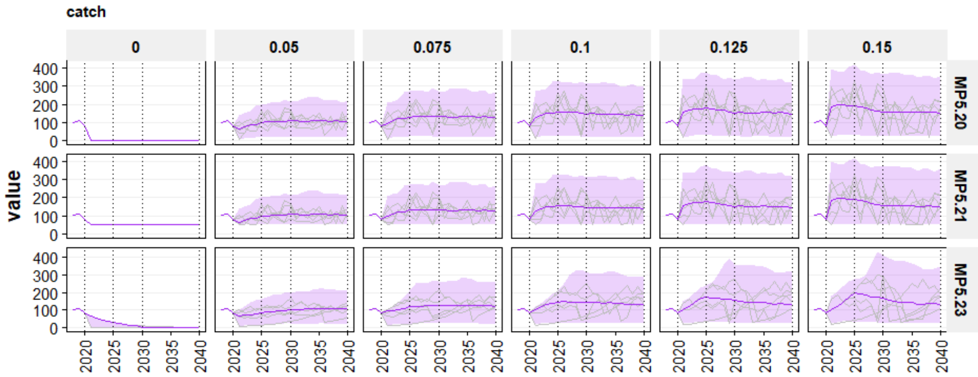
- MP5.20 Double BP
- MP5.11 Double BP with minimum TAC of 50kT;

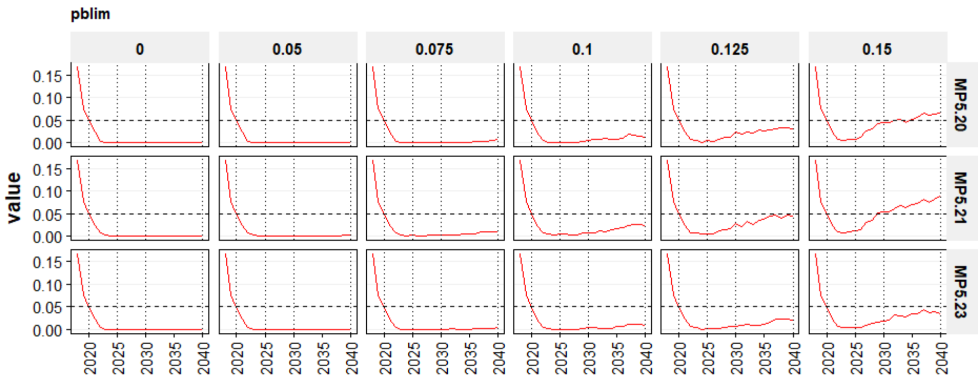
- MP5.13 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

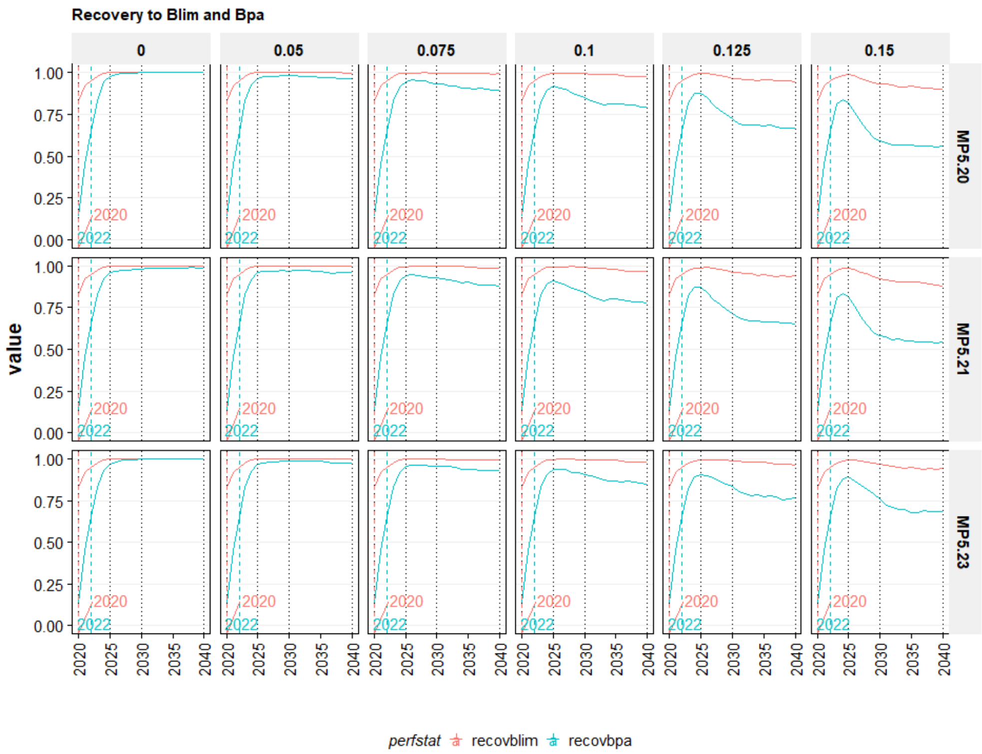
Generally, what we find is that the SAM assessment has a somewhat more optimistic view of the stock size in relation to the reference points. This means that the stock is estimated to be above Blim with a high probability in most of the scenarios. It also means that expected recovery to Bpa is in 2022 in all scenarios.



-







8.4 SAM HCR forecast tool

8.4.1 Description of the method

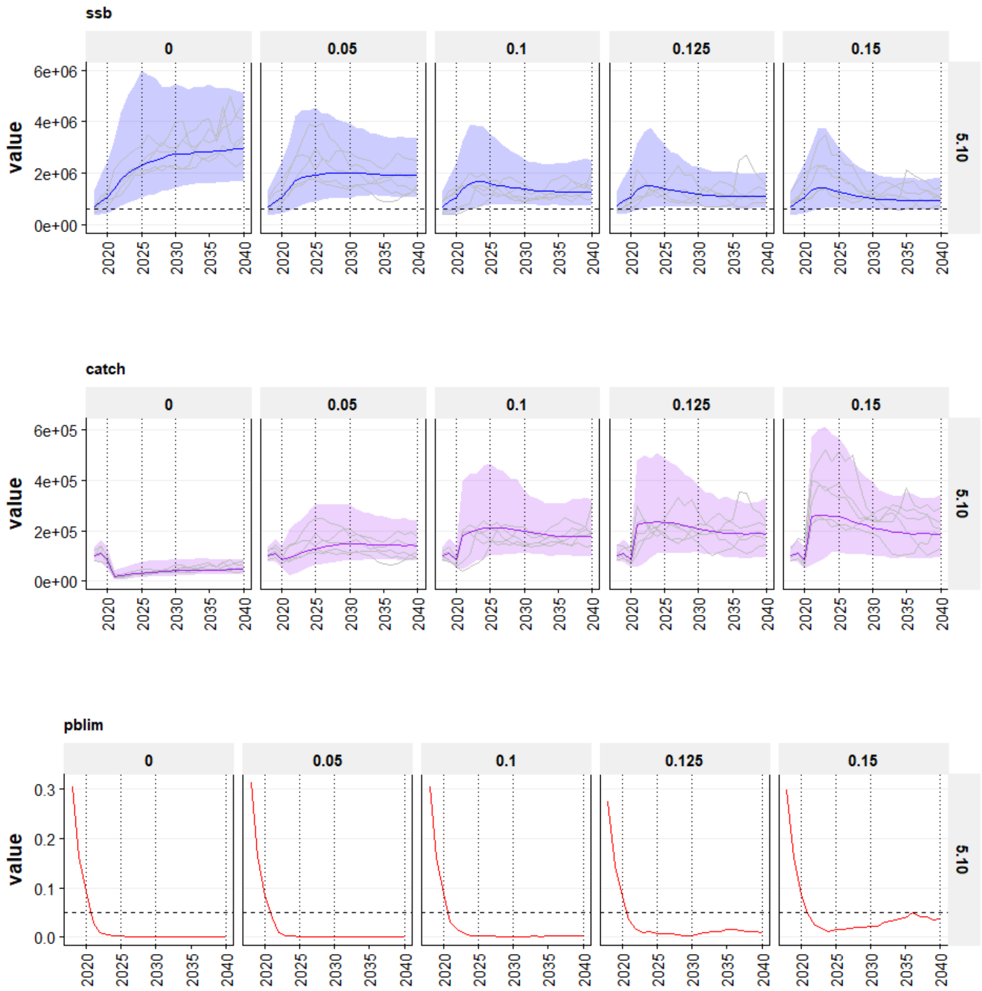
The SAM HCR was applied to the exploratory SAM assessment (IBPWHM 2019) that was also used for the EqSim with SAM analysis. The SAM HCR forecast can only be run on a SAM assessment[4].

8.4.2 SAM HCR with ICES Advice Rule

Here we only present the simple ICES AR scenario without any additional constraints as the main purpose is only to show the feasibility of using this simple method while generating similar results from more complicated methods.

- MP5.10 ICES AR.

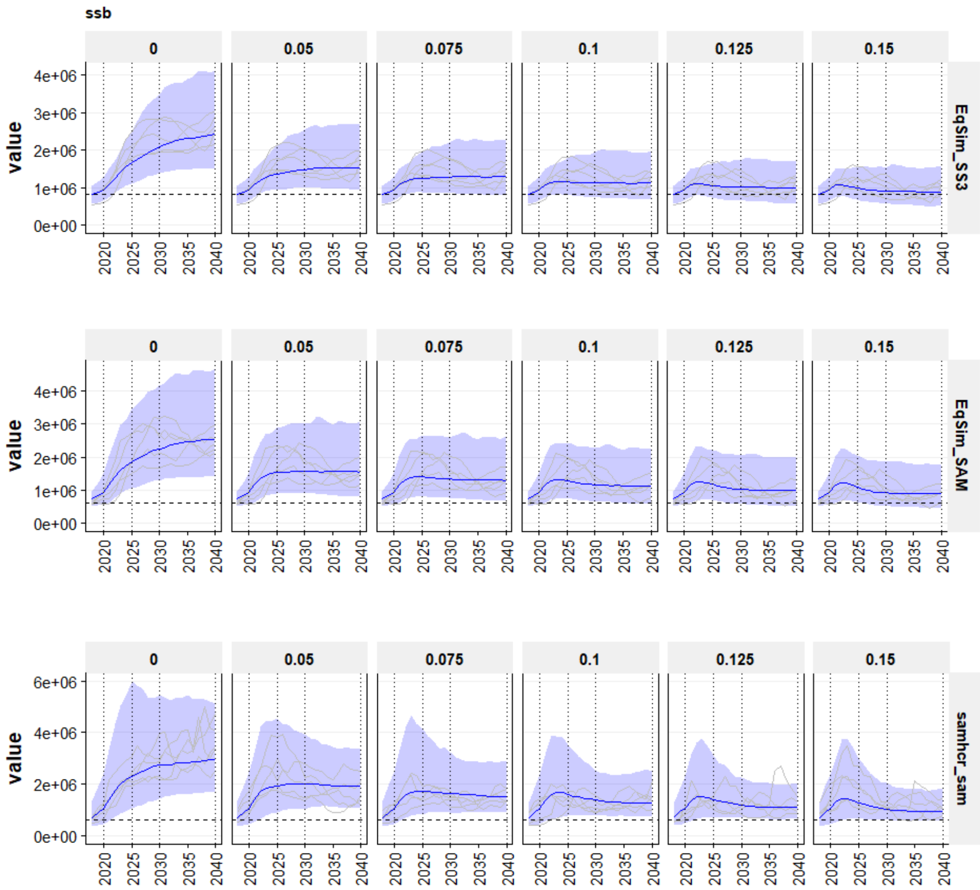
[4] Note that with the SAM HCR it was not possible to run the forecast with F = 0; therefore F = 0.01 has been run for the results denoted below with F = 0 .

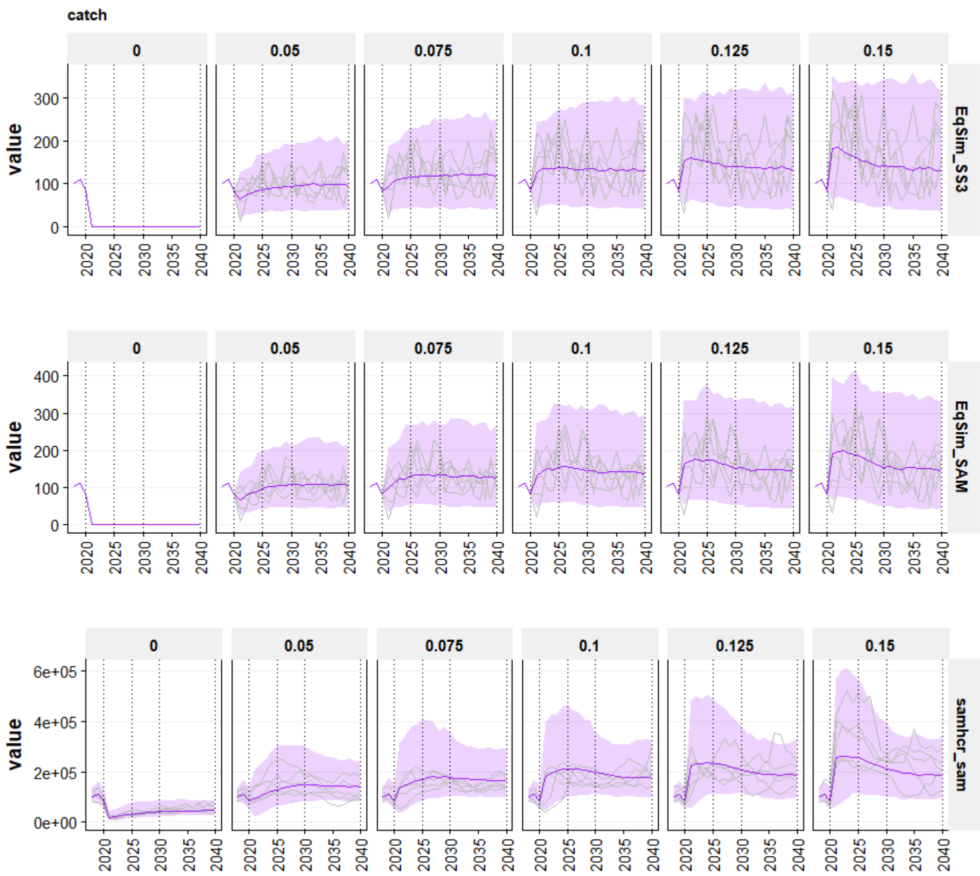


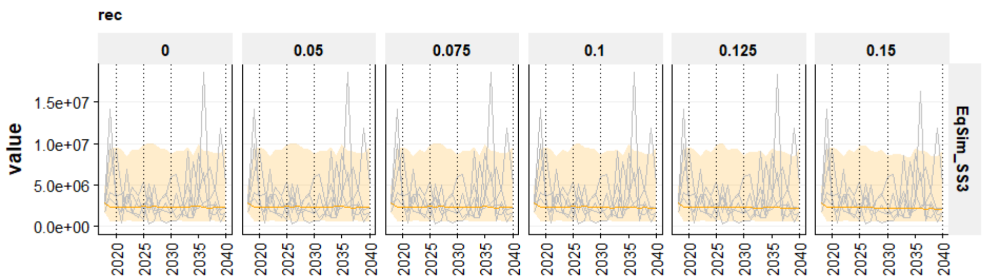
8.5 Comparison of results for different simulation tools and assessments

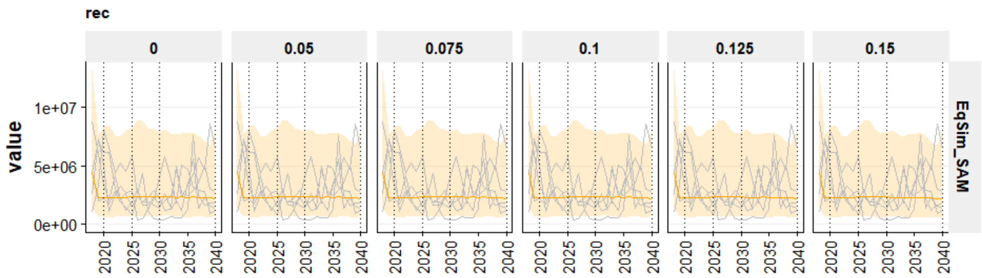
To compare the behaviour of evaluation tools (EqSim or SAM HCR) and assessment method (SAM or SS3), we compared the simple ICES AR scenarios for the three possible combinations:

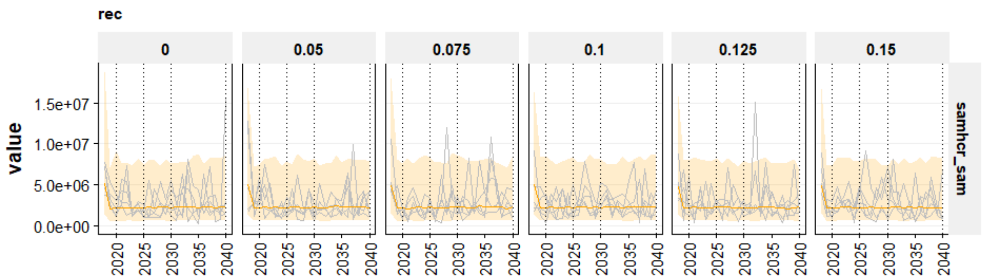
- EqSim – SAM – MP5.1 (ICES AR)
- EqSim – SS3 – MP5.1 (ICES AR)
- SAM HCR – SAM – MP5.1 (ICES AR)

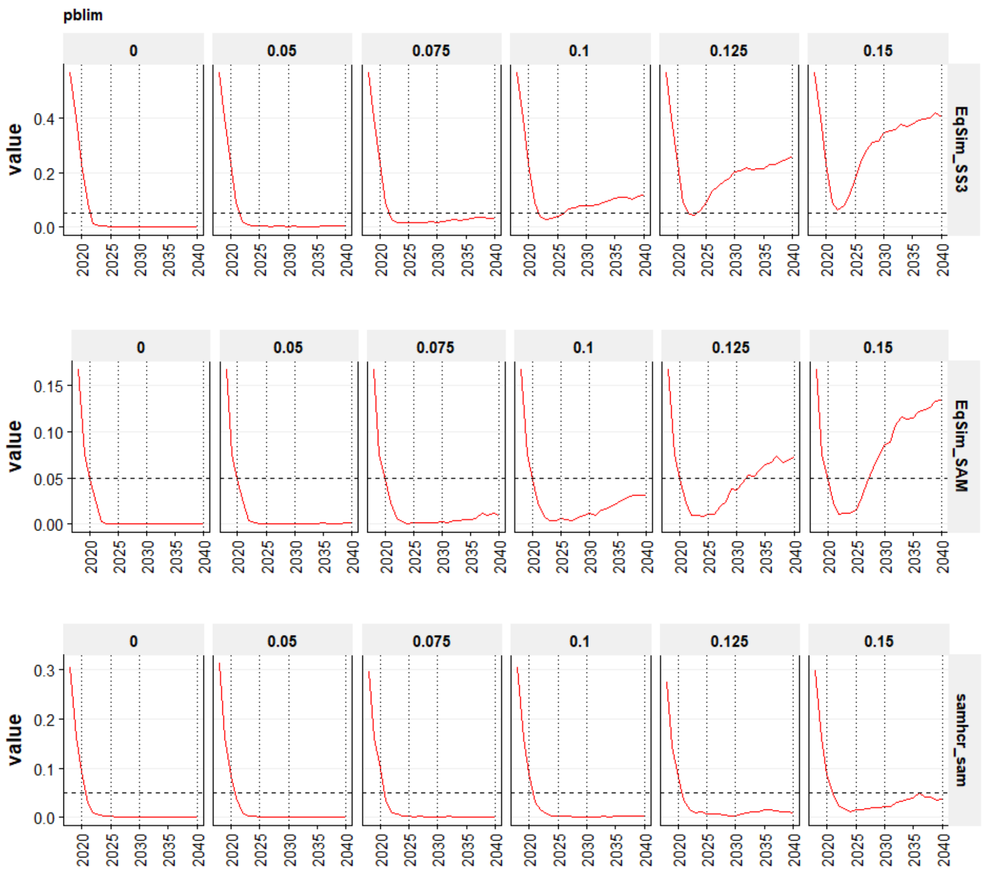

ssb
0
0.05
0.075
0.1
0.125
0.15
EqSim_SS3
EqSim_SAM
samhcr_sam
value
0e+00
1e+06
2e+06
3e+06
4e+06
2e+06
4e+06
6e+06
2020
2025
2030
2035
2040


catch
0
0.05
0.075
0.1
0.125
0.15
EqSim_SS3
EqSim_SAM
samhcr_sam
value
2020
2025
2030
2035
2040
0
100
200
300
400
0e+00
2e+05
4e+05
6e+05


rec
0
0.05
0.075
0.1
0.125
0.15
EqSim_SS3
value
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2020
2025
2030
2035
2040


rec
0
0.05
0.075
0.1
0.125
0.15
EqSim_SAM
value
1e+07
5e+06
0e+00
2020
2025
2030
2035
2040


rec
0
0.05
0.075
0.1
0.125
0.15
samhcr_sam
value
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2020
2025
2030
2035
2040



The probability of being below Blim broadly follows the same pattern across the three different evaluation method although the levels do differ between the evaluations. Because the SAM assessment estimates the most recent SSBs higher than year where Bloss was calculated, the probability of currently being below Blim is smaller. The patterns observed for the EqSim_SS and EqSim_SAM runs are qualitatively similar albeit at different levels. The SAMHCR_SAM run exhibits a slightly different pattern because the forecasted SSB is expected to remain above Blim with a high probability in all F scenarios. This may be due to the fact that the SAMHCR is operating as a forecast only and therefore lacks the feature that the management perception of the stock differs from the real stock, so that the implemented HCR in the simulation does not suffer from the mismatch between perception and reality.

9 Selection of preferred HCRs for Western Horse mackerel

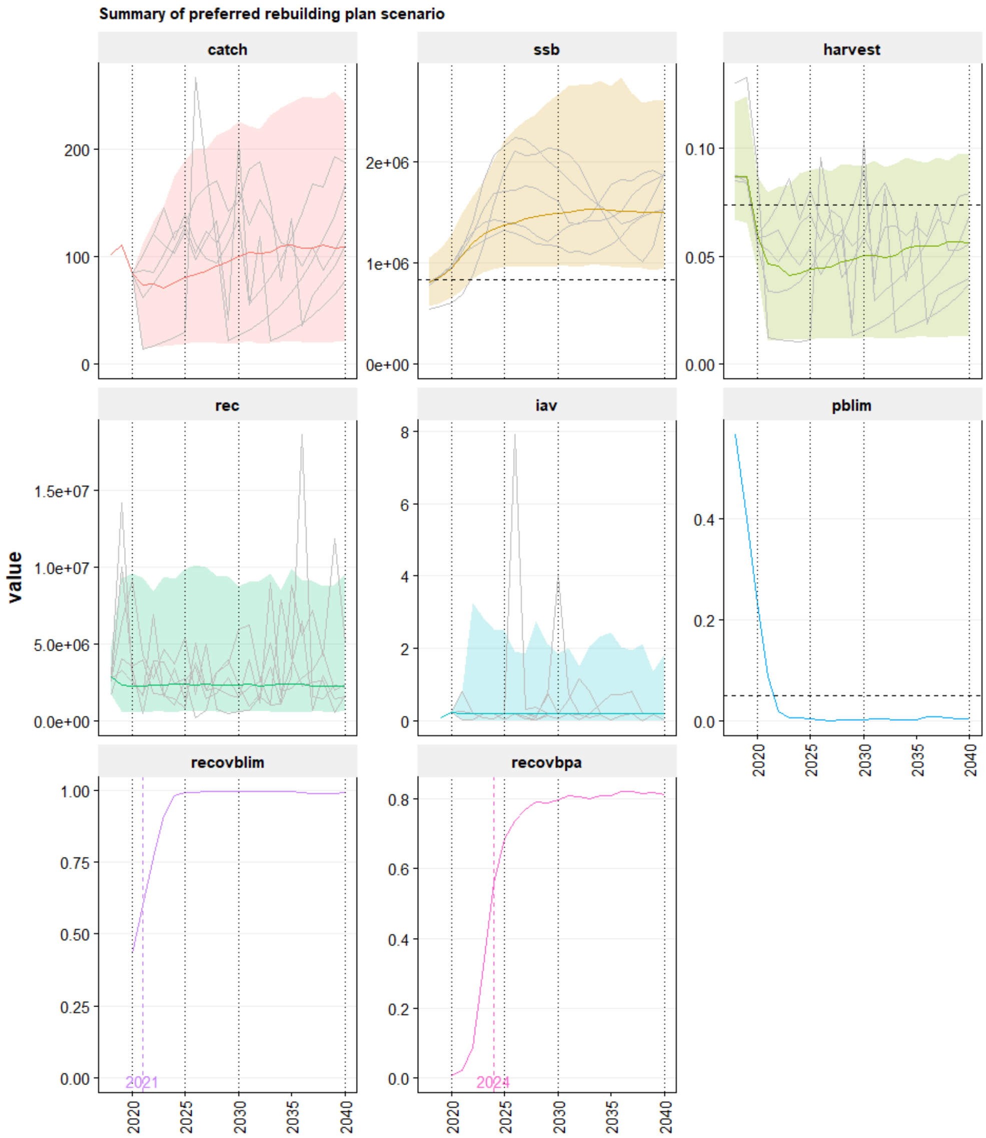
The PELAC selected the following preferred option for the Western horse mackerel rebuilding plan:

- Evaluation method: EqSim
- Assessment: Stock Synthesis (WGWIDE 2019), because this is the basis for the assessment and advice.
- Target fishing mortality at Fmsy = 0.074 (approximated by 0.075 in the simulations)
- Blim at ICES Blim (834 480 t)
- Btrigger at ICES MSY Btrigger (1 168 272 t)

- Double breakpoint rule with 20% constraint on IAV above Btrigger
- Minimum F when stock is below Blim at 20% of Fmsy = 0.015

The selected rebuilding plan has a 50% probability of rebuilding to Blim by 2021 (similar to zero catch option) and a 50% probability of rebuilding to Bpa/MSY Btrigger by 2024 (similar to the zero-catch option). Furthermore, the probability of being below Blim remains well below 5% for the duration of the simulation.

In this scenario, the average catch in the years 2021-2025 is expected to be lower than recent catches. However, after rebuilding, catches should be able to be maintained around 100 000 tonnes.



Summary of results of the preferred rebuilding plan

```
statistic  yearrange   period   median      range
---------- ----------- -------- ----------- ---------------------

catch      2018-2020   CU       102         84 - 110             * in kT
catch      2021-2025   ST       75          17 - 167
```

catch	2026-2030	MT	92	20 - 210
catch	2031-2040	LT	107	21 - 242
ssb	2018-2020	CU	872,454	608,164 - 1,210,564
ssb	2021-2025	ST	1,249,710	832,465 - 1,902,950
ssb	2026-2030	MT	1,451,882	966,840 - 2,506,102
ssb	2031-2040	LT	1,514,418	958,213 - 2,740,040
harvest	2018-2020	CU	0.080	0.048 - 0.118
harvest	2021-2025	ST	0.044	0.011 - 0.085
harvest	2026-2030	MT	0.047	0.012 - 0.092
harvest	2031-2040	LT	0.054	0.012 - 0.095
rec	2018-2020	CU	2,599,180	696,645 - 7,944,499
rec	2021-2025	ST	2,363,631	606,888 - 9,317,602
rec	2026-2030	MT	2,361,298	599,077 - 9,438,791
rec	2031-2040	LT	2,321,690	612,371 - 9,088,107
iav	2018-2020	CU	0.162	0.086 - 0.239
iav	2021-2025	ST	0.200	0.021 - 2.576
iav	2026-2030	MT	0.200	0.018 - 2.083
iav	2031-2040	LT	0.200	0.017 - 2.032
pblim	2018-2020	CU	0.401	0.243 - 0.560
pblim	2021-2025	ST	0.006	0.005 - 0.082
pblim	2026-2030	MT	0.002	0.001 - 0.003
pblim	2031-2040	LT	0.004	0.002 - 0.009

Table of settings used in the evaluation

class	desc	value
OM	code	OM2.2
OM	desc	WGWIDE19
OM	IM	
OM	SRR	SRR.WG19.SegReg_Blim.exterm
OM	RecAR	TRUE
OM	maxRecRes1	3
OM	maxRecRes2	-3
OM	BioYrs1	2008
OM	BioYrs2	2017
OM	BioConst	FALSE
OM	SelYrs1	2008
OM	SelYrs2	2017
OM	SelConst	FALSE
OM	Obs	
OM	refPts.Fpa	0.074
OM	refPts.Flim	0.103
OM	refPts.Fmsy	0.074
OM	refPts.Bpa	1168272
OM	refPts.Blim	834480
OM	refPts.MSYBtrigger	1168272
OM	refPts.Bloss	761613
OM	pBlim	0.05
MP	code	MP5.23
MP	desc	Double BP HCR
MP	xlab	Double BP IAVBtrig
MP	HCRName	DoubleBP
MP	F_target1	0
MP	F_target2	0.025
MP	F_target3	0.05
MP	F_target4	0.075
MP	F_target5	0.1
MP	F_target6	0.125
MP	F_target7	0.15
MP	B_trigger	MSYBtrigger
MP	minTAC	

```
MP      maxTAC
MP      TAC_IAV1            0.2
MP      TAC_IAV2            0.2
MP      Obs.cvF             0.22
MP      Obs.phiF            0.03
MP      Obs.cvSSB           0.36
MP      Obs.phiSSB          0.51

OTHER   niters              1000
OTHER   nyr                 23
OTHER   CU                  2018-2020
OTHER   ST                  2021-2025
OTHER   MT                  2026-2030
OTHER   LT                  2031-2040
OTHER   flstock             WGWIDE19.RData
OTHER   flstock_sim         MSE_WGWIDE19_FLStocks_1k15PG.RData
```

10 Summary and conclusions

This report has brought together many different topics that are related to the western horse mackerel stock in an attempt to develop a potential rebuilding plan for the stock. Even though western horse mackerel was not classified by ICES as in need of rebuilding in their latest advice (ICES, 2019a), the general perception within the fishing industries has been that the stock has been in a poor state but that there have been some positive signals in recent recruitment. Using the new recruitments to improve the stock status requires a careful management approach. The PELAC has been a proponent of developing management plans for all stocks in their remit. In this case, the PELAC has termed the approach a rebuilding plan because of the current stock status of the stock.

Substantial progress has been made over the past few years on horse mackerel stock ID (Farrell et al., 2020). The full genome sequencing of horse mackerel from samples taken all the way from the Skagerrak to the Mediterranean and North Africa, has yielded a suitable panel of SNP markers that can be used to differentiate between the different horse mackerel stocks. The strongest differentiation between populations was between the northern and southern populations, with the boundary being in the middle of Portugal. The North Sea population is clearly distinct from the Western population and it should be possible to tell the difference from mixed samples with a high probability (>93%). This would also allow screening of catches in 7d and 7e on the contribution of western and North Sea populations. The separation between the northern and southern populations could mean that the current division between western and southern horse mackerel is not adequate, at the northern part of 9a is currently included in the southern population. A similar split in the middle of Portugal has also been observed for boarfish (Farrell et al., 2016) and could indicate a biogeographical feature.

Length compositions of the catches are an important element of the assessment approach for western horse mackerel, because Stock Synthesis uses length composition in combination with age-length key to estimate the age compositions within the model. Part of a rebuilding plan for western horse mackerel could be to evaluate differences in length compositions in the catches in certain areas and to take specific measures to protect incoming recruitment. Therefore, we planned to carry out an analysis of length compositions by area and season. However, we found that such data is not currently available for all years. Length data for western horse mackerel is not included in the ICES InterCatch database. Instead, length data has been processed on a year by year basis in non-standardized Excel spreadsheets. A time series of length compositions by area and season can therefore only be derived by manually working through the spreadsheets and extracting the required information. This was not feasible as part of the project to develop and evaluate a rebuilding plan for western horse mackerel. We recommend to WGWIDE that the full time series of catch at length by country is recreated from the Excel spreadsheets and converted in a standardized database format to allow for future interrogations of the data and an underpinning of the input data to the stock assessment.

In order to understand how a stock would respond to recovery measures, it is useful to consider the age composition in the spawning stock which illustrates how recruitment in the previous years contributed to the present spawning stock. To this end, an SSB per recruit analysis has been carried out. As one should expect for a relatively long-lived species with low mortality, the spawning stock is currently rather old. At F =0.075, the mean age is about 9 years, 80% is older than 5 years and 20% older than 12 years. So, an improved recruitment will take some time to materialize as increased SSB. The results also indicate that with a low F, the plus group still does matter.

The current stock assessment method for western horse mackerel is Stock Synthesis 3, as agreed in the WKWIDE benchmark of 2017 (ICES, 2017b). Reference point were also set at WKWIDE 2017 but have subsequently been updated in the IBPWHM 2019 (ICES, 2019b). In addition, an exploratory SAM assessment has been carried out as part of IBPWHM 2019. This was done in order to get a second view on stock trends but also to be able to run the SAM HCR forecast as part of the development of a potential rebuilding plan. The exploratory SAM assessment (https://www.stockassessment.org/setStock.php?stock=WHOM2018) was initiated with the same input data as was used for the Stock Synthesis assessment of WGWIDE 2018 (ICES, 2018) with the exception of the length frequency data, which was not used. The PELACUS survey data was therefore only used as an index of biomass within SAM. The process of fine-tuning the assessment lead to the binding of the observation variances for certain variables and to the application of a fixed selectivity pattern (correlation coefficient ϱ=1 in the F random process (https://github.com/martinpastoors/wgwide/blob/master/R/HOM%20optimization_SAM.R). A comparison of Fbar and SSB between the SS3 assessments of WG2018 and 2019 with the SAM assessment (WG18SAM, WG19SAM), shows that the general trends are the same but that there are some deviations in certain periods (e.g. the SSB in the late 1980s is estimated substantially higher in SAM compared to SS3). The Stock Synthesis results are in general a bit smoother compared to SAM.

In order to be able to use the SAM assessment as an alternative assessment in the rebuilding plan evaluation, we needed to estimate reference point for this assessment. In doing so, we aimed to follow the same procedure as during IBPWHM 2019 (ICES, 2019b). However, one of the elements of the reference point estimation, triggered a more in-depth study: the role of assessment uncertainty parameter Fcv and Fphi. There has been little standardization in how Fcv and Fphi have been calculated in different benchmarks where reference points were estimated. Fcv is expected to capture the assessment error in the advisory year and Fphi is the autocorrelation in assessment error in the advisory year (ICES, 2014a). We documented the method for generating the input data for the calculations and explored the sensitivity of Fcv and Fphi to the assessment that was used (both for western horse mackerel and for Atlantic mackerel). We found that there can be a high dependence of Fphi on the assessment that is used to compare against the Fset. When the assessment that is used has values that are all higher or lower than the Fset values, then Fphi will be close to zero. To our knowledge, this behaviour of Fphi was unknown so far. We also found that the number of years that is used for calculating Fcv and Fphi may have an impact on the values. In the recommendations from WKMSYREF3 it is stated that 10 years (or more) should be taken. A further study should be undertaken to assessment the impacts of using different time periods for estimating Fcv and Fphi.

During the IBPWHM 2019, reference points were estimated for western horse mackerel based on the 2018 WGWIDE assessment and using default values for Fcv and Fphi (0.212 and 0.423) and using a segmented regression through Blim (segregBlim). In order to calculate reference points for the exploratory SAM assessment and to explore the sensitivity to the assessment year, reference points were calculated on the basis of the 2018 or 2019 assessments for SS and SAM. The reference points for the SAM assessment are based on the 2018 assessment. Bpa and Blim are lower than the values for the SS assessment, while the Fmsy is higher. The changes due the assessment year were minor for both the SS and SAM assessments.

RP	WG18	WG18SAM	WG19	WG19SAM
Blim	834480	611814	885341	612635
Flim	0.1107	0.1612	0.1049	0.1756
Fpa	0.07909	0.1152	0.07493	0.1254
MSYBtrigger	1168272	856540	1239478	857689
FMSY	0.09102	0.1262	0.08665	0.1353
FP05	0.08398	0.1255	0.07826	0.1402
FMSY_final	0.07909	0.1152	0.07493	0.1254

HCR evaluations

The HCR analyses represent two different assessment methods (SS3 and SAM) and two different HCR evaluation tools (EqSim and SAM HCR). Both HCR evaluation tools are of the type 'short-cut' with appropriate conditioning of the uncertainties in the assessment based on historical CV and autocorrelation in line with the recommendations from WKMSYREF3 and WKMSYREF4 . The evaluations followed the guidelines from WKGMSE2 (ICES, 2019c) and WKREBUILD (ICES, 2020).

Three different types of harvest control rules were evaluated:

- Constant F strategy: fixed Ftarget independent of biomass level
- ICES Advice Rule: breakpoint at Btrigger and straight decline in F to zero below Btrigger.
- Double Breakpoint rule: breakpoint at Btrigger and straight decline in F to 20% of Ftarget at Blim. Below Blim continued fishing at F = 0.2 * Ftarget.

For each of the HCRs, a number of different target fishing mortalities were explored (0.0, 0.05, 0.075, 0.1, 0.125, 0.15). No evaluation of different Btrigger values was carried out, so that all evaluations used MSY Btrigger as the trigger point. All HCRs where evaluated with three variants:

- Without any additional constraints

- With a minimum TAC of 50 kT
- With a maximum 20% inter-annual variation (IAV) in TAC, but only when the stock is above Btrigger)

Two simulation tools were used: the EqSim simulator and the SAM HCR forecast. The EqSim simulator is a further worked up version of the SimpSIM approach that was used for the blue whiting MSE in 2016 (ICES, 2016). The code was further developed by Andrew Campbell and Martin Pastoors to improve standardization, documentation and visualization of results. EqSim makes use of an Operating Model (OM) and a Management Procedure (MP). The SAM HCR forecast is a simple stochastic forecast with HCR to evaluate management for fish stocks that need rebuilding in the short-term. The stochastic forecasts start from what we believe is the current level of the stock with appropriate uncertainty, i.e. the assessment estimates currently used for tactical management advice, with consideration of the uncertainty in these estimates. Rebuilding is evaluated forward for a specified number of years and for different target fishing mortality values.

The EqSim with SS3 results indicate that the constant F strategy is the least cautious rule and the double breakpoint rule is the most cautious rule. Under the F strategy rule with a Ftarget of 0.075, rebuilding to Bpa is expected to be achieved is only just being achieved (probability just above 50%) by 2025, while in the double breakpoint rule this is expected to be achieved in 2024 with substantially higher probabilities of remaining above Bpa. The first year of rebuilding to Bpa in the double breakpoint rule with target fishing mortalities up to 0.1 is the same as the first year of rebuilding under the zero fishing scenarios.

Similar results have been obtained with the EqSim with SAM evaluations although the levels of SSB are slightly higher and risk to Blim is slightly lower. According to these evaluations, rebuilding to Bpa could be obtained by 2022 in all scenarios.

The SAM HCR with SAM evaluations have only been carried out for the ICES Advice Rule scenario, as this was intended more as a contrasting study rather than a full analysis of HCR evaluation. Again, we find similar patterns in simulated stock trends, but SSB is estimated higher than in the EqSim with SAM evaluations and risk to Blim stays below Blim for all target fishing mortalities that have been explored.

Given that the EqSim with SS3 evaluation is closest to the ICES advisory practice, this was used as the basis for the preferred rebuilding plan by the PELAC. The PELAC preferred options are:

- Target fishing mortality at Fmsy = 0.074 (approximated by 0.075 in the simulations)
- Blim at ICES Blim (834 480 t)
- Btrigger at ICES MSY Btrigger (1 168 272 t)
- Double breakpoint rule with 20% constraint on IAV above Btrigger
- Minimum F when stock is below Blim at 20% of Fmsy = 0.015

The selected rebuilding plan has a 50% probability of rebuilding to Blim by 2021 (similar to zero catch option) and a 50% probability of rebuilding to Bpa/MSY Btrigger by 2024 (similar to the zero-catch option). Furthermore, the probability of being below Blim remains well below 5% for the duration of the simulation.

In this scenario, the average catch in the years 2021-2025 is expected to be lower than recent catches. However, after rebuilding, catches should be able to be maintained around 100 000 tonnes.

11 References

Berg, C. W., and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES Journal of Marine Science: Journal du Conseil, 73: 1788-1797.

Campbell, A., De Oliveira, J. A. A., Kelly, C., and Roel, B. 2015. Report on Western Horse mackerel Management Strategy Evaluations (MSE). Report for Pelagic Fishing Industry. .

Cox, S. P., Benson, A. J., Doherty, B., and Johnson, S. 2018. Evaluation of potential rebuilding strategies for the Western Horse Mackerel (Trachurus trachurus) fishery.

Cox, S. P., and Kronlund, A. R. 2008. Practical stakeholder-driven harvest policies for groundfish fisheries in British Columbia, Canada. Fisheries Research, 94: 224-237.

Farrell, E. D., Carlsson, J. E. L., and Carlsson, J. 2016. Next Gen Pop Gen: implementing a high-throughput approach to population genetics in boarfish (Capros aper). Royal Society Open Science, 3.

Farrell, E. D., Fuentes-Pardo, A. P., Pettersson, M., Sprehn, C. G., and Andersson, L. 2020. Population structure of the Atlantic horse mackerel (Trachurus trachurus) revealed by whole-genome sequencing.

ICES. 2008. Western horse mackerel (Trachurus trachurus) (Divisions IIa, IVa, Vb, VIa, VIIa–c,e–k, and VIIIa–e), Section 9.4.3 in Advice 2008. December 2008.

ICES 2013a. EC request to ICES to evaluate possible modifications of the long-term management arrangement for the Western horse mackerel stock. Section 9.3.3.4 in Advice October 2013. *In* Report of the ICES Advisory Committee, 2013, p. 7 pp. ICES advice, 2013.

ICES. 2013b. Report of the Workshop to consider reference points for all stocks (WKMSYREF), 23 - 25 January 2013, Copenhagen, Denmark. ICES C.M. 2013 / ACOM:37. 17 pp. pp.

ICES. 2014a. Report of the Joint ICES-MYFISH Workshop to consider the basis for FMSY ranges for all stocks (WKMSYREF3), 17–21 November 2014, Charlottenlund, Denmark. ICES C.M. 2014 / ACOM:64.

ICES. 2014b. Report of the Workshop to consider reference points for all stocks (WKMSYREF2), 8-10 January 2014, Copenhagen, Denmark. . ICES C.M. 2014 / ACOM:47. 91 pp. pp.

ICES. 2015. Report of the Workshop to consider FMSY ranges for stocks in ICES categories 1 and 2 in Western Waters (WKMSYREF4), 13–16 October 2015, Brest, France. ICES C.M. 2015 / ACOM:58.

ICES. 2016. Report of the Workshop on Blue Whiting Long Term Management Strategy Evaluation (WKBWMS), 30 August 2016, Copenhagen. ICES C.M. 2016 / ACOM: 53.

ICES. 2017a. 12.4.3.1 ICES fisheries management reference points for category 1 and 2 stocks. ices.pub.3036.

ICES. 2017b. Report of the Benchmark Workshop on Widely Distributed Stocks (WKWIDE). ICES C.M. 2017 / ACOM:36.

ICES 2019a. Horse mackerel (Trachurus trachurus) in Subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a–c, and 7.e–k (the Northeast Atlantic). *In* ICES Advice 2019, p. 7 pp.

ICES. 2019b. Interbenchmark Protocol on reference points for western horse mackerel (Trachurus trachurus) in subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a-c,e-k (the Northeast Atlantic) (IBPWHM). ICES VOLUME X | ISSUE X (FORTHCOMING).

ICES. 2019c. Report of the Workshop on Guidelines for Management Strategy Evaluations (WKGMSE2). Volume 1, Issue 33.

ICES. 2019d. Working Group on Widely Distributed Stocks (WGWIDE 2019) Volume 1, Issue 36.

ICES. 2020. Workshop on guidelines and methods for the evaluation of rebuilding plans (WKREBUILD).2:55. 79 pp. . Volume 2, Issue 55. 79 pp pp.

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

Annex 3: Working Documents

Summary

This working document has been produced by the group of experts that developed the original evaluation of a rebuilding plan of Western Horse mackerel on behalf of the Pelagic Advisory Council (Pastoors et al., 2020). The evaluation was submitted to ICES for a scientific review by WKWHMRP in February 2021. The reviewers and workshop participants concluded that various elements of the paper and rebuilding plan could not be evaluated due to lack of detail. Therefore, it was not possible to determine from the report whether all criteria listed under the ToR of WKWHMRP were met.

The review requested that the report authors provide clarifications on some elements, and address several additional questions. These are addressed within this working document to WKWHMRP. Detailed results of the simulations are available on the WKHMRP sharepoint site.

The additional work consists of the following key elements:

1) **Better documentation on the simulation approaches, the parameterization of the models, the results achieved and the metrics being used (specifically the re-building metrics)**

The expert group involved in this work acknowledges that the documentation of the simulation work on horse mackerel could have been better and more complete in the first place. On reworking the material, we also carried out a number of checks on the simulation code and improved the standard documentation from the model. We have now included the standard WKGMSE table for MSE documentation in section **Error! Reference source not found.** of the Working Document.

We also streamlined the code on github[1], separating it into code for processing of input data, code for general settings and loading of relevant functions, code for simulations and code for presentation of the results.

No assessment or forecast is conducted during the simulation i.e. this is a 'short-cut' MSE approach. The parameterisation of uncertainty in observed SSB and F are incorporated into the simulations following the methodology developed for EqSim whereby SSB and FBar outputs from historic assessments and short term forecasts are compared to the most recent assessment and described via a pair of parameters (CV and autocorrelation) for each of SSB and FBar. These parameters are then used to generate an error and derived "observed" values of SSB and Fbar for the management procedure.

Three different types of harvest control rules were evaluated:

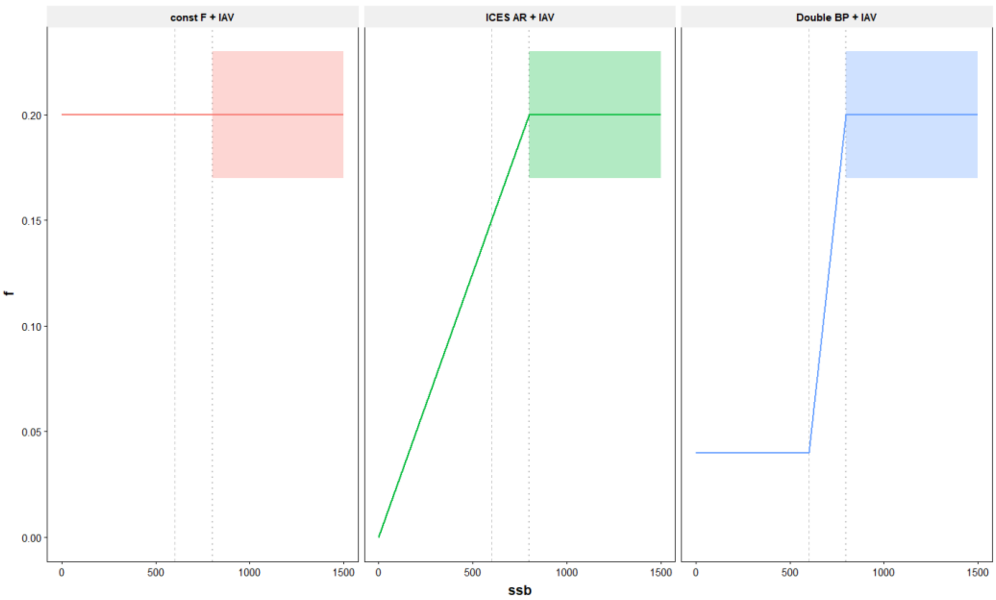
- Constant F strategy: fixed Ftarget independent of biomass level
- ICES Advice Rule: breakpoint at Btrigger and straight decline in F to zero below Btrigger.
- Double Breakpoint rule: breakpoint at Btrigger and straight decline in F to 20% of Ftarget at Blim. Below Blim continued fishing at F = 0.2 * Ftarget.

For each of the HCRs, a number of different target fishing mortalities were explored (0.0, 0.05, 0.075, 0.1, 0.125, 0.15). No evaluation of different Btrigger

[1] https://github.com/ices-eg/wk_WKREBUILD/tree/master/EqSimWHM

values was carried out, so that all evaluations used MSY Btrigger as the trigger point. All HCRs where evaluated with three variants:

- Without any additional constraints
- With a minimum TAC of 50 kT
- With a maximum 20% inter-annual variation (IAV) in TAC, but only when the stock is above Btrigger)

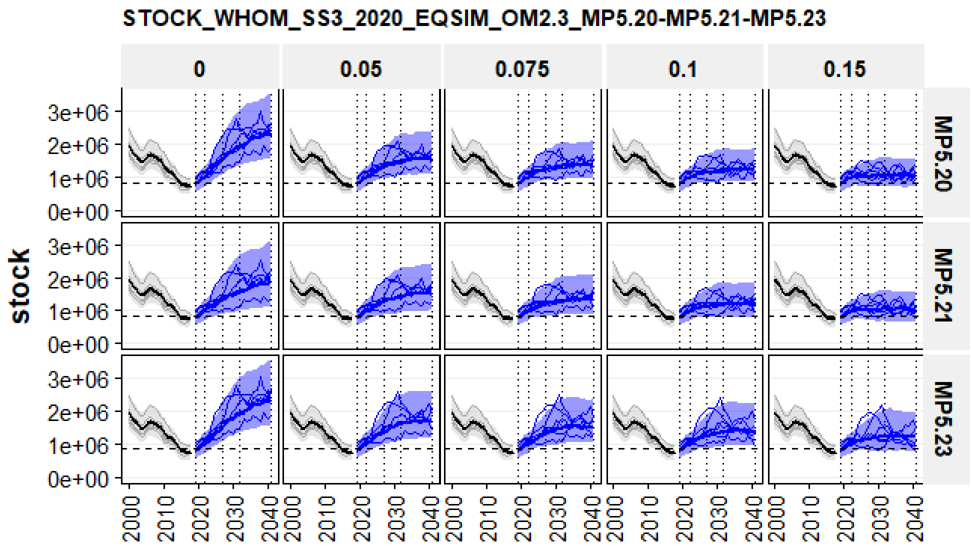


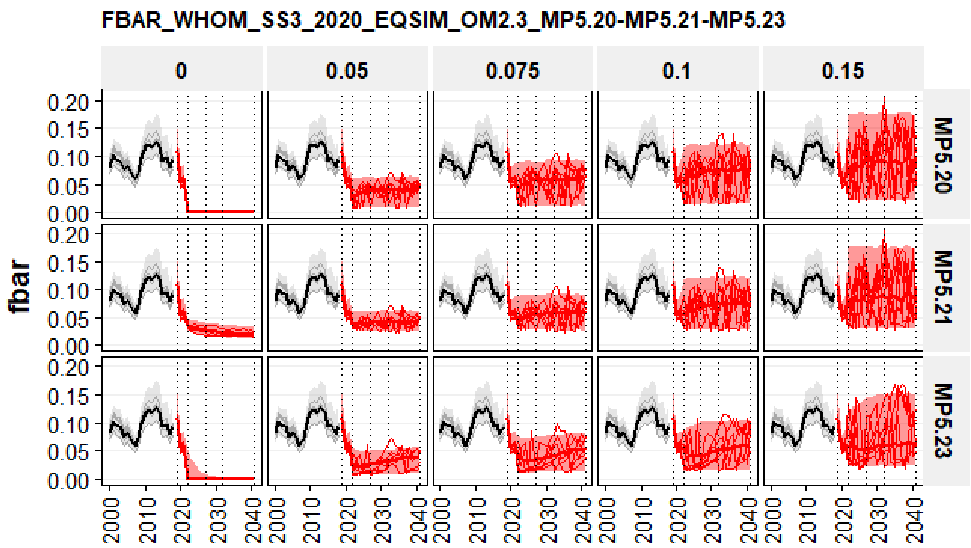
Presentation of input data has been included in Working Document sections **Error! Reference source not found.**, **Error! Reference source not found.** and **Error! Reference source not found.**. Presentation of the results has been included in the Working Document Annex (section 1). In the presentation of the results, we now report all the relevant metrics and we connected the recent past (from 2000 onwards) to the simulated futures. New figures have been included to present the first year of rebuilding to certain biomass reference points under different assumptions or conditions.

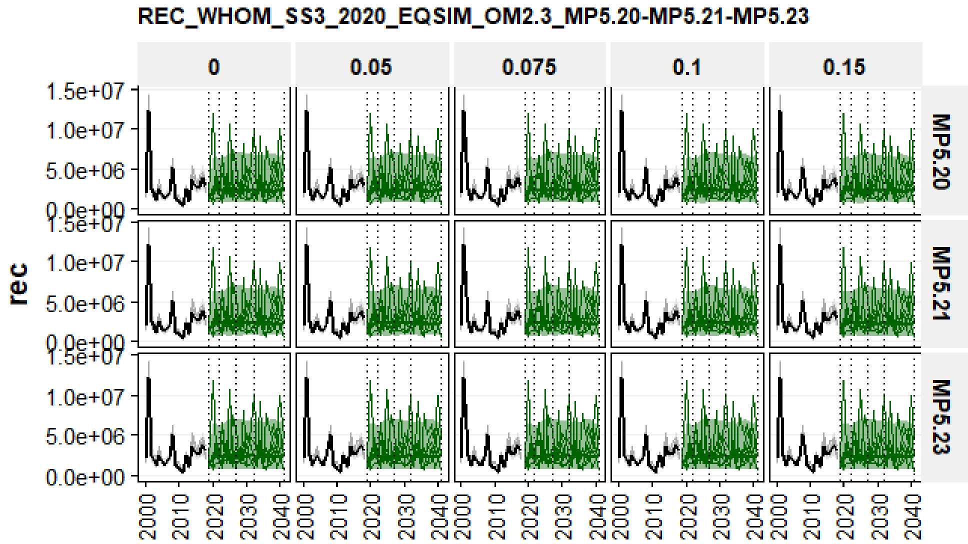
For each iteration, rebuilding is defined as being above Bpa for three consecutive years. When the probability of rebuilt stocks in a particular year for all iterations together increases above 50%, the stock is considered to have rebuilt.

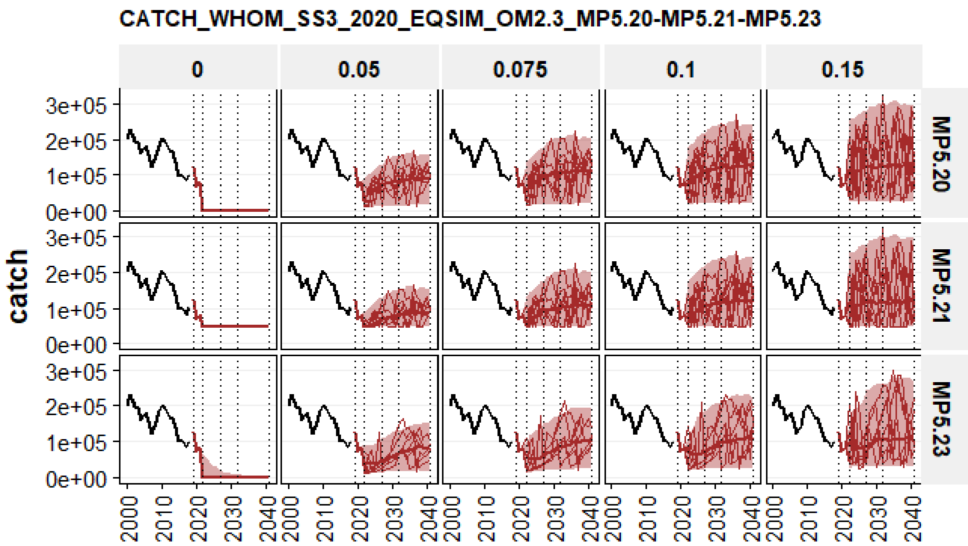
Main results are presented, using the 2020 SS3 assessment as the base case. Sensitivity exploration have been carried out for the 2019 assessments, similar to the original PELAC document.

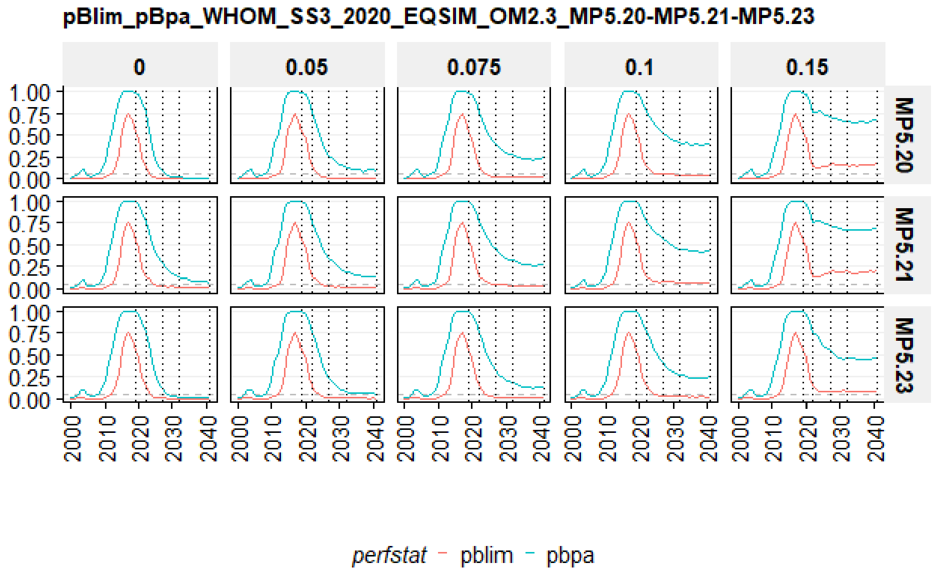
Results using the 2020 SS3 assessment for three scenarios (MP5.20 Double BP; MP5.21 Double BP with minimum TAC of 50kT; MP5.23 Double BP with 20% IAV constraint above Btrigger) and for five different target fishing mortalities (0, 0.05, 0.75, 0.1, 0.15) are presented for SSB, mean fishing mortality, recruitment, catch, probability of being below Blim or Bpa and rebuilding probability by year. Expected rebuilding at the target fishing mortality 0.075 using the double breakpunt rule MP5.23 is in year 2027.

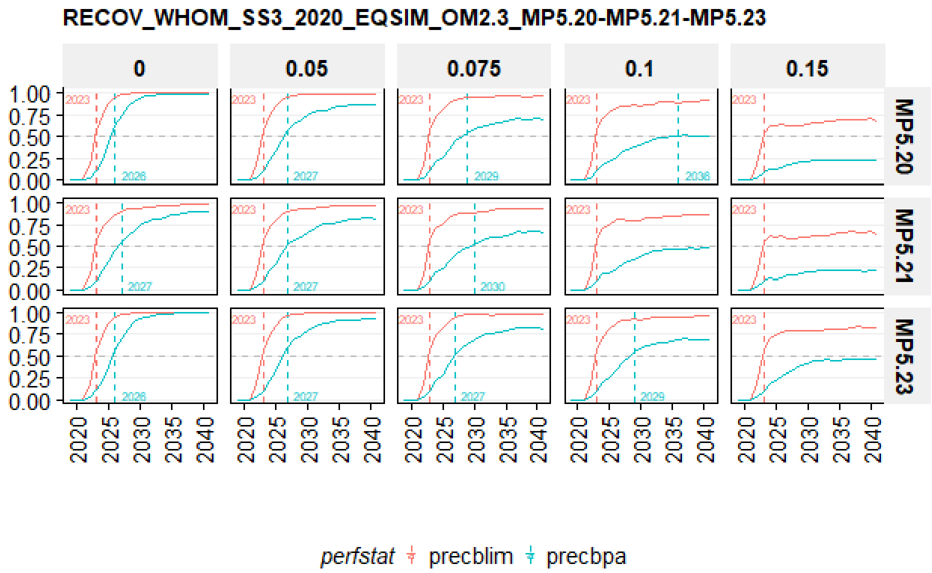

STOCK_WHOM_SS3_2020_EQSIM_OM2.3_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
stock
3e+06
2e+06
1e+06
0e+00
2000
2010
2020
2030
2040


FBAR_WHOM_SS3_2020_EQSIM_OM2.3_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
fbar
0.20
0.15
0.10
0.05
0.00
2000
2010
2020
2030
2040


REC_WHOM_SS3_2020_EQSIM_OM2.3_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
rec
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2000
2010
2020
2030
2040

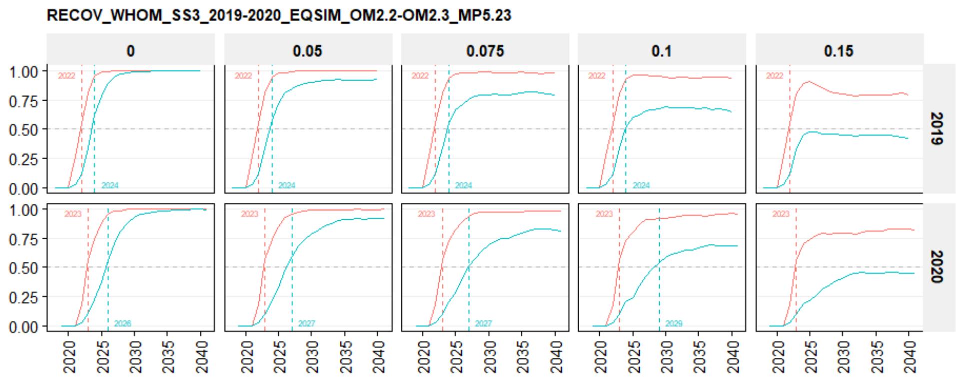

CATCH_WHOM_SS3_2020_EQSIM_OM2.3_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040





2) Exploration of the sensitivity of the results to the assessment year (2019 or 2020)

The exploration to the sensitivity to the assessment year has been explored both for the SS3 and SAM assessments. Results are summarized in Working Document sections **Error! Reference source not found.**, 1.4.1 and 1.4.2. Using the 2019 assessment as the OM for the simulations leads to a faster rebuilding to Bpa than using the 2020 assessment. For the target F of 0.075 (~Fmsy) with SS3, the rebuilding to Bpa would be in 2024 using the 2019 assessment and 2027 using the 2020 assessment. The main factor contributing to the delay is the lower estimates of recruitment in the 2020 assessment.



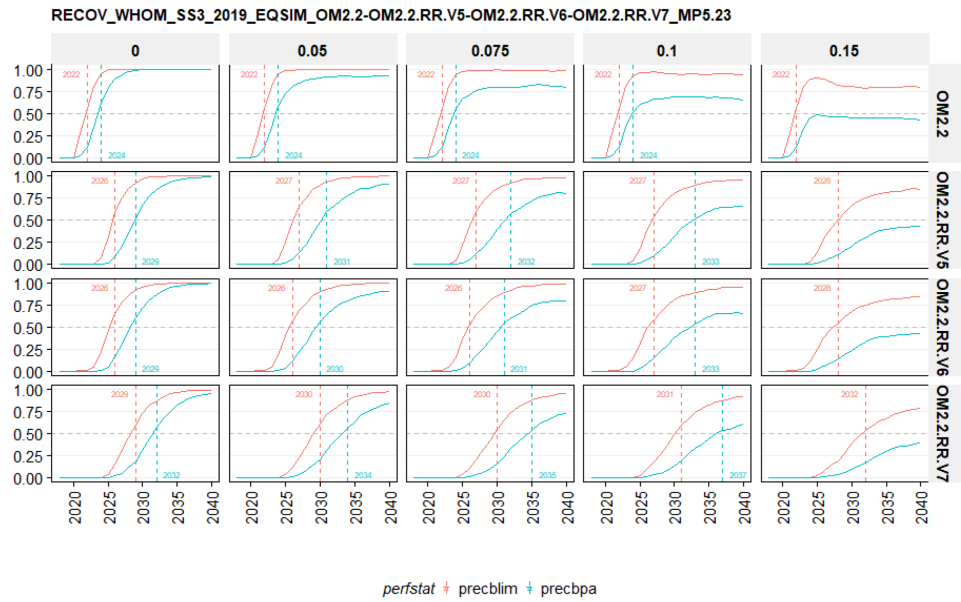
3) Exploration of the sensitivity to the recent (relatively stronger) recruitment

The exploration of sensitivity to the recent (stronger) recruitment has been explored both for the SS3 and SAM assessments, using the assessments of 2019 (note that this is different from the assessment that is used as the basis for the advice).

Results are summarized in Working Document sections **Error! Reference source not found.**, 1.4.3 and 1.4.4. A base case and three alternative recruitment assumptions were tested:

- Base case: use recruitment estimates directly from the assessment(s)
- Annual recruitments for 2014-2018 are equivalent to one half of the baseline estimate (i.e. that from the final assessment)
- Each of the 2014-2018 recruitments are assumed to be equivalent to the geometric mean of the time series from 2002-2013. Aside from 2008, this period is representative of the lowest recruitment observed.
- Each of the 2014-2018 recruitments are assumed to be equivalent to the mean of the 5 lowest recruitments from the full time series.

Under the reduced recruitment scenarios, risk to Blim is increased in the initial year of the simulation as a greater proportion of the replicates are below Blim. The effect of suppressing the recruitment for a period of 5 years leads to an initial further increase in risk in the first years of the simulation, during which a catch constraint applies for all replicates and management procedures. However, risk starts to reduce rapidly within 5 years for the most pessimistic scenario falling below 50% in 2023/24 for the baseline/2 and GM 20-14 scenarios and by 2027 for the scenario with the lowest recruitment. It can be concluded that lower recruitment scenarios are expected to lead to longer rebuilding periods, compared to the base case of the 2019 SS3 assessment. This is also confirmed by the exploration of using the 2020 assessment as the basis for the operating model instead of the 2019 assessment.



Under the preferred scenario by PELAC (Double BP with IAV constraint, target F=0.075), the first year of rebuilding to Bpa is expected around 2024 in the base case and would be delayed to 2030-2034 under the reduced recruitment regimes.

perfstat	scenario	recscenario	0	0.025	0.05	0.075	0.1	0.125	0.15
precbpa	SS3 2019	def	2024	2024	2024	2024	2024	2025	Inf
precbpa	SS3 2019	RR.V5	2029	2030	2031	2032	2033	2036	Inf
precbpa	SS3 2019	RR.V6	2029	2029	2030	2031	2033	2037	Inf
precbpa	SS3 2019	RR.V7	2032	2033	2034	2035	2037	Inf	Inf

4) Exploration of the sensitivity to the uncertainty and autocorrelation in advice error

The exploration to the sensitivity to the uncertainty and autocorrelataion in advice error has been explored for the SS3 assessment only. Results are summarized in section **Error! Reference source not found.**. In the focus group report (Pastoors et al., 2020), a review of the methods to generate cv and autocorrelation in advice error was presented. This concluded that the method of estimating advice error has some potential drawbacks. A similar conclusion has recently been reached in the North Sea herring assessment (HAWG 2021, forthcoming).

To explore the sensitivity of the results to the uncertainty (cv) and autocorrelation (phi), additional sensitivity simulations were conducted assuming "perfect knowledge" such that zero error is assumed in the assessment process and also with an arbitrary high correlated error (cv=0.6, phi=0.6). As expected, the removal of assessment error from the simulation leads to a lowering of the risk to Blim. The inflated and highly correlated error assumption ("High Err") scenario is associated with the highest risk. For both HCRs, risk reduces over the initial years due to the relatively strong recruitment between 2014 and 2018. In the PELAC select double BP rule, the difference between these scenarios is relatively small. A more discernible difference can be seen with the constant F rule where the use of the default values (with a higher autocorrelation in F error) is associated with an increase in risk to Blim.

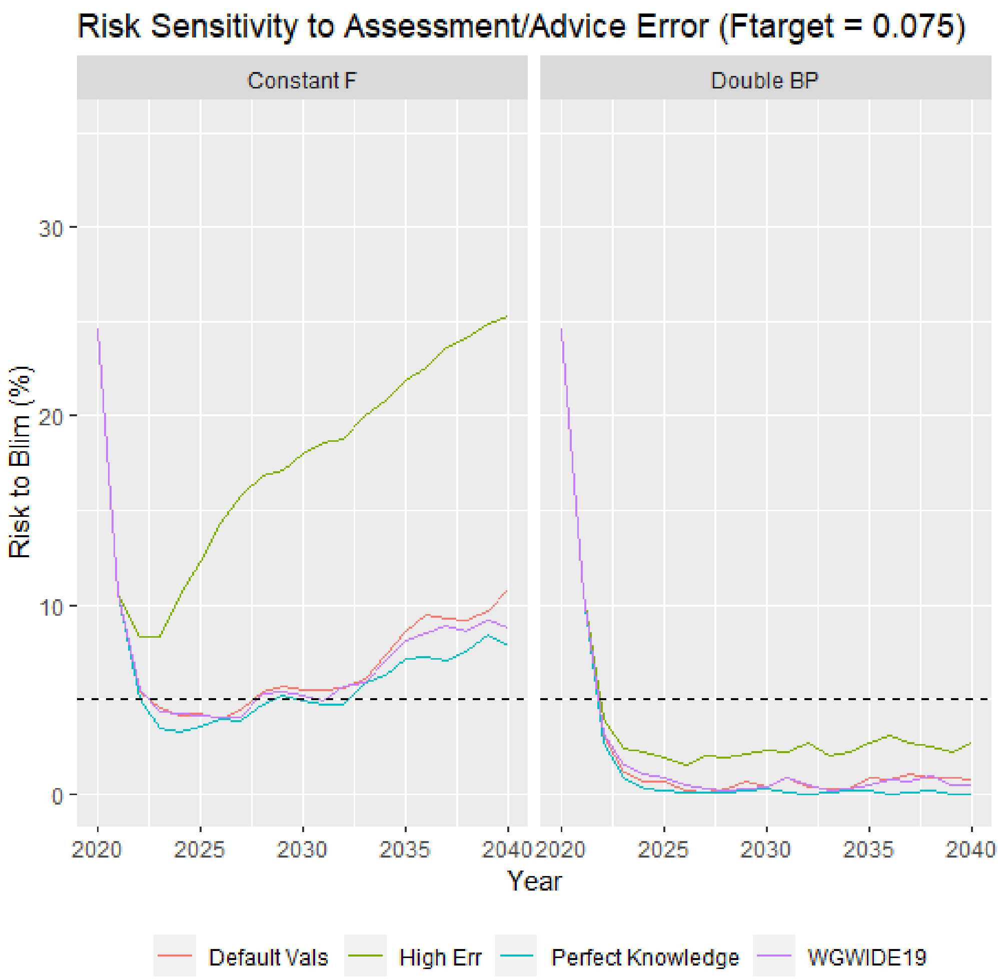


Figure 1 Comparison of the effect of varying assessment error parameters on the temporal development of risk to Blim for the constant F and double breakpoint HCR types (Ftarget = 0.075, TAC change limits of 20% when above Btrigger).

5) Exploration of the sensitivity to the reference points used

The exploration to the sensitivity to the reference points used has been explored for both the SS3 and SAM assessments. Results are summarized in sections **Error! Reference source not found.** and 1.4.5. The biomass reference points in the SAM assessment are lower than the biomass reference points in the SS3 assessment. This is due to the overall lower SSB and higher F in the SAM assessment compared to SS3. As could be expected, the application of the SS3 reference points to the SAM assessment lead to a substantial longer period of rebuilding that is required (because the bar is set higher) and vice versa for the SS3 assessment with SAM reference points.

A table linking the MSE work in this working document the guidelines for MSE from WKGMSE has been included (Working Document section **Error! Reference source not found.**) and similarly for the guidelines on rebuilding plans from WKREBUILD has been included (Working Document section 0).

BACKGROUND		
Motive/initiative/background.	There is no agreed management plan currently in place for western horse mackerel. The Pelagic Advisory Council (PELAC) initallly proposed a plan in 2007 and this was used for the purposes of catch advice until 2010. Since this time, advice has been based on the MSY approach and the PELAC has continued efforts to develop a new management plan. The evaluation that is presented, has been conducted by a group of experts, on request from the PELAC. On the basis of the evaluation, the PELAC has selected a preferred management procedure.	
Main objectives	Rebuilding of the stock to levels well above Blim. Recently the stock has been only just above Blim (median value). Stable and sustainable catches also during periods of lower than average recruitment.	
Formal framework	EC request to ICES to evaluate the MSE by the PELAC technical focus group	
Who did the evaluation work	PELAC technical focus group on horse mackerel (Pastoors et al., 2020)	
Method		
Software Name, brief outline include ref. or documentation	1. A further adaptation of SimpSIM which was, initially developed from the *EqSim* and FLR R packages for the 2016 blue whiting MSE (ICES, 2016). Code of the adapted version is on github (https://github.com/ices-eg/wk_WKREBUILD; folder EqSimWHM) . The *EqSim* based simulator was conditioned on the SS3 assessment model as carried out during WGWIDE 2019. Additional runs were also conditioned with the WGWIDE 2020 assessment and with the exploratory SAM assessments (2019, 2020) as reported in the relevant expert groups (ICES, 2019; ICES, 2020). The *EqSim* simulator is based on an age-structured operating model, with catches, numbers and F at age derived from simulated assessments. Assessment/advice error on SSB and F is incorporated following guidelines from WKMSYREF3 (ICES, 2014) 2. SAM HCR forecast based on the exploratory SAM assessment. The WHOM assessment model scripts and results used in this study are available at https://www.stockassessment.org/set-Stock.php?stock=WHOM2019. The HCR forecast in SAM that is used in this study was first developed and presented in Trijoulet et al. (Trijoulet et al., 2021). The code is available at https://github.com/vtrijoulet/SAM/tree/master2 and the main function is forecast2(). The forecast is a stochastic forecast that allows HCRs in the management procedure. It was set up to replicate as closely as possible the settings in the *EqSim* simulator.	
Type of stock	Medium life span (>20 years), pelagic	
Knowledge base *	Analytic assessment (SS3) and exploratory assessment (SAM)	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Segmented regression with a breakpoint at Blim. Four scenarios considered (base case and three alternative (more pessimistic) scenarios whereby recent increases in recruitment (since 2014) is reduced).	Log-normal, CV from residuals
Weight at age	Random selection from stochastic time series of catch/stock weight at age.	Weight at age estimates from each of the 1000 model replicates were used to generate a time series stochastic weights via application of an error term derived from variability in the historic weight at age datasets to the replicate mean. The error term is considered age specific and autocorrelated.
Maturity	Maturity during the simulations is time invariant (as in the assessment)	No
Natural mortality	0.15 – all ages and years	No

Selectivity	From assessment	The MSE draws at random from the 1000 individual selection profiles available from the replicates for each year of the projection period.
Initial stock numbers	From assessment	Using 1000 model iterations to derive starting values.
Reference points	OM assessment specific reference points are estimated. because of different partioning between SSB and F for the exploratory SAM and SS3 assessmens	
	WGWIDE/SS3 reference points Blim 834,480t Bpa 1,168,272t Fmsy 0.074	SAM reference points Blim 611,814t Bpa 856,540t Fmsy 0.115
Decision basis **	SSB in the TAC year	
Number of iterations	1000	
Projection time	23 years	
Observation and implementation models (no assessment in the loop)		
Type of noise	EqSim: historic advice error on F and SSB SAMHCR: uncertainty directly taken from SAM assessment (function: simulate)	
Comparison with ordinary assessment?	Advice uncertainty is larger than model parametric uncertainty (SS3 estimates relative low uncertainties). The historic advice error for the last ten years was derived according to the methodology detailed in WKMSYREF3 (ICES, 2014) SAMHCR is based on uncertainty estimated within the SAM model.	
Projection: If yes - how?	*EqSim*: no forecast included. TAC derived directly from SSB estimate taking into account the advice error that was estimated for the forecast. SAMHCR: is essentially a multi-annual forecast using different HCR types.	
Projection: Deviations from WG practice?	Yes, WG runs a deterministic projection based on the SS3 model	
Implementation	Catches in numbers at age from projection according to the rule. No implementation error assumed.	
Harvest rule		
Harvest rule design	Proposed HCR: Double breakpoint on SSB: If SSB < Blim, Flow = 0.2 * Ftarget If Blim<SSB<Bpa: linear slope between Flow and Ftarget If SSB > Bpa: F = Ftarget Fixed F and ICES advice rule (with breakpoint at Bpa) HCRs were also evaluated.	
Stabilizers	Optionally, 20% TAC change limits (only applied above $B_{trigger}$ (B_{pa}) Scenarios with fixed minimum catch (50 kT) were run.	
Duration of decisions	Annual	
Revision clause	LTMP specifies a revision in no more than 5 years from first year of implementation.	
Presentation of results		
Interest parameters	Rebuilding probability above biomass reference points. Rebuilding was defined as at least three consecutive years with at least 50% probability of being above the biomass reference point (Blim or Bpa) Catch, Inter-annual variation	
Risk type and time interval	Type 3	
Precautionary risk level	5%	
Experiences and comments		
Review, acceptance:	Under review (WKWHMRP 2021)	
Experiences and comments	Running two different operating models (SS3 and SAM) was useful to explore the sensitivity of the rebuilding probability to the different model formulations. Running two different evaluation methods (EqSim and SAMHCR) was useful to explore how the results of the management procedure are dependent on the evaluation method. Model results were comparable but not identical. Main difference is due to the different perception of the stock relative to reference points.	

Form of rebuilding plan in the light of WKREBUILD

WKREBUILD (2020) made a list of recommendations for rebuilding plan guidelines that will be further refined in the future as cases studies are developed. In this context, we have noted below the recommendation made by WKREBUILD in 2020 and added comments on how this work fits into the recommendations or can help refining them.

Table 1.1. Recommendations for guidelines from WKREBUILD (2020) and comments from the rebuilding plan evaluation carried out for WHOM.

Criteria	Recommendations from WKREBUILD	Comments from work in WKWHMRP
Rebuilding targets	Defining rebuilding biomass and fishing targets is critical to the evaluation of rebuilding plans and should be clearly defined at the beginning of the evaluation. While ICES is currently using Blim as the limit reference point, MSY Btrigger as the trigger reference point, and FMSY as the target fishing mortality; other targets could also be considered in a rebuilding plan if relevant.	A number of different target fishing mortalities were explored (0.0, 0.05, 0.075, 0.1, 0.125, 0.15). Simulations used Blim as the limit reference point and MSY Btrigger as the trigger reference point. Evaluations of different values for the biomass reference points were not evaluated. Risk was evaluated with Blim as target biomass.
Reference points	WKREBUILD raised some concerns regarding the estimation of reference points in ICES. If used in the evaluation of the rebuilding plan, reference points must be suitable, i.e. consistent with the cur-rent productivity of the stock at low SSB and current environmental conditions. Evaluation of rebuilding targets that may differ from current reference points, may be necessary for depleted stocks when those were estimated including periods of high productivity and optimistic stock-recruitment relationships.	FMSY was estimated for both simulation methods (SAM and EqSim) using recruitment pairs from 1995-2017. The reference points were therefore consistent with the simulation method used and taking recruitment from 1995 excluded the early years of very high productivity. FMSY is therefore consistent with the current productivity of the stock.
Time frame leading to a rebuilding plan	Development of rebuilding plan options should be initiated, when not already in the management plan, as soon as the median SSB of a stock is estimated below MSY Btrigger at the beginning of the advice year and the forecast based on the ICES rule (F from the slope) does not reverse the decline in SSB at the end of the forecast year. The effect of retrospective patterns on the possible future forecast of stock biomass should be taken into account when this is deter-mined.	Work on the rebuilding plan evaluation was initiated in 2019 when the stock was estimated between MSY Btrigger and Blim (2018 advice). After re-evaluation of reference point at the 2019 inter-benchmark, the stock was estimated below MSY Btrigger and close to Blim. Since then (2020 assessment), the stock is still above but close to Blim. Therefore, the development of the rebuilding plan was initiated when the stock was below MSY Btrigger at the beginning of the advice year and was still below it at the end of the forecast year (2019 advice) as recommended by WKREBUILD.
	Implementation of a rebuilding plan, when not already in the management plan, should be advised if the median SSB of a stock is estimated below Blim at the beginning of the advice year and the forecast based on the ICES rule (F from the slope) does not allow the stock to get above Blim at the end of the forecast year.	The expert group advise on the implementation of the rebuilding plan despite the fact that the stock is currently close to but not below Blim, because there are concerns regarding retrospective patterns in the model that might bring the stock below Blim in recent future.

Time frame leading out from a re-building plan	The exit strategy should be embedded in the rebuilding plan. Leading out from the rebuilding plan too early or too late should be avoided. The exit strategy should preferably contain element on how to ensure a "smooth" transition between the rebuilding phase and the post-rebuilding phase (i.e., ICES advice rule or a LTMP) to reduce the risk of inversion of positive trends. The exit from a rebuilding plan should be robust to uncertainty in the estimation of the stock status to reduce the risk of falling back to a rebuilding phase soon after the exit. Robustness to uncertainty could include setting a certain probability of SSB being above re-building reference points, being above rebuilding targets for a number of consecutive years, a consistent positive trend in SSB, evidences of a strong year class confirmed by independent observations (i.e., survey and commercial fishery) and through time. Maintaining F below FMSY for a sufficient time (at least one generation) then smoothly transitioning to FMSY could also be a possible strategy to exit a re-building plan.	The rebuilding plan may be superseded by a long-term strategy for the stock when, according to ICES, the spawning stock biomass is assessed to have been above Bpa for three consecutive years.
Time frame for the evaluation of a re-building plan	The evaluation period represents the time window between TMIN and TMAX which is used to assess the level of re-building achieved by alternative rebuilding strategies. TMIN is defined as the time taken for the stock to rebuild with zero fishing to above Blim, or the agreed rebuilding target with 95% probability, or other level of probability depending on the state of depletion of the stock. TMAX, defined as the maximum amount of time for rebuilding the stock, is usually specified by managers/requesters but could be ex-pressed as x* TMIN with x > 1. WKREBUILD was not able to conclude on a value for x in the estimation of TMAX. x=2 is often used in other jurisdictions.	Tmin could be derived from the zero catch options in the simulations. Tmin may be dependend on the assumptions used during the simulations, e.g. the base year for the simulations and the recruitment assumptions.
Checking the progress of the rebuilding plan	Re-evaluation of the rebuilding plan may be necessary if the stock trajectory is outside the range of expected performance relative to timelines of the rebuilding plan or if other exceptional circumstances arise such as unexpected data or a new understanding of the stock. The new rebuilding plan evaluation will need to adapt to the new data or findings. A re-evaluation of the rebuilding targets or objectives may also be necessary.	Checks on the stock response to the re-building plan if accepted will be done annually with the update assessment. If the stock patterns deviate from the predictions, the rebuilding plan will be re-evaluated. If changes in data occur when the rebuilding plan is in place, the HCR will be also re-evaluated given the assessment results given the new data.

Probability of achieving rebuilding	The default probability for rebuilding above the target is 95% but for certain stocks a lower probability may be more relevant in the short- to medium-term depending on the nature of the fishery and socio-economic considerations. This would be notably relevant for short-lived stocks with high recruitment variability that are estimated to be below Blim with a probability larger than 5% even if unfished	The probability to fall below Blim was estimated every year of the forecast. Rebuilding to Blim or Bpa was assumed to have occurred when three consecutive years above the reference points were achieved with a probability of at least 50%.
Harvest rules in rebuilding phase	Several harvest rules should be evaluated during a rebuilding plan. These should be compared against the zero catch scenario and the ICES advice rule.	These was respected. A total of 3 HCRs were compared and all runs included a F=0 scenario.
Evaluation tools	Rebuilding plans necessitate a prompt management response. Evaluation tools should be available when the evaluation starts. Multiple tools already exist to evaluate rebuilding plans. Rebuilding plan evaluation should use tools that have been reviewed or validated.	The evaluation used two existing tools (EqSim simulator and SAM HCR forecast) that were further developed for the exercise.
Uncertainty considerations	Alternative operating models should be evaluated to account for stock specific uncertainties. Typical uncertainties to consider in the rebuilding plan context are uncertainties in stock productivity (e.g. recruitment), in the assessment model (e.g. stock perception, bias such as retrospective patterns) and in implementation error.	The main method (EqSim) considered uncertainty in: - Initial population size - Stock & catch weights - Selectivity at age - Recruitment - Assessment Model The SAM HCR forecast considered uncertainty in: - Current perception of the stock (simulated starting values for stock and fishing mortality taken from the assessment results) - Uncertainty in recruitment (segmented regression with fixed inflection point at Blim and variance in stock-recruitment relationship) - Process errors in stock survival (same as in the assessment)
Special considerations	The context of the rebuilding plan may be framed based on mixed stocks, mixed fisheries and socio-economic objectives.	[more to come on stock structure implications]

References

ICES. 2014. Report of the Joint ICES-MYFISH Workshop to consider the basis for FMSY ranges for all stocks (WKMSYREF3), 17–21 November 2014, Charlottenlund, Denmark. ICES C.M. 2014 / ACOM:64.

ICES. 2016. Report of the Workshop on Blue Whiting Long Term Management Strategy Evaluation (WKBWMS), 30 August 2016, Copenhagen. ICES C.M. 2016 / ACOM: 53.

ICES. 2019. Interbenchmark Protocol on reference points for western horse mackerel (Trachurus trachurus) in subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a-c,e-k (the Northeast Atlantic) (IBPWHM). ICES VOLUME 2 | ISSUE 95.

ICES. 2020. Working Group on Widely Distributed Stocks (WGWIDE 2020) Volume 2, Issue 82.

Pastoors, M. A., Campbell, A., Trijoulet, V., Skagen, D., Gras, M., Lambert, G., Sparrevohn, C. R., et al. 2020. Western Horse Mackerel Technical Focus Group on Harvest Control Rule Evaluations 2020. 43 pp pp.

Trijoulet, V., Berg, C. W., Sparrevohn, C. R., Nielsen, A., Pastoors, M. A., and Mosegaard, H. 2021. Can we fish on stocks that need rebuilding? Illustrating the trade-offs between stock conservation and fisheries considerations. bioRxiv preprint, 2021.2002.2025.432880.

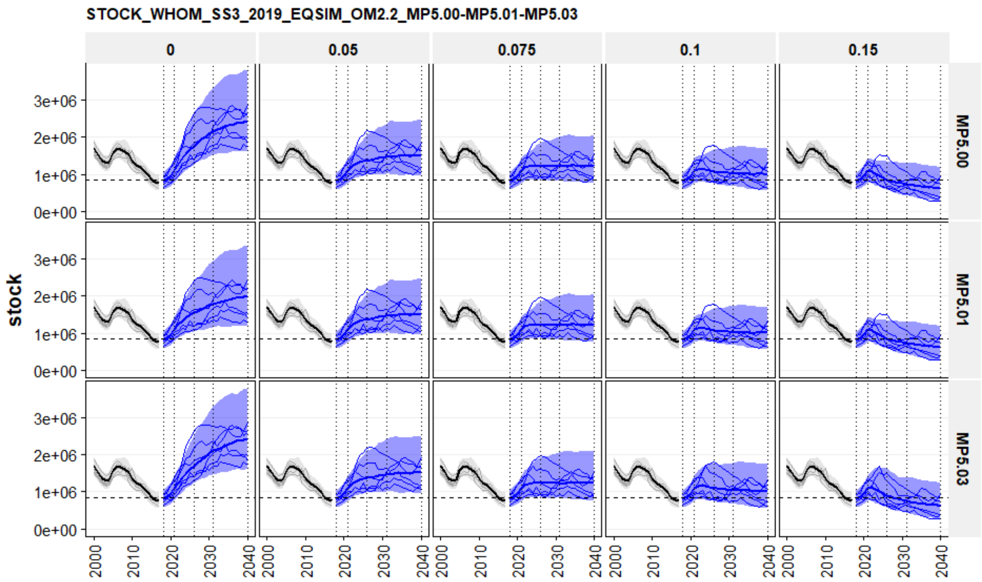
1 Annex: detailed simulation results

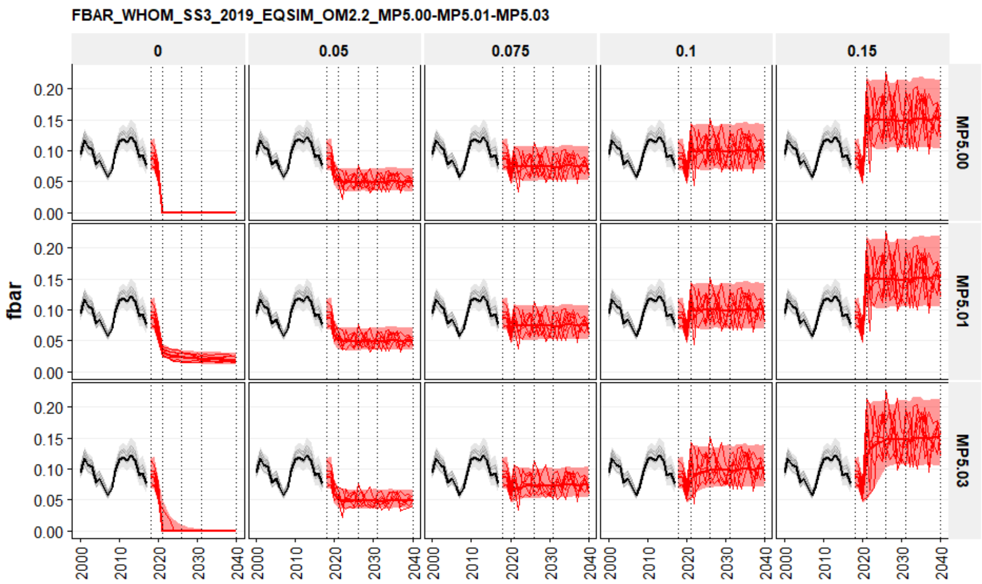
1.1 SS3 and EqSim

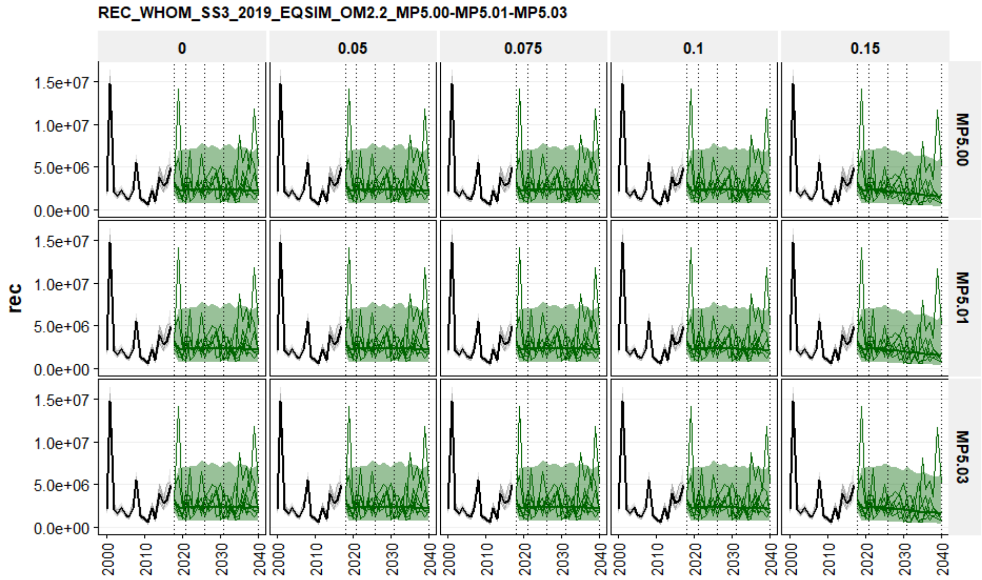
1.1.1 Constant F strategy

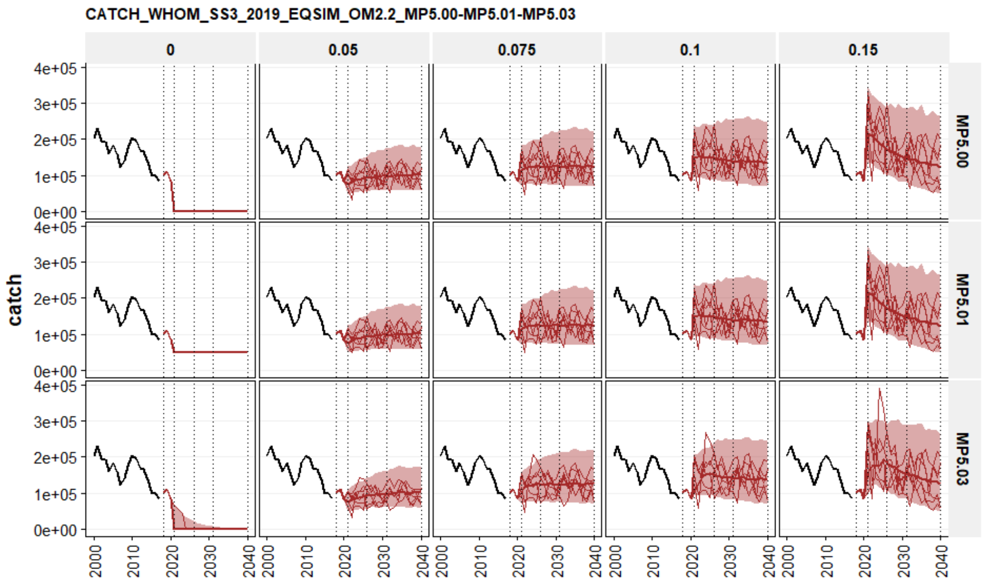
Simulation summary 2019 assessment, SS3, EqSim.

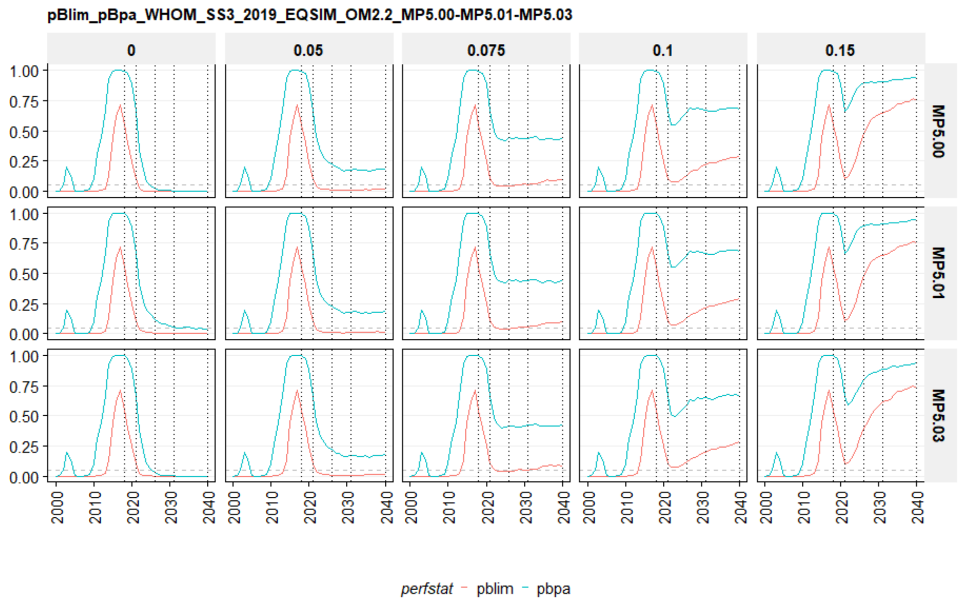
MP5.00 constant F; MP5.01 constant F with minimum TAC of 50kT; MP5.03 constant F with 20% IAV constraint above Btrigger.






REC_WHOM_SS3_2019_EQSIM_OM2.2_MP5.00-MP5.01-MP5.03
0
0.05
0.075
0.1
0.15
MP5.00
MP5.01
MP5.03
rec
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2000
2010
2020
2030
2040


CATCH_WHOM_SS3_2019_EQSIM_OM2.2_MP5.00-MP5.01-MP5.03
0
0.05
0.075
0.1
0.15
MP5.00
MP5.01
MP5.03
catch
4e+05
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040



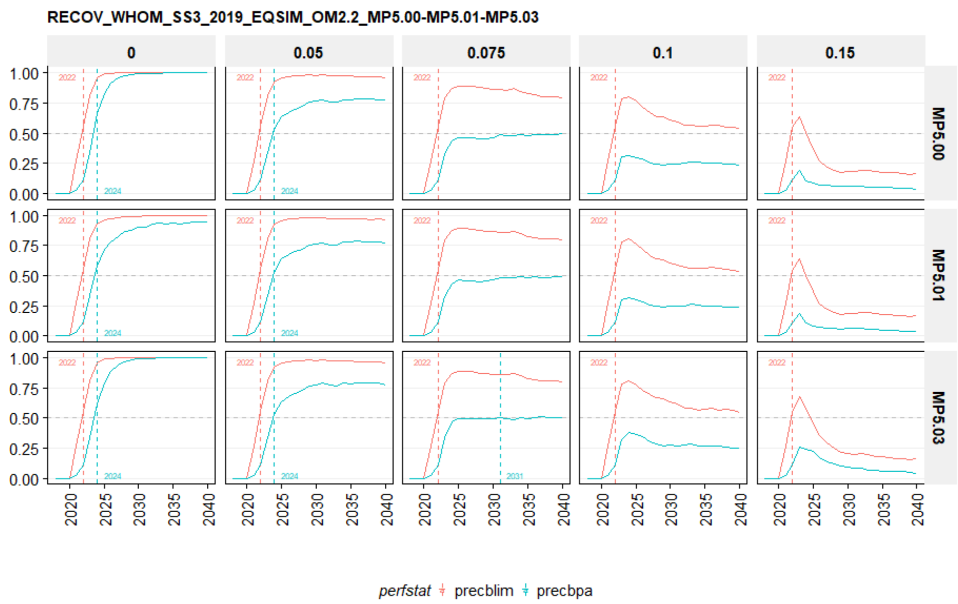
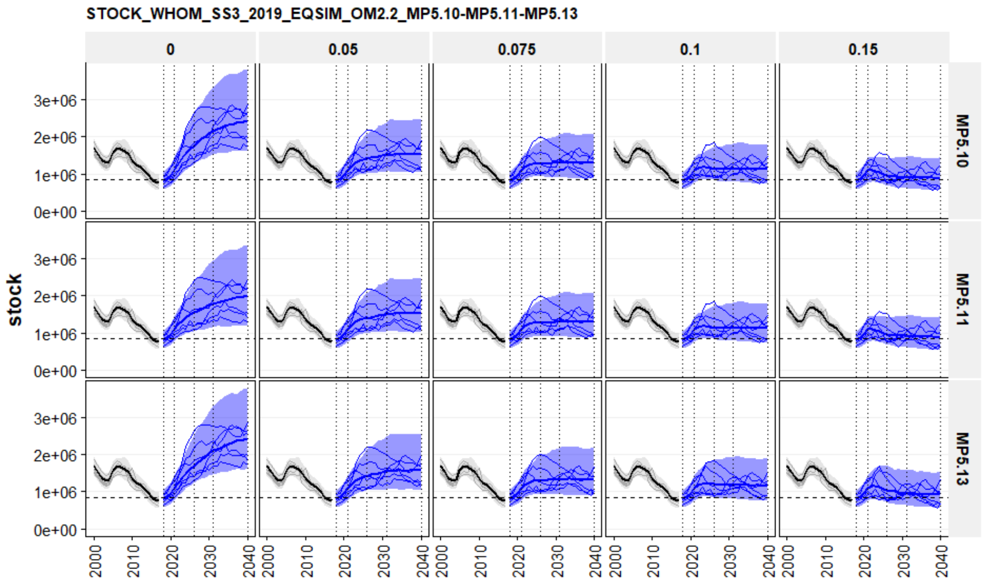


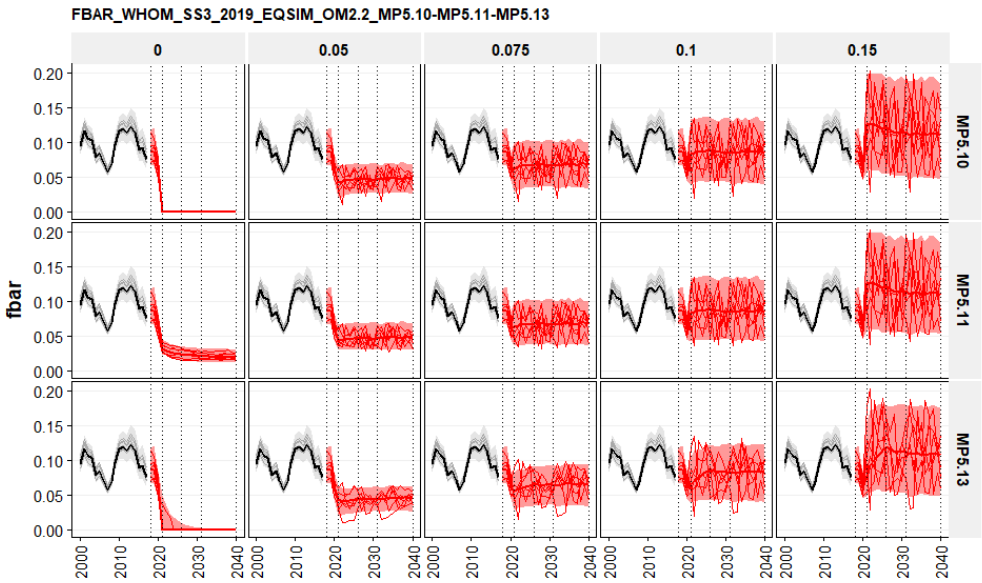
Figure 2 Simulation summary 2019 assessment, SS3, EqSim. MP5.00 constant F; MP5.01 constant F with minimum TAC of 50kT; MP5.03 constant F with 20% IAV constraint above Btrigger.

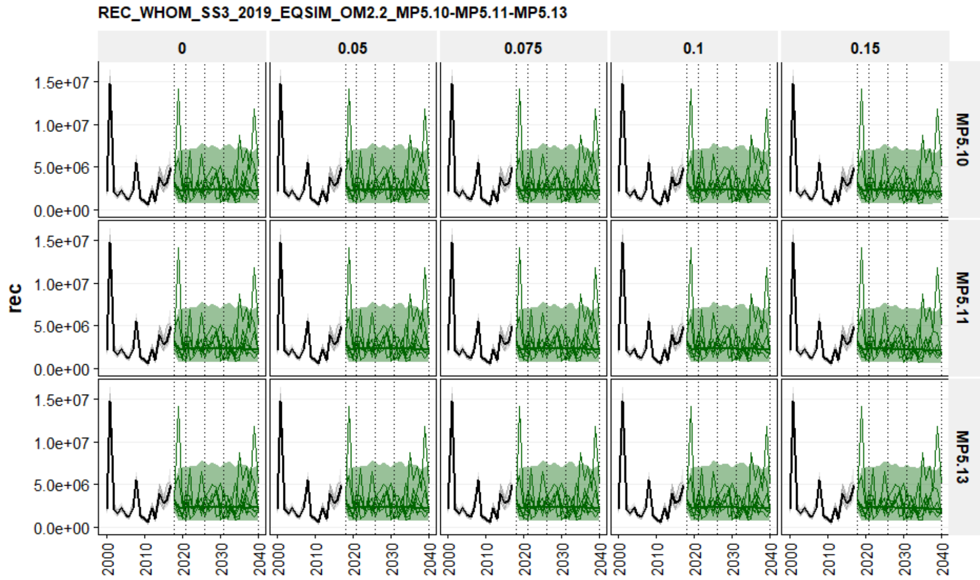
1.1.2 ICES advice rule

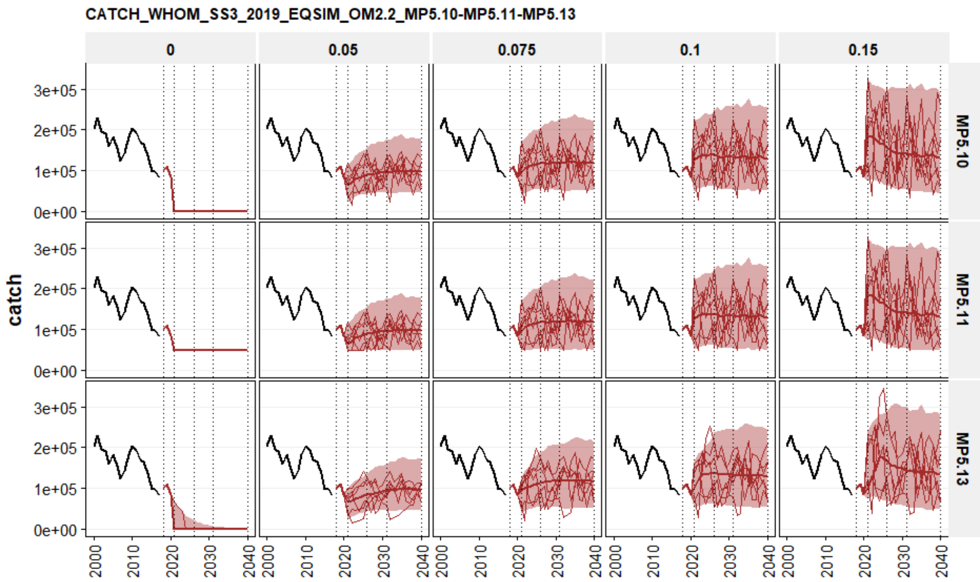
Simulation summary 2019 assessment, SS3, EqSim.

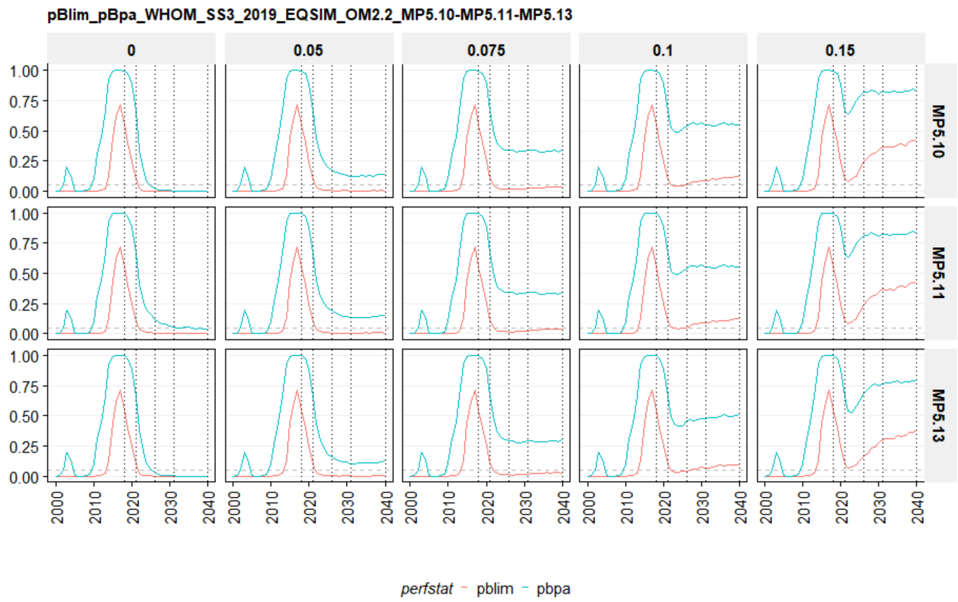
MP5.10 ICES AR; MP5.11 ICES AR with minimum TAC of 50kT; MP5.13 ICES AR with 20% IAV constraint above Btrigger.






REC_WHOM_SS3_2019_EQSIM_OM2.2_MP5.10-MP5.11-MP5.13
0
0.05
0.075
0.1
0.15
MP5.10
MP5.11
MP5.13
rec
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2000
2010
2020
2030
2040


CATCH_WHOM_SS3_2019_EQSIM_OM2.2_MP5.10-MP5.11-MP5.13
0
0.05
0.075
0.1
0.15
MP5.10
MP5.11
MP5.13
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040



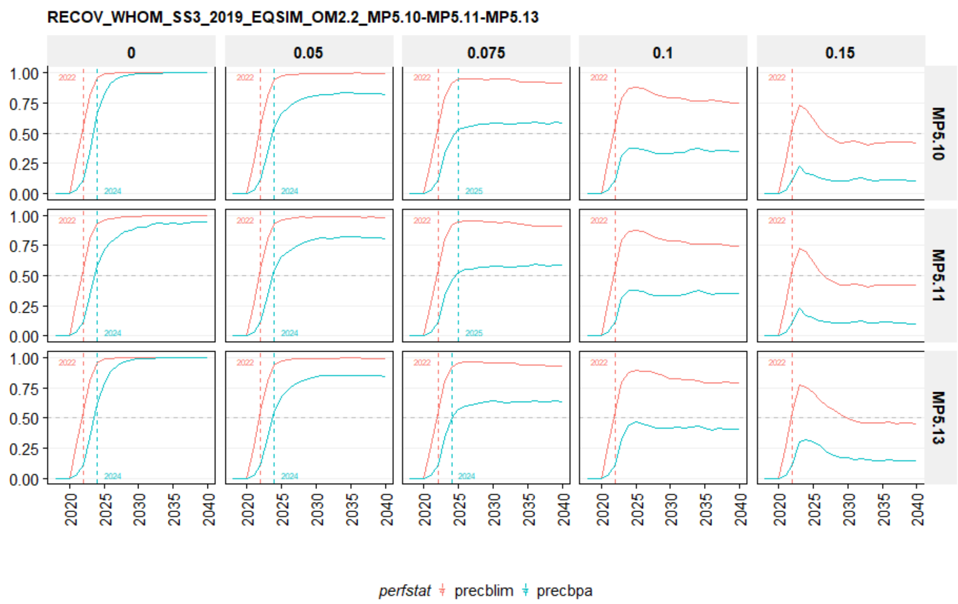
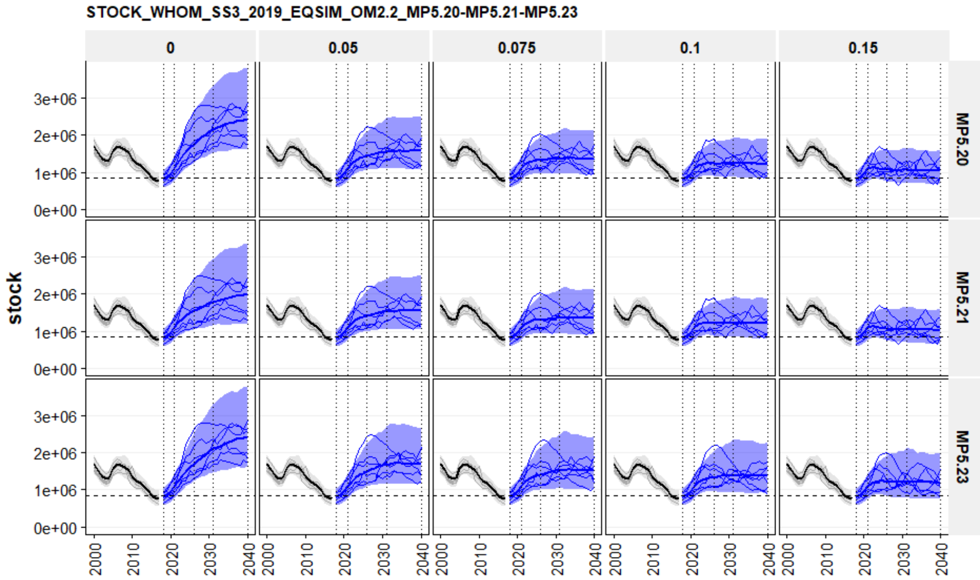


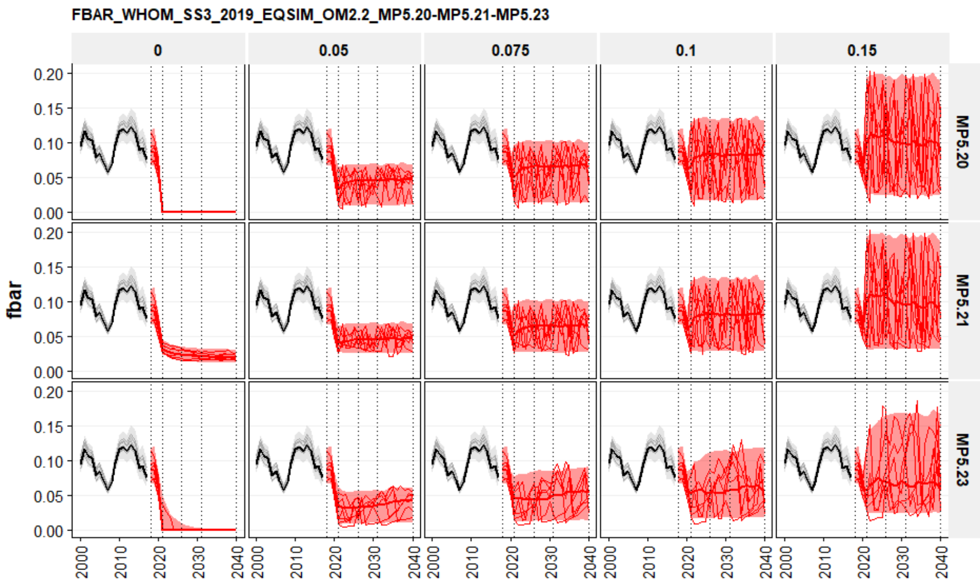
Figure 3 Simulation summary 2019 assessment, SS3, EqSim. MP5.10 ICES AR; MP5.11 ICES AR with minimum TAC of 50kT; MP5.13 ICES AR with 20% IAV constraint above Btrigger.

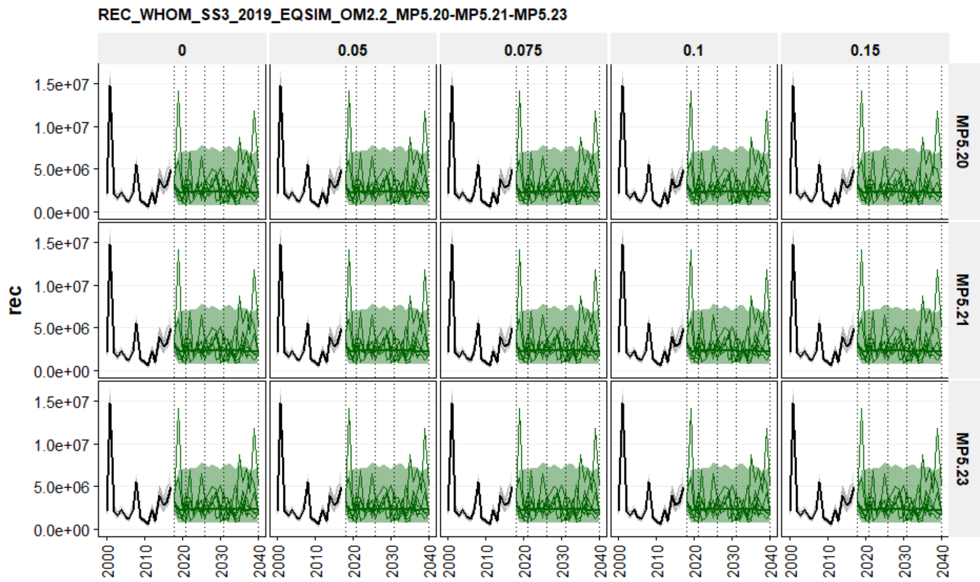
1.1.3 Double breakpoint rule

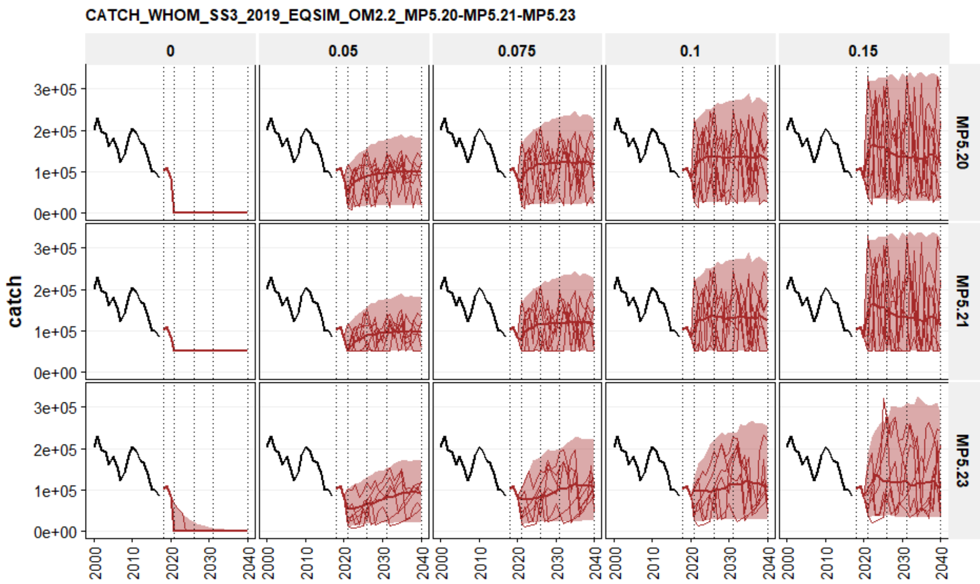
Simulation summary 2019 assessment, SS3, EqSim.

MP5.20 Double BP; MP5.11 Double BP with minimum TAC of 50kT; MP5.13 Doubl BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.






REC_WHOM_SS3_2019_EQSIM_OM2.2_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
rec
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2000
2010
2020
2030
2040


CATCH_WHOM_SS3_2019_EQSIM_OM2.2_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040

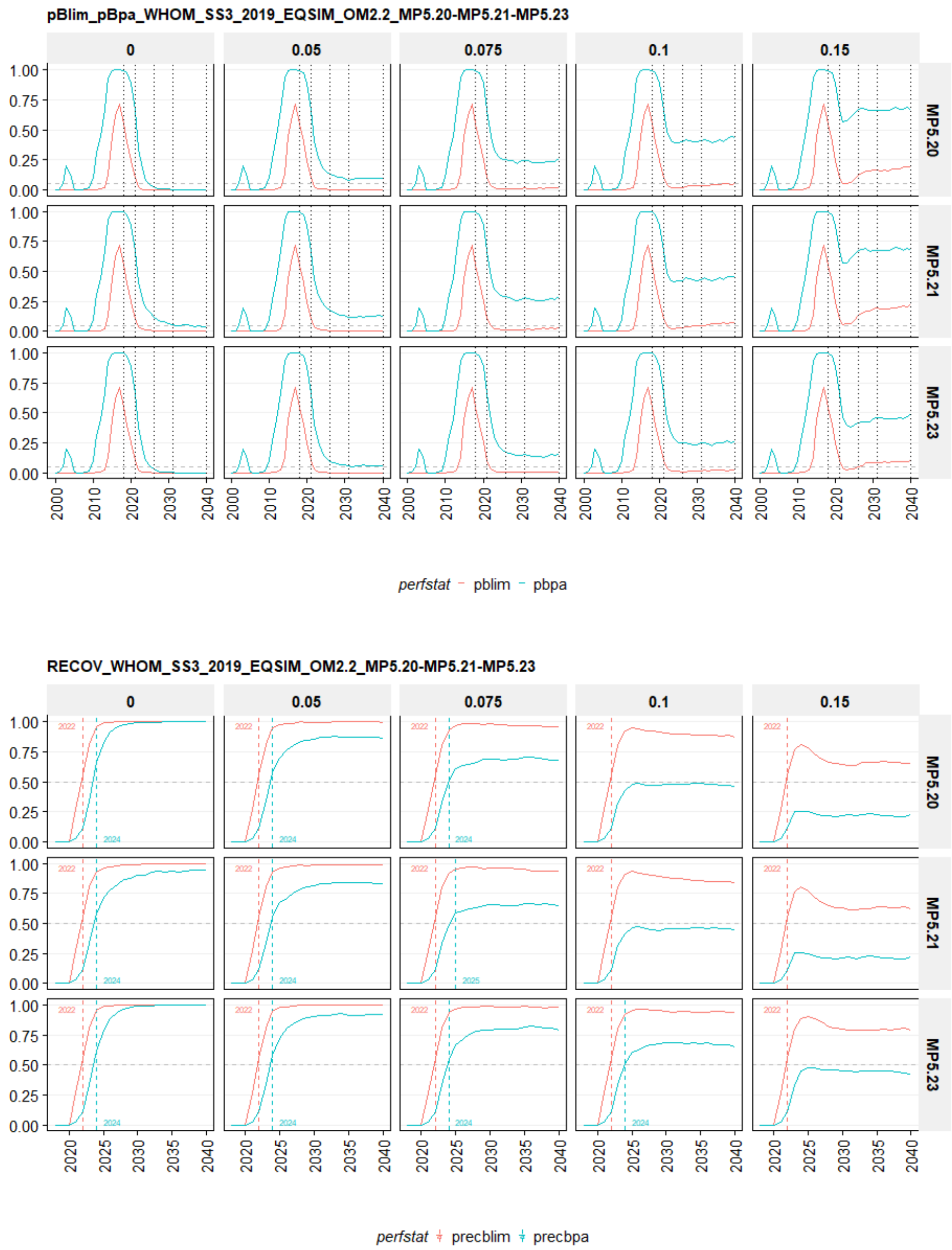


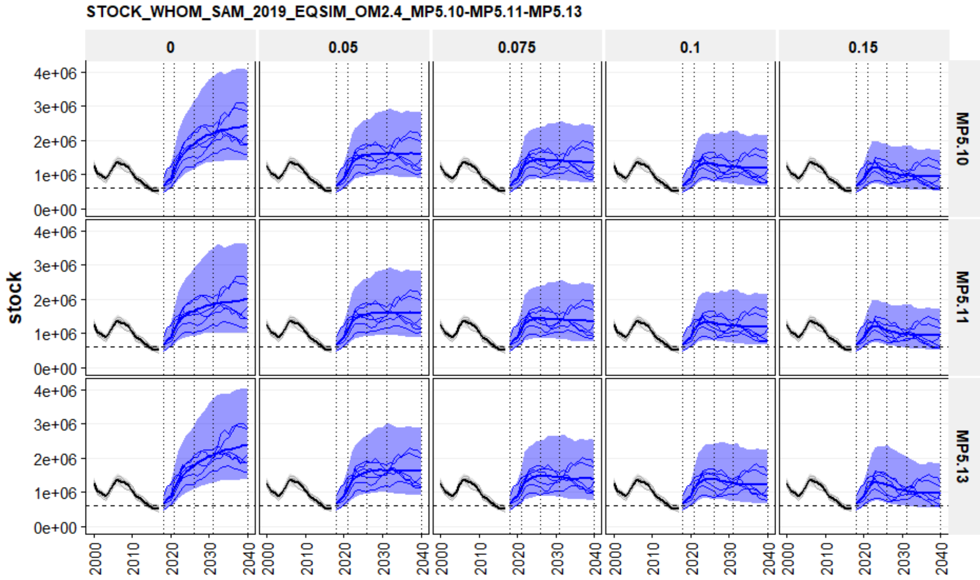
Figure 4 Simulation summary 2019 assessment, SS3, EqSim. MP5.20 Double BP; MP5.21 Double BP with minimum TAC of 50kT; MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

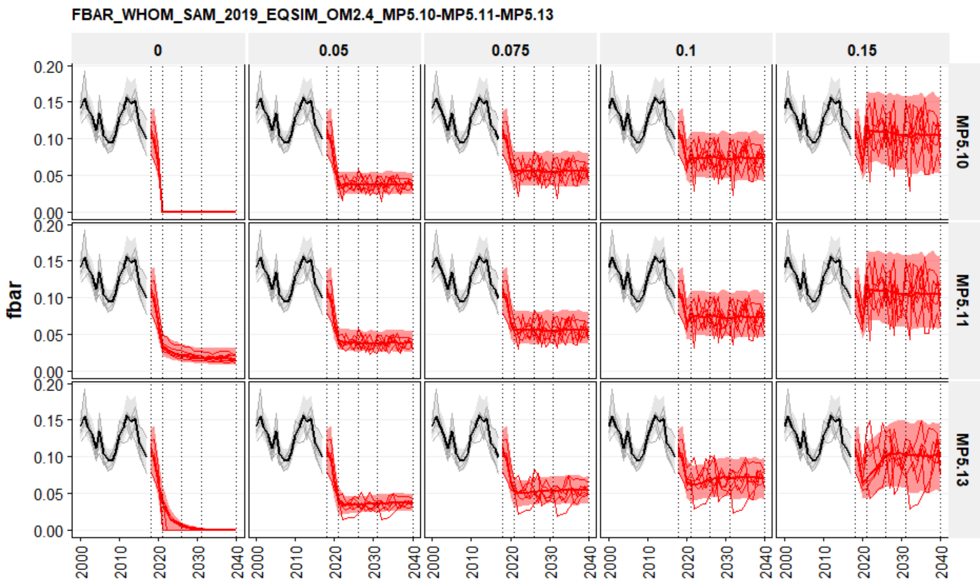
1.2 SAM assessments in EqSim simulator

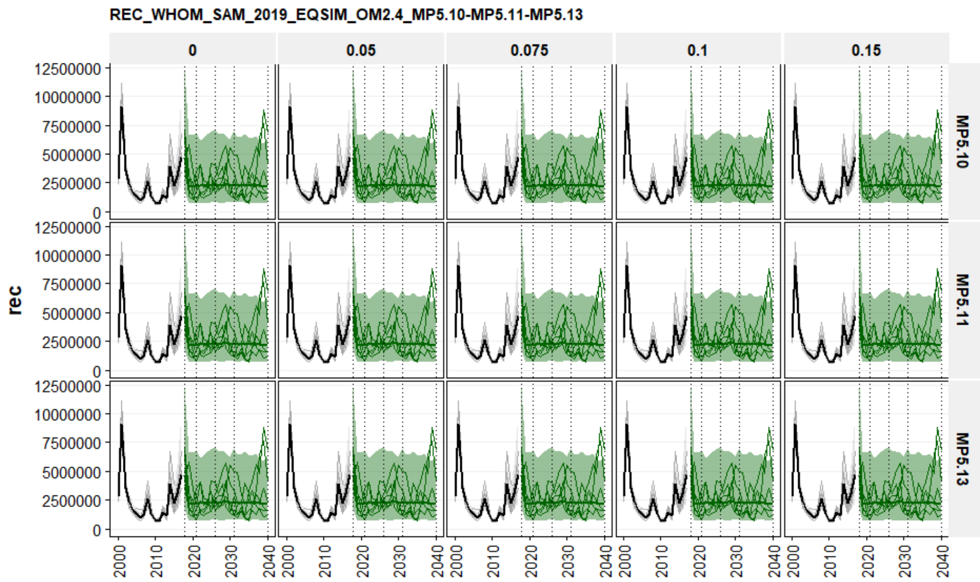
1.2.1 ICES advice rule

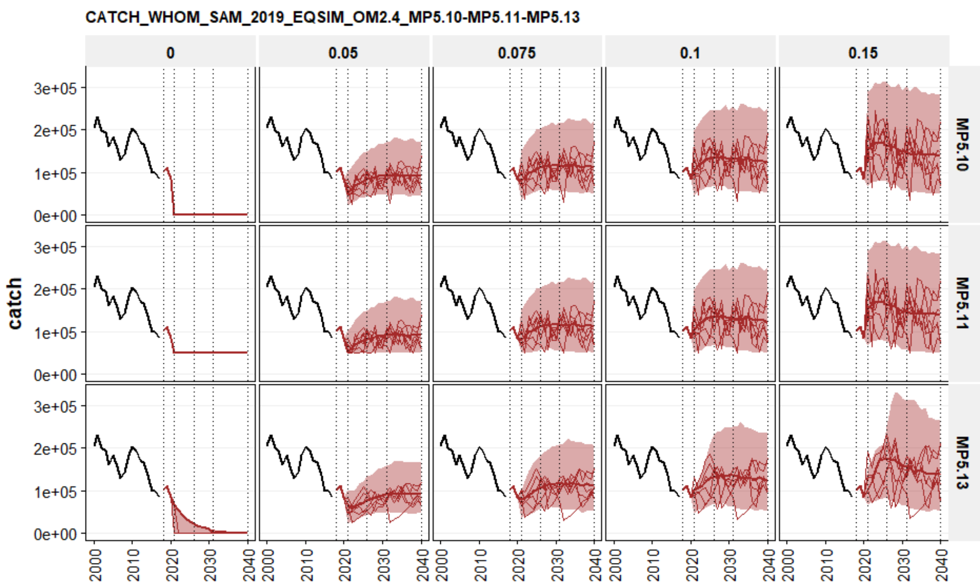
Simulation summary 2019 assessment, SAM, EqSim.

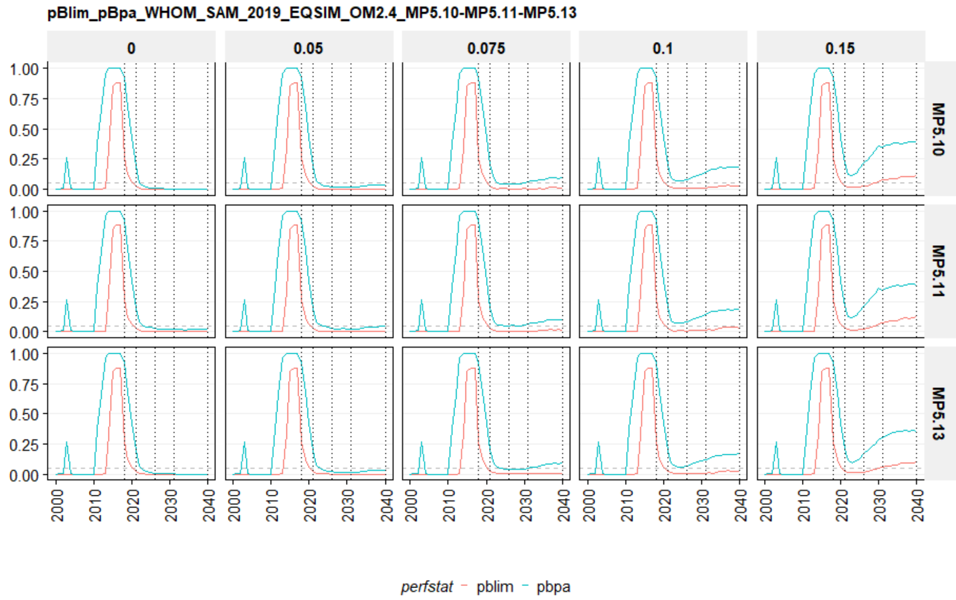
MP5.10 ICES AR; MP5.11 ICES AR with minimum TAC of 50kT; MP5.13 ICES AR with 20% IAV constraint above Btrigger.






REC_WHOM_SAM_2019_EQSIM_OM2.4_MP5.10-MP5.11-MP5.13
0
0.05
0.075
0.1
0.15
MP5.10
MP5.11
MP5.13
rec
12500000
10000000
7500000
5000000
2500000
0
2000
2010
2020
2030
2040


CATCH_WHOM_SAM_2019_EQSIM_OM2.4_MP5.10-MP5.11-MP5.13
0
0.05
0.075
0.1
0.15
MP5.10
MP5.11
MP5.13
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040



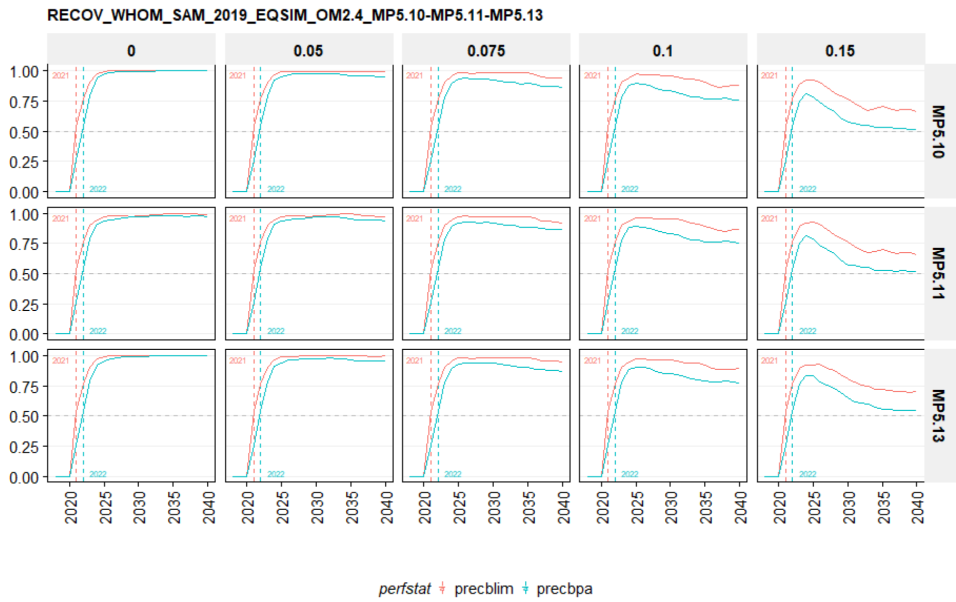
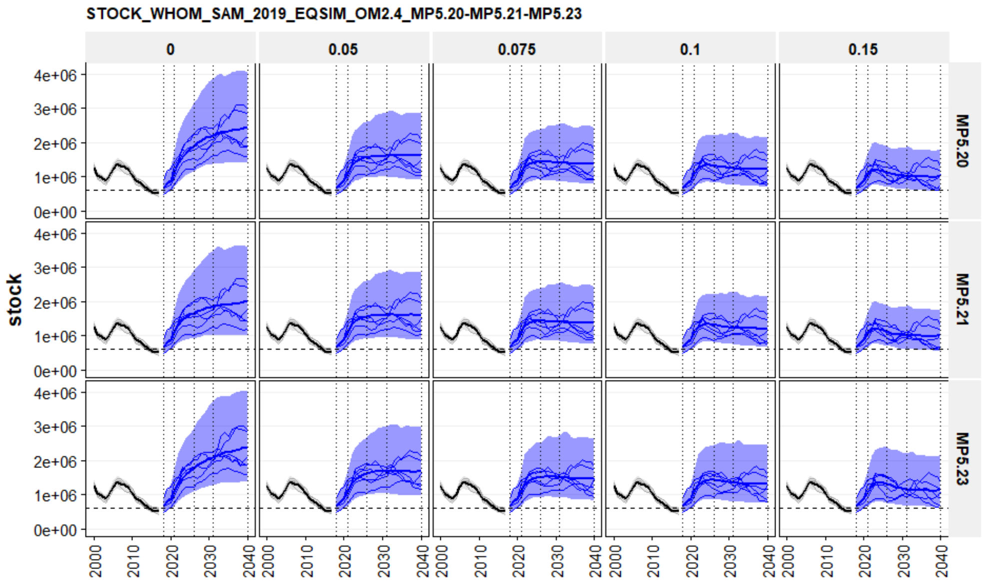


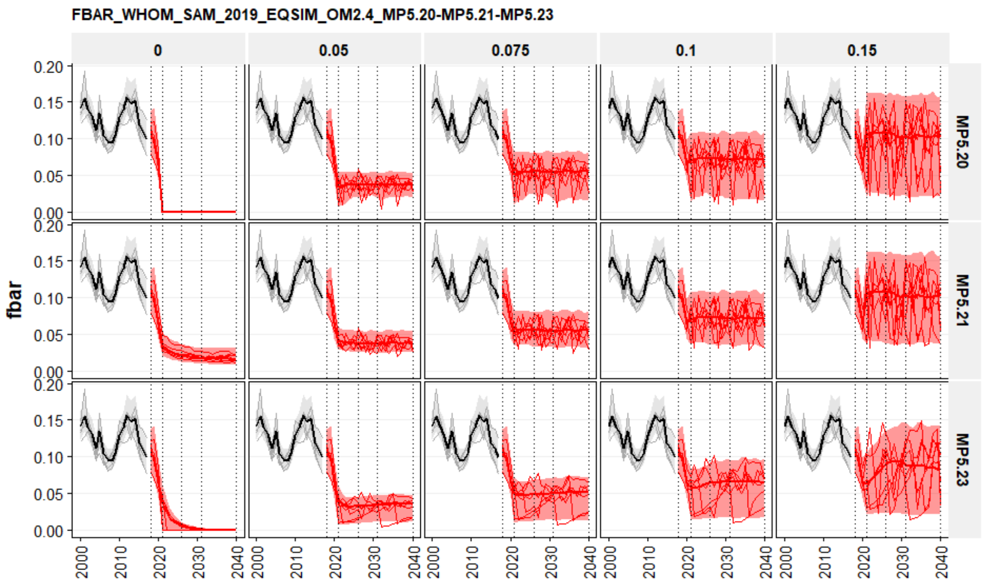
Figure 5 Simulation summary 2019 assessment, SAM, EqSim. MP5.10 ICES AR; MP5.11 ICES AR with minimum TAC of 50kT; MP5.13 ICES AR with 20% IAV constraint above Btrigger.

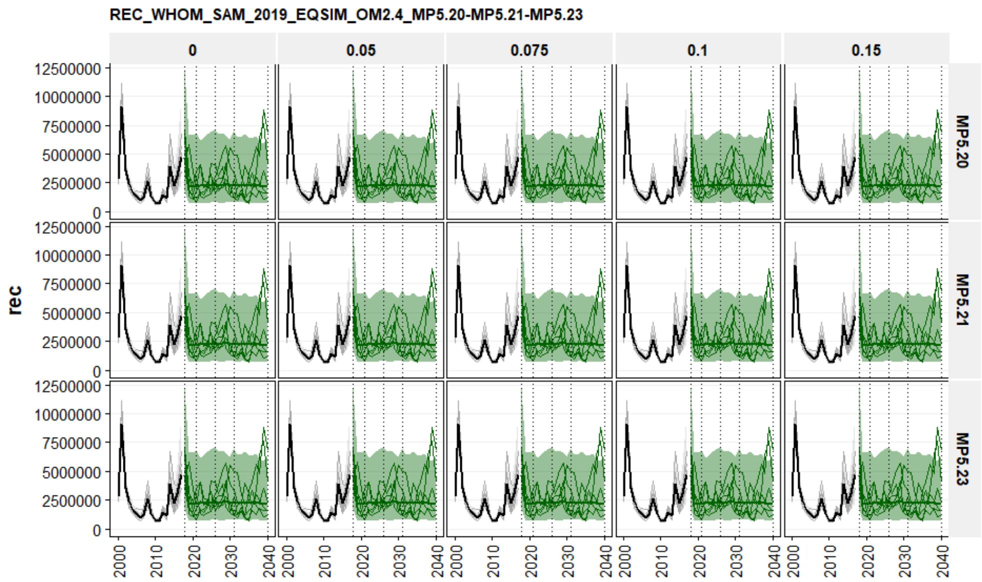
1.2.2 Double breakpoint rule

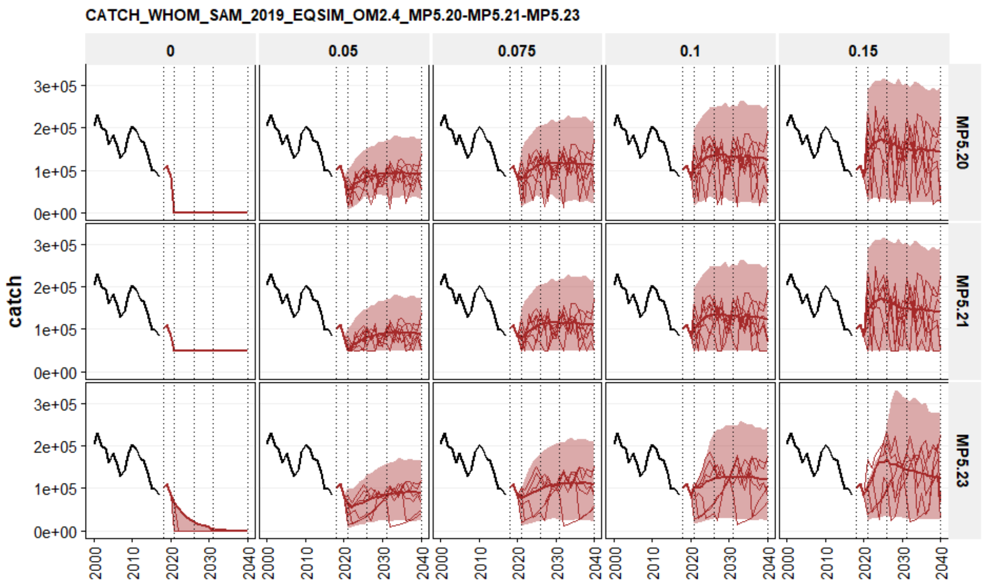
Simulation summary 2019 assessment, SAM, EqSim.

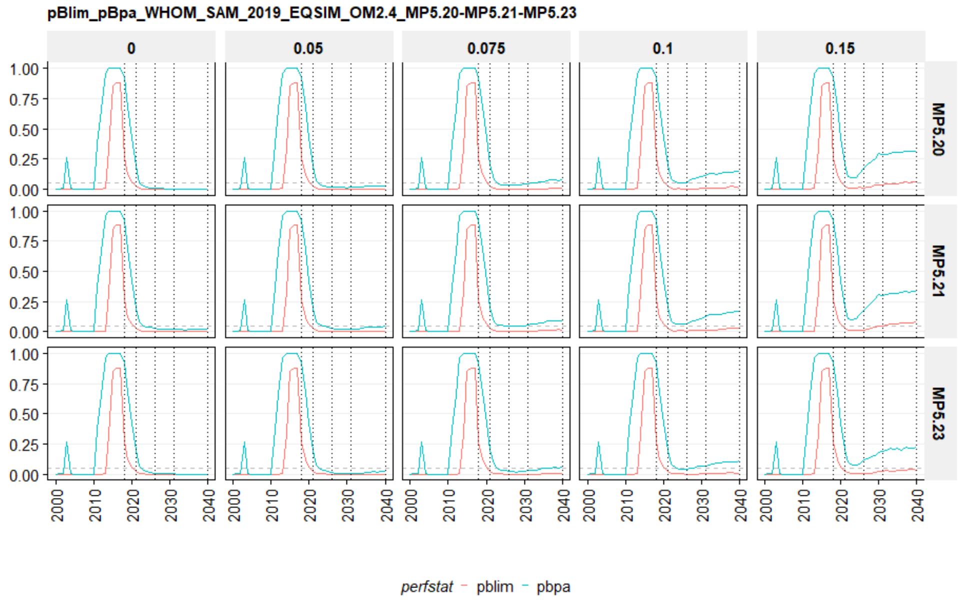
MP5.20 Double BP; MP5.21 Double BP with minimum TAC of 50kT; MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.






REC_WHOM_SAM_2019_EQSIM_OM2.4_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
rec
12500000
10000000
7500000
5000000
2500000
0
2000
2010
2020
2030
2040


CATCH_WHOM_SAM_2019_EQSIM_OM2.4_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040



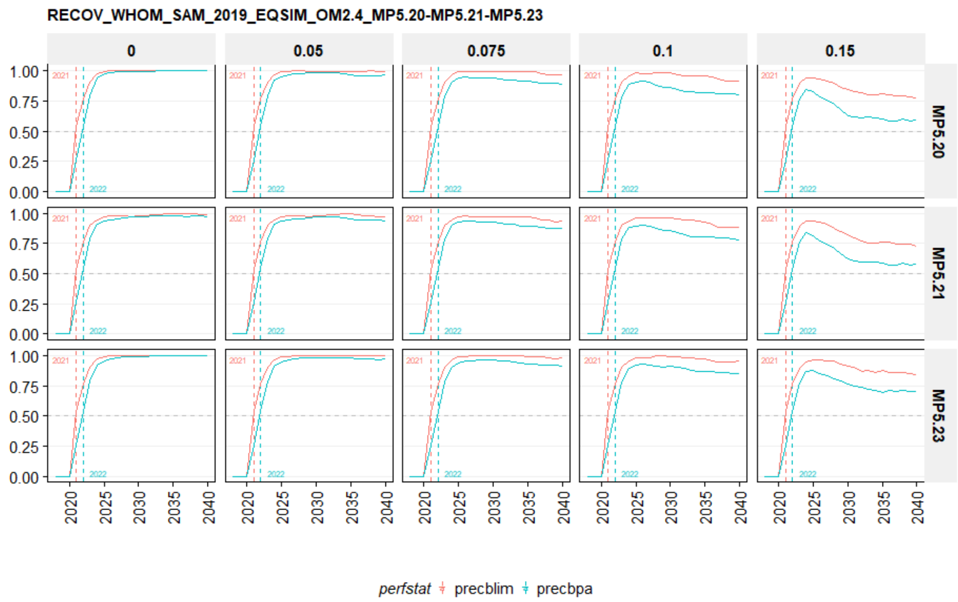
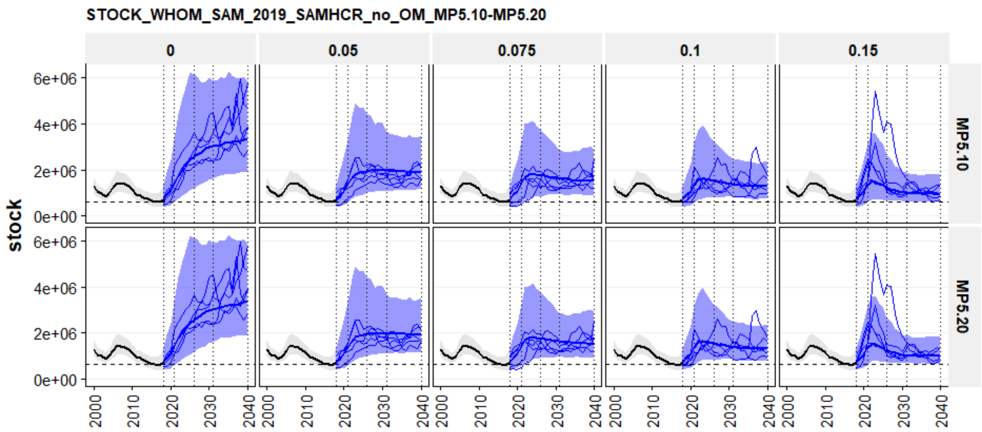


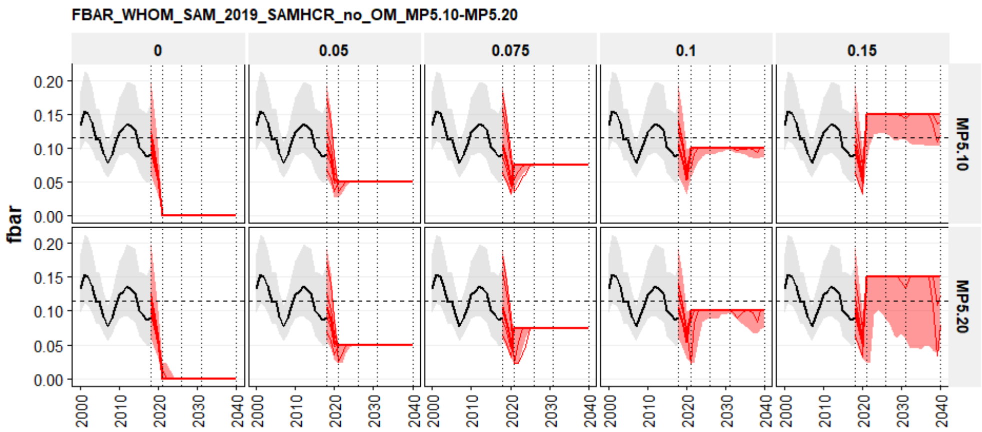
Figure 6 Simulation summary 2019 assessment, SAM, EqSim. MP5.20 Double BP; MP5.21 Double BP with minimum TAC of 50kT; MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

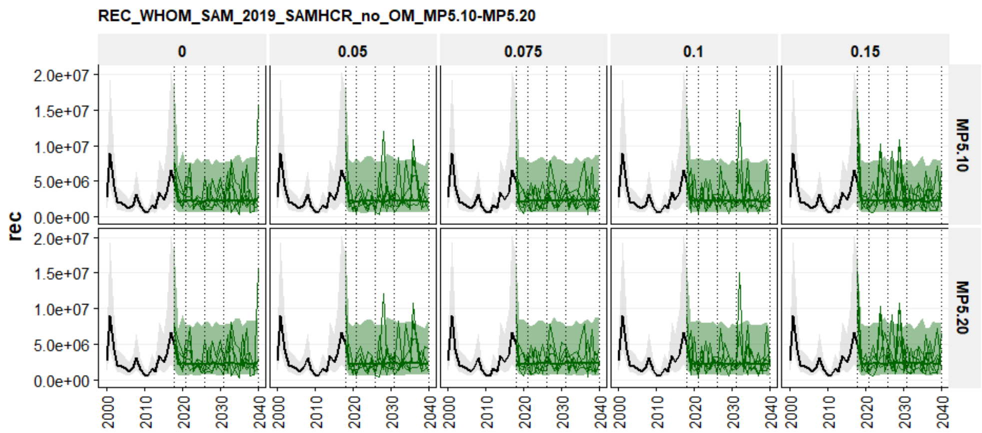
1.3 SAM and SAM HCR forecast

SAM Assessment and SAM HCR forecast.

MP5.10 ICES Advice Rule; MP5.20 Double BP; No IAV of fixed catch options. Minimum F in Double BP is 20% of Fmsy.







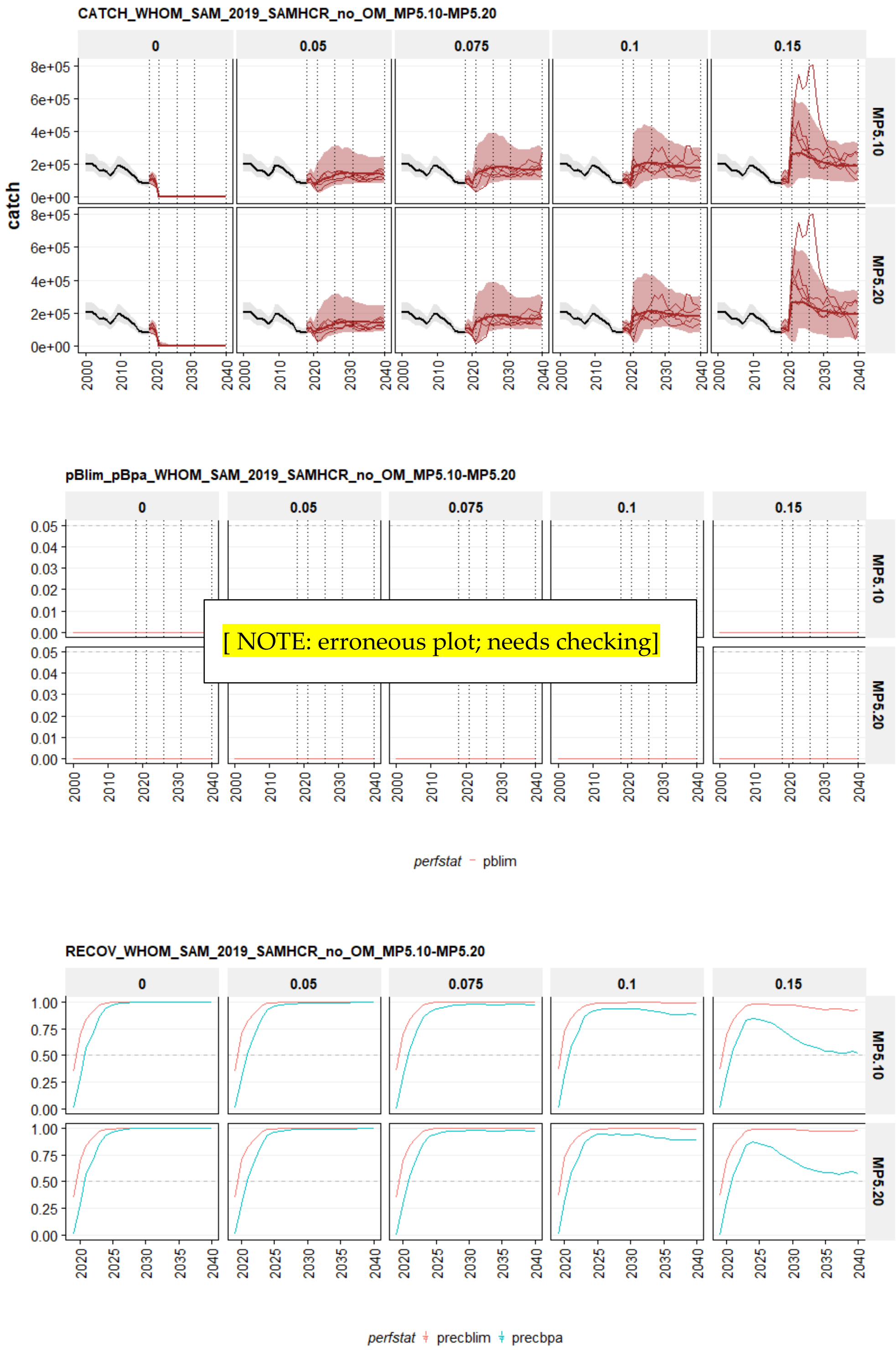


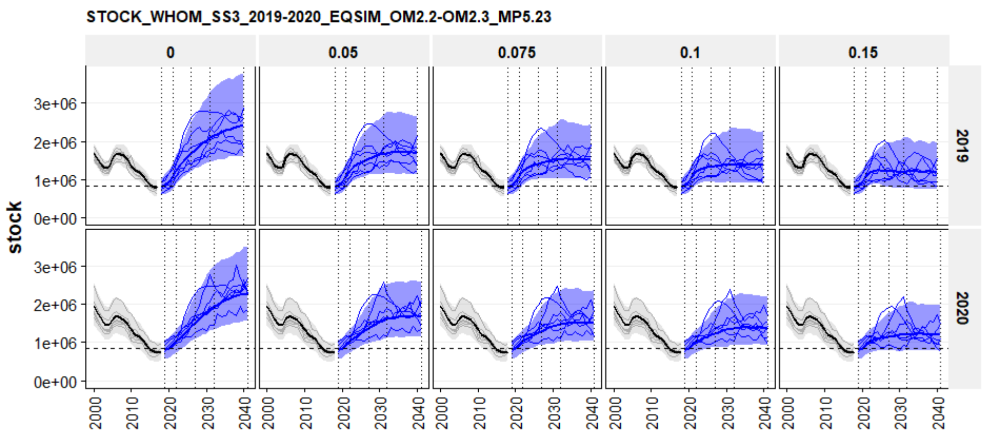
Figure 7 Simulation summary of SAM Assessment and SAM HCR forecast. MP5.10 ICES Advice Rule; MP5.20 Double BP; No IAV of fixed catch options. Minimum F in Double BP is 20% of Fmsy.

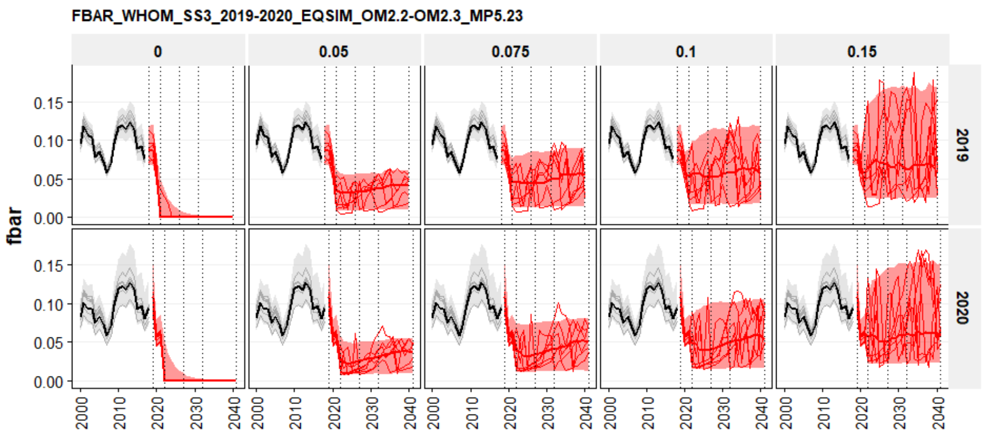
1.4 Sensitivity analysis to assessment year as basis for the simulation

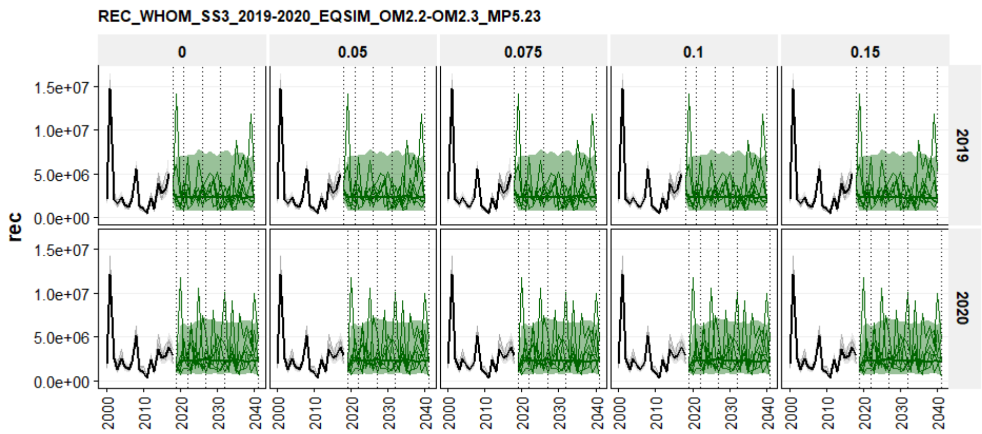
1.4.1 Comparison of 2019 and 2020 SS3 assessment and EqSim simulator

MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

SS3 assessment of 2019 and 2020 using Eqsim simulator







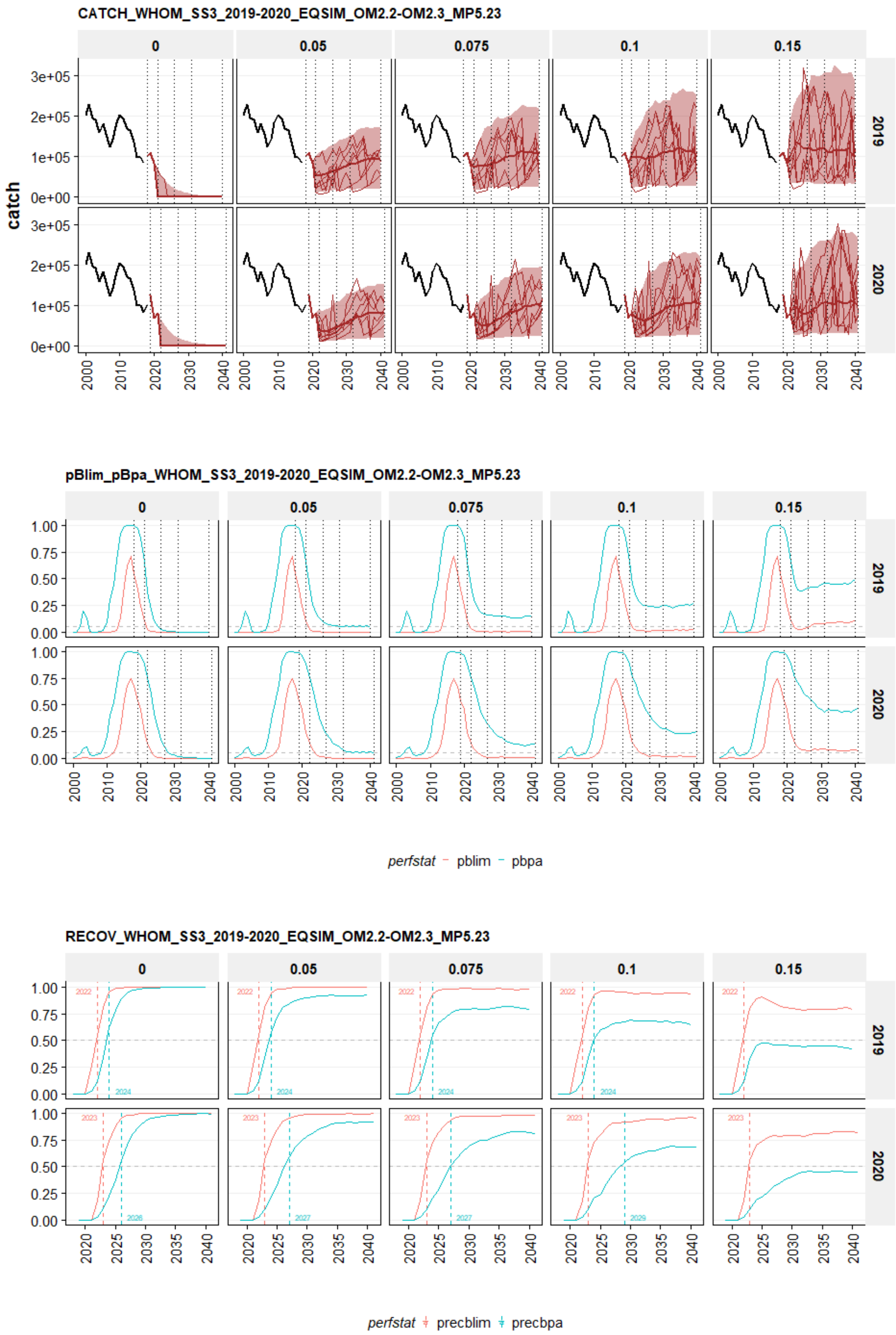
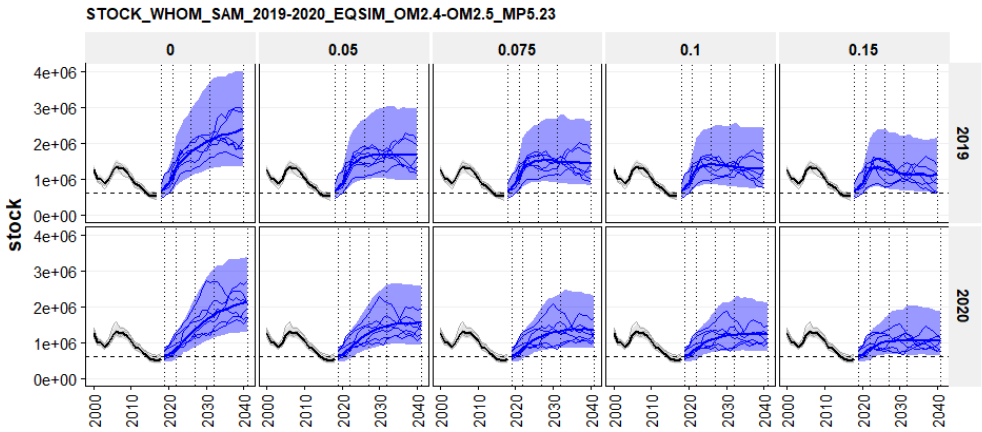


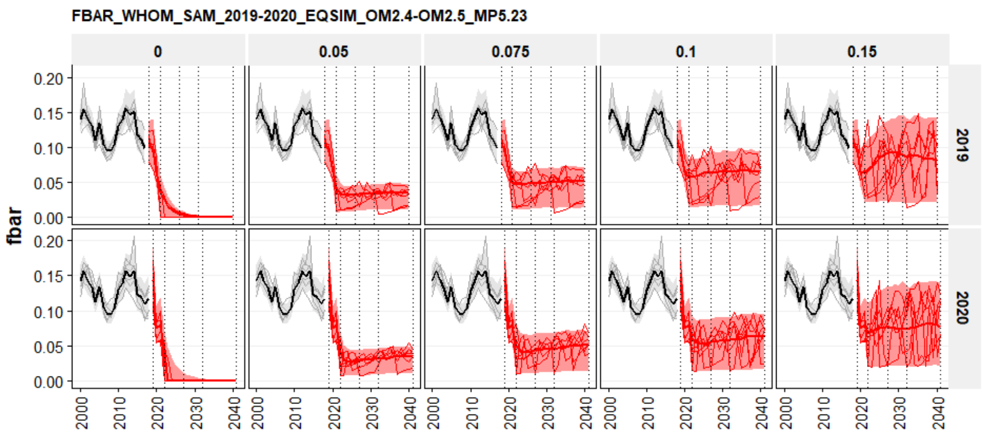
Figure 8 Simulation summary of SS3 Assessment 2019 and 2020 and EqSim simulator. MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

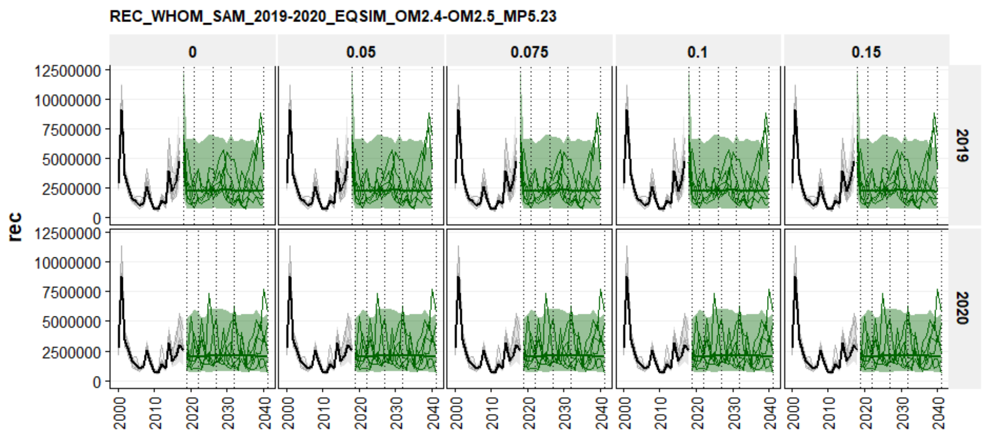
1.4.2 Comparison of 2019 and 2020 SAM assessment and EqSim simulator

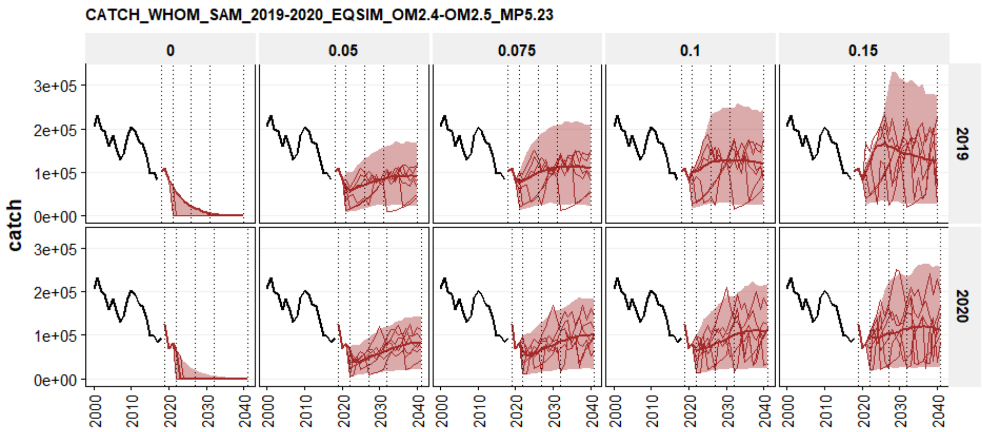
MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

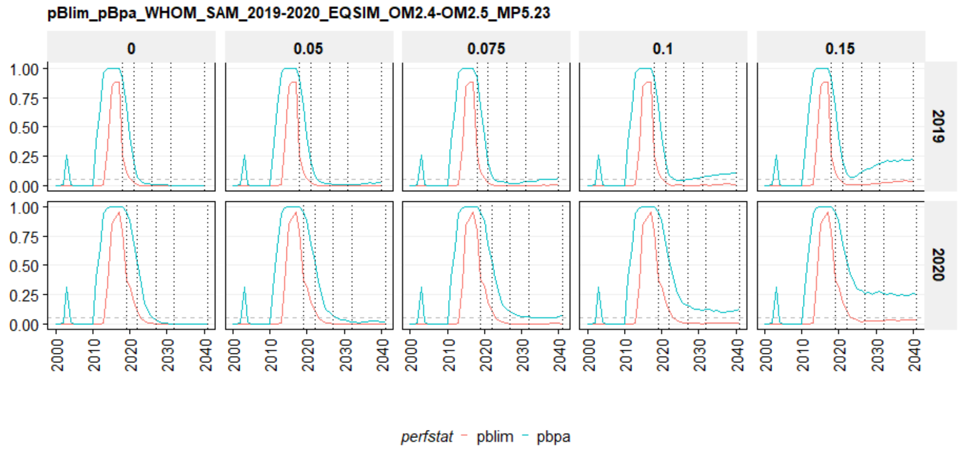
SAM assessment of 2019 and 2020 using Eqsim simulator











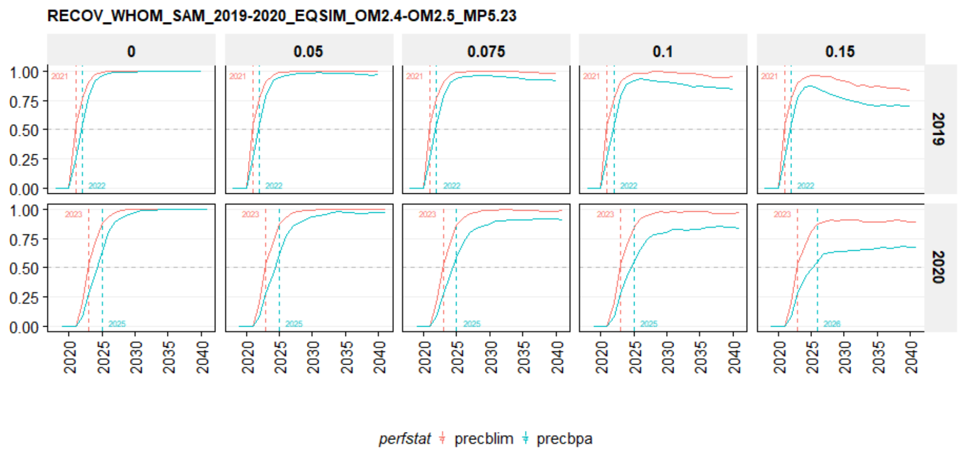


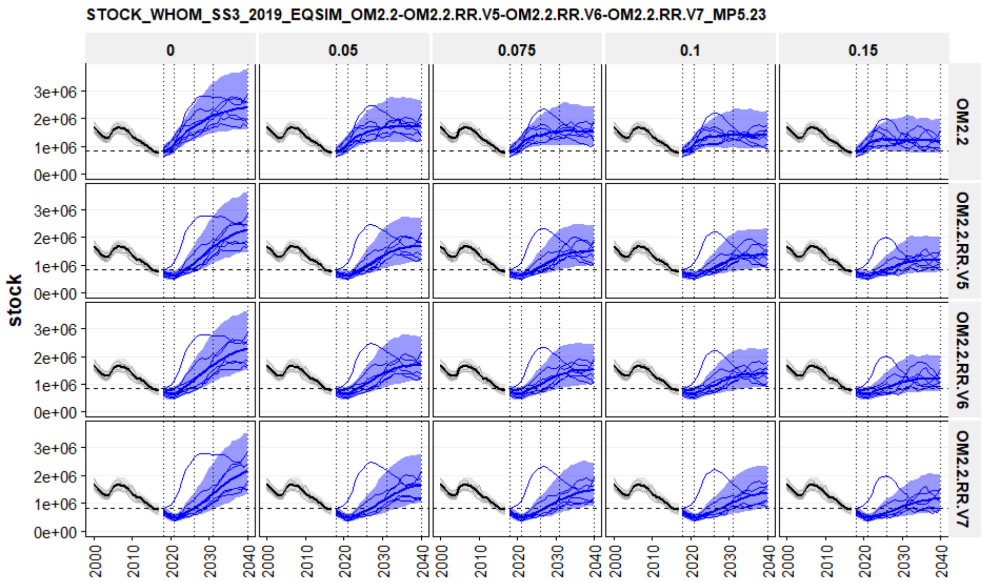
Figure 9 Simulation summary of SAM Assessment 2019 and 2020 and EqSim simulator. MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

1.4.3 Sensitivity of 2019 SS3 with different recruitment assumptions

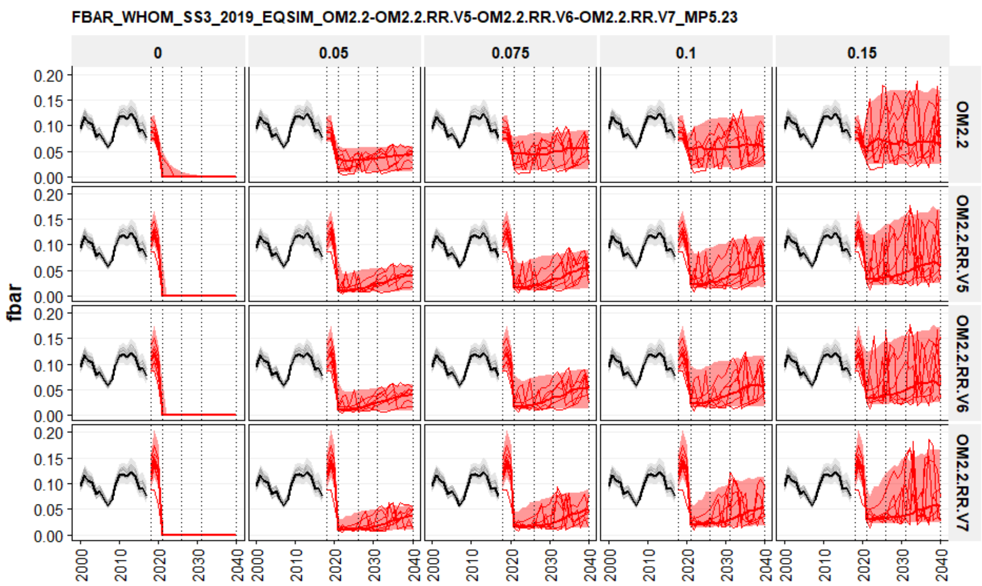
2019 SS3 assessment using Eqsim simulator and different recruitment assumptions. MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

Recruitment assumptions:

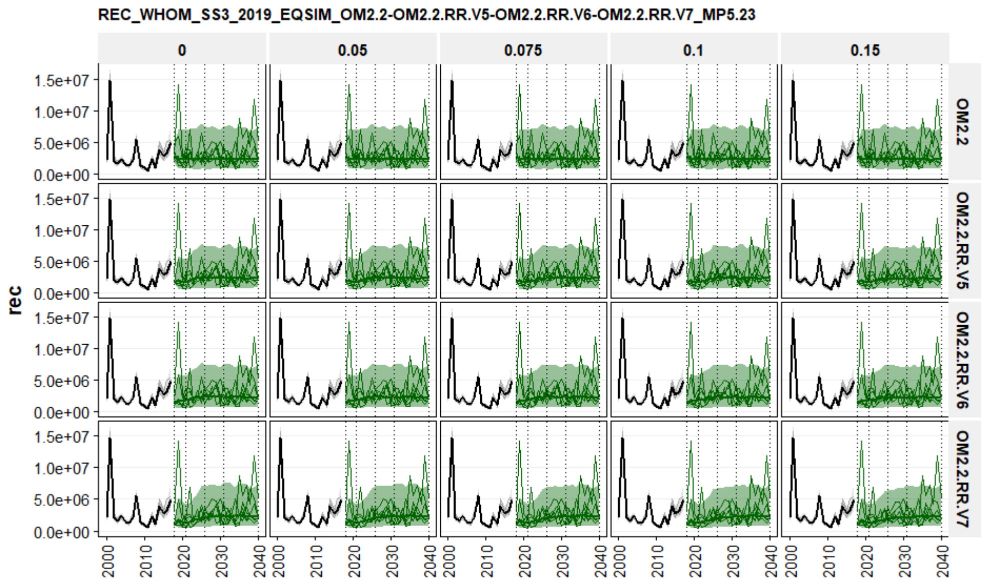
- base case: recruitment from 2019 SS3 assessments. The geo mean of the full time series (ex 1982) is 2,572,649. The mean of the last 5 years from the 2019 assessment is 3,612,642
- RR.V5: 2014-2018 recruitments reduced to the GM of the time series from 2002-2013. Apart from 2008, this represents an extended period of low recruitment. The geo mean for 2002-2013 is 1,620,516
- RR.V6: 2014-2018 recruitments reduced to ½ of the baseline estimate. The mean of the last 5 years is then 1 806 321, similar to scenario RR.V5.
- RR.V7: 2014-2018 recruitments reduced to the mean of the 5 lowest from the full time series. The mean of the 5 lowest is 1,001,051. This is the most pessimistic scenario.

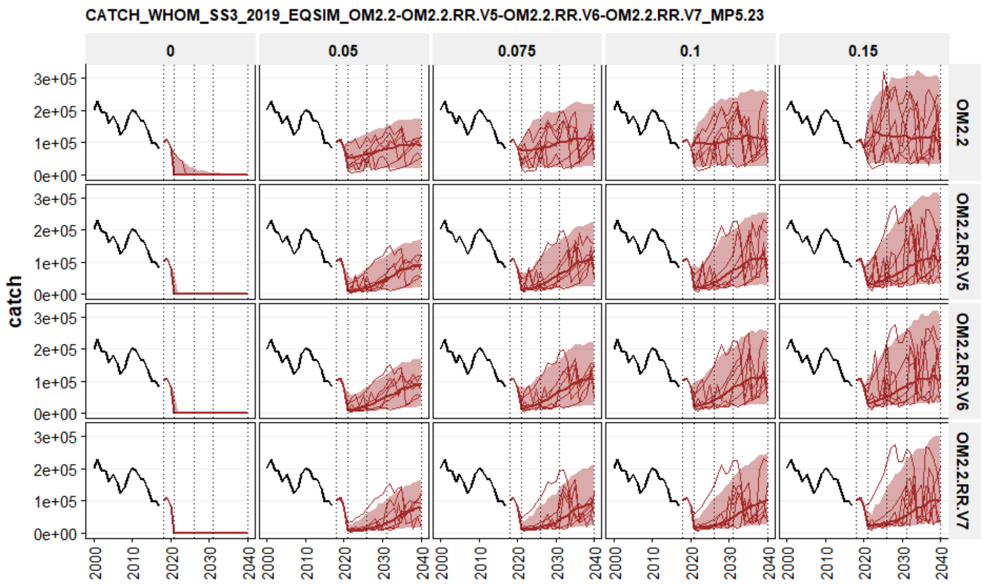


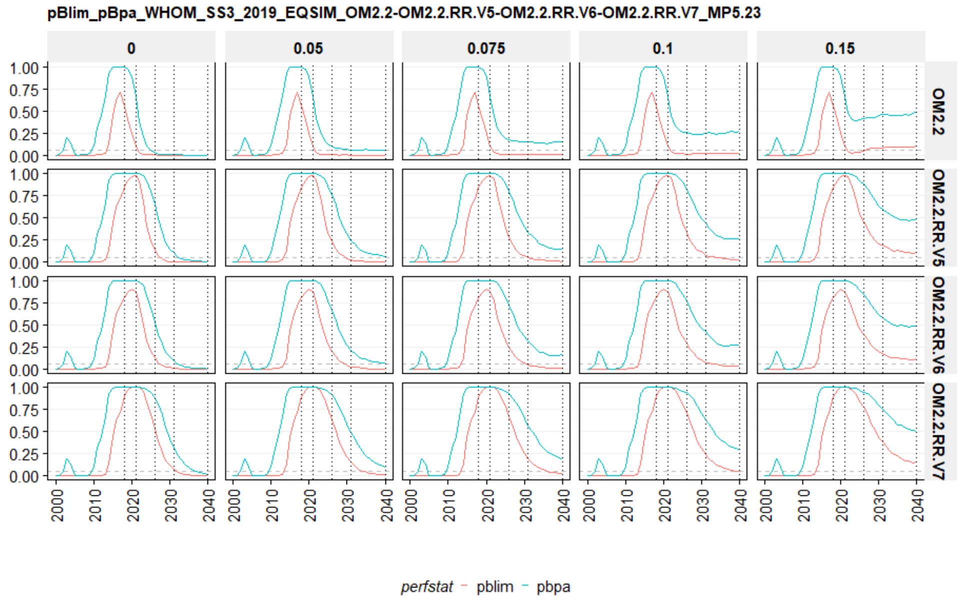
STOCK_WHOM_SS3_2019_EQSIM_OM2.2-OM2.2.RR.V5-OM2.2.RR.V6-OM2.2.RR.V7_MP5.23
0
0.05
0.075
0.1
0.15
OM2.2
OM2.2.RR.V5
OM2.2.RR.V6
OM2.2.RR.V7
stock



FBAR_WHOM_SS3_2019_EQSIM_OM2.2-OM2.2.RR.V5-OM2.2.RR.V6-OM2.2.RR.V7_MP5.23
0
0.05
0.075
0.1
0.15
OM2.2
OM2.2.RR.V5
OM2.2.RR.V6
OM2.2.RR.V7
fbar


REC_WHOM_SS3_2019_EQSIM_OM2.2-OM2.2.RR.V5-OM2.2.RR.V6-OM2.2.RR.V7_MP5.23
0
0.05
0.075
0.1
0.15
OM2.2
OM2.2.RR.V5
OM2.2.RR.V6
OM2.2.RR.V7
rec
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2000
2010
2020
2030
2040


CATCH_WHOM_SS3_2019_EQSIM_OM2.2-OM2.2.RR.V5-OM2.2.RR.V6-OM2.2.RR.V7_MP5.23
0
0.05
0.075
0.1
0.15
OM2.2
OM2.2.RR.V5
OM2.2.RR.V6
OM2.2.RR.V7
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040



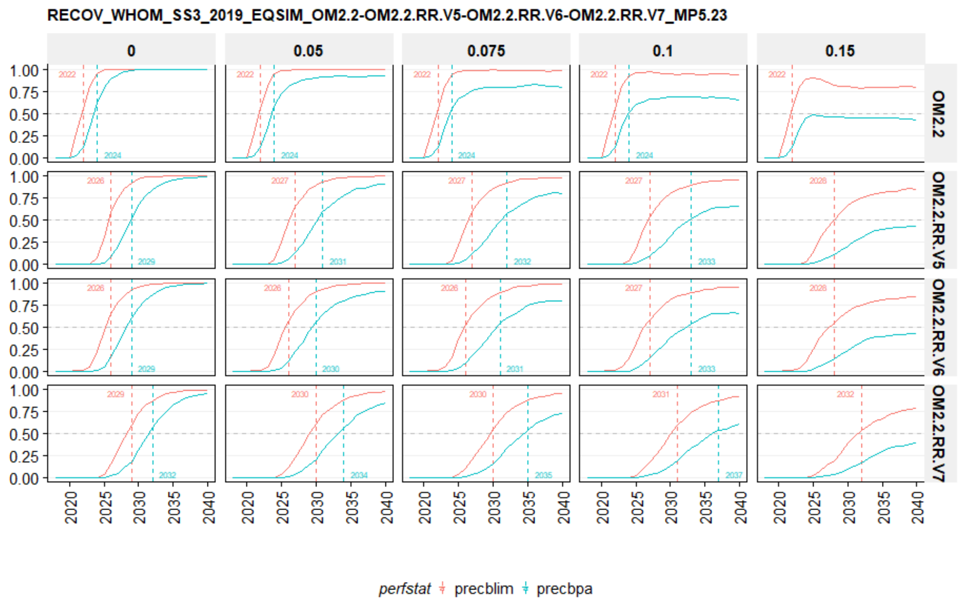


Figure 10 Simulation summary of SS3 Assessment 2019, EqSim simulator and different recruitment as-sumptions. MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

1.4.4 First year rebuild to a biomass reference point under different recruitment assumptions

Continued from section 1.4.3.

perfstat	scenario	recscenario	0	0.025	0.05	0.075	0.1	0.125	0.15
precblim	SAM 2019	def	2021	2021	2021	2021	2021	2021	2021
precblim	SAM 2020	def	2023	2023	2023	2023	2023	2023	2023
precblim	SS3 2019	def	2022	2022	2022	2022	2022	2022	2022
precblim	SS3 2019	RR.V5	2026	2026	2027	2027	2027	2028	2028
precblim	SS3 2019	RR.V6	2026	2026	2026	2026	2027	2027	2028
precblim	SS3 2019	RR.V7	2029	2029	2030	2030	2031	2031	2032
precblim	SS3 2020	def	2023	2023	2023	2023	2023	2023	2023
precbpa	SAM 2019	def	2022	2022	2022	2022	2022	2022	2022
precbpa	SAM 2020	def	2025	2025	2025	2025	2025	2025	2026
precbpa	SS3 2019	def	2024	2024	2024	2024	2024	2025	Inf
precbpa	SS3 2019	RR.V5	2029	2030	2031	2032	2033	2036	Inf
precbpa	SS3 2019	RR.V6	2029	2029	2030	2031	2033	2037	Inf
precbpa	SS3 2019	RR.V7	2032	2033	2034	2035	2037	Inf	Inf
precbpa	SS3 2020	def	2026	2026	2027	2027	2029	2031	Inf

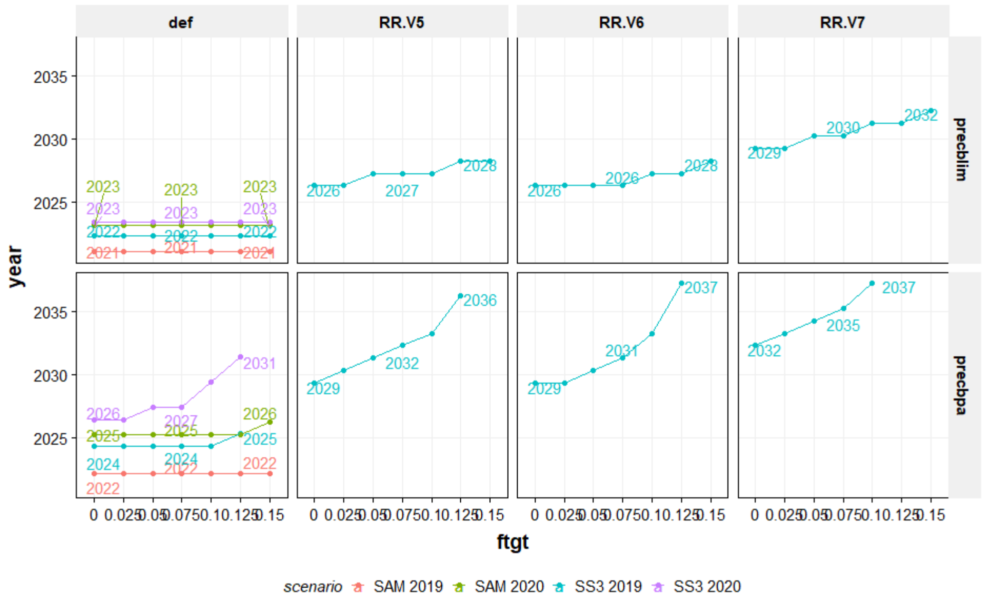


Figure 11 First year to rebuild to a biomass reference point under different recruitment assumptions

1.4.5 First year rebuild to a biomass reference point when ‘wrong’ reference points are used

What happens if SAM reference points are used for SS3 assessment or when SS3 reference points are used for the SAM assessment.

perfstat	scenario	rpscenario	0	0.025	0.05	0.075	0.1	0.125	0.15
precblim	SAM 2019	def	2021	2021	2021	2021	2021	2021	2021
precblim	SAM 2019	WR	2022	2022	2022	2022	2022	2022	2022
precblim	SAM 2020	def	2023	2023	2023	2023	2023	2023	2023
precblim	SAM 2020	WR	2025	2025	2025	2025	2025	2025	2025
precblim	SS3 2019	def	2022	2022	2022	2022	2022	2022	2022
precblim	SS3 2019	WR	2022	2022	2022	2022	2022	2022	2022
precblim	SS3 2020	def	2023	2023	2023	2023	2023	2023	2023
precblim	SS3 2020	WR	2023	2023	2023	2023	2023	2023	2023
precbpa	SAM 2019	def	2022	2022	2022	2022	2022	2022	2022
precbpa	SAM 2019	WR	2024	2024	2024	2024	2024	2024	2024
precbpa	SAM 2020	def	2025	2025	2025	2025	2025	2025	2026
precbpa	SAM 2020	WR	2027	2028	2028	2029	2030	2032	Inf
precbpa	SS3 2019	def	2024	2024	2024	2024	2024	2025	Inf
precbpa	SS3 2019	WR	2022	2022	2022	2022	2022	2022	2022
precbpa	SS3 2020	def	2026	2026	2027	2027	2029	2031	Inf
precbpa	SS3 2020	WR	2023	2023	2023	2023	2023	2023	2023

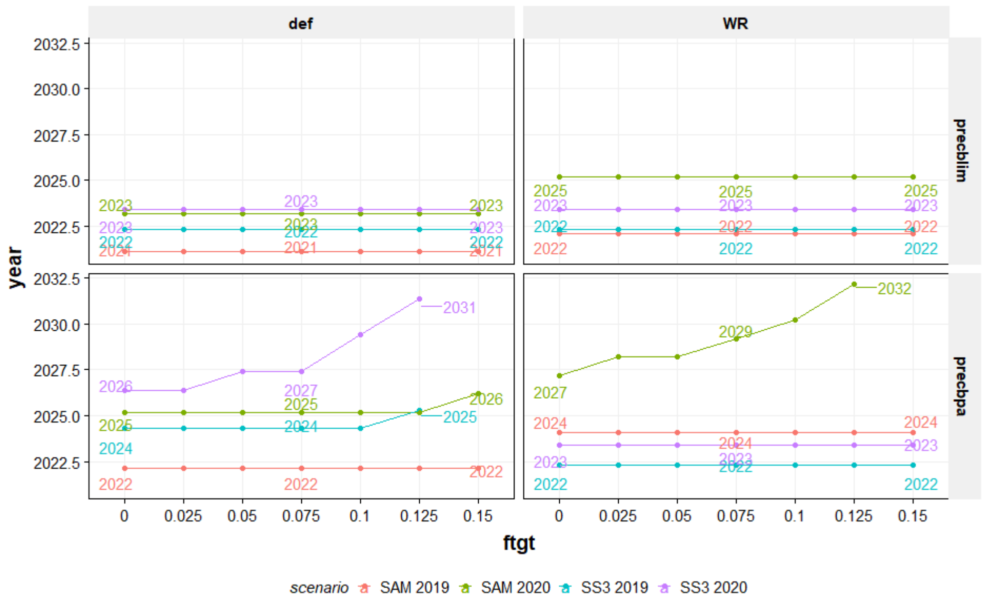


Figure 12 First year to rebuild to a biomass reference point when ‘wrong’ reference points are used

Working document 01, WKWHMRP 2021

Updated Diagnostics and Simulations of the Western horse mackerel Rebuilding Plan evaluations

M.A. Pastoors, A. Campbell, V. Trijoulet, M. Gras, G.I. Lambert, C.R. Sparrevohn, S. Mackinson, Rosana Ourens

6/5/2021

Contents

Figures

Tables

Summary

This working document has been produced by the group of experts that developed the original evaluation of a rebuilding plan of Western Horse mackerel on behalf of the Pelagic Advisory Council (Pastoors et al., 2020). The evaluation was submitted to ICES for a scientific review by WKWHMRP in February 2021. The reviewers and workshop participants concluded that various elements of the paper and rebuilding plan could not be evaluated due to lack of detail. Therefore, it was not possible to determine from the report whether all criteria listed under the ToR of WKWHMRP were met.

The review requested that the report authors provide clarifications on some elements, and address several additional questions. These are addressed within this working document to WKWHMRP. Detailed results of the simulations are available on the WKHMRP sharepoint site.

The additional work consists of the following key elements:

1) **Better documentation on the simulation approaches, the parameterization of the models, the results achieved and the metrics being used (specifically the rebuilding metrics)**

   The expert group involved in this work acknowledges that the documentation of the simulation work on horse mackerel could have been better and more complete in the first place. On reworking the material, we also carried out a number of checks on the simulation code and improved the standard documentation from the model. We have now included the standard WKGMSE table for MSE documentation in section 2.4.

   We also streamlined the code on github[1], separating it into code for processing of input data, code for general settings and loading of relevant functions, code for simulations and code for presentation of the results.

   Presentation of input data has been included in sections 2.2, 2.3 and 2.8. Presentation of the results has been included in the Annex (section 8). In the presentation of the results, we now report all the relevant metrics and we connected the recent past (from 2000 onwards) to the simulated futures. New figures have been included to present the first year of rebuilding to certain biomass reference points under different assumptions or conditions.

2) **Exploration of the sensitivity of the results to the assessment year (2019 or 2020)**

   The exploration to the sensitivity to the assessment year has been explored both for the SS3 and SAM assessments. Results are summarized in section 2.10, 8.4.1 and 8.4.2. Using the 2020 assessment as the OM for the simulations leads to a delay in achieving rebuilding to both Blim and Bpa. For the target F of 0.075 (~Fmsy) with SS3, the rebuilding to Blim would be delayed to from 2022 to 2023 and the rebuilding to Bpa from 2024 to 2027. The main factor contributing to the delay is the lower estimates of recruitment in the most recent years.

3) **Exploration of the sensitivity to the recent (relatively stronger) recruitment**

[1] https://github.com/ices-eg/wk_WKREBUILD/tree/master/EqSimWHM

The exploration of sensitivity to the recent (stronger) recruitment has been explored both for the SS3 and SAM assessments. Results are summarized in sections 2.11, 8.4.3 and 8.4.4. A base case and three alternative recruitment assumptions were tested:

- Base case: use recruitment estimates directly from the assessment(s)
- Annual recruitments for 2014-2018 are equivalent to one half of the baseline estimate (i.e. that from the final assessment)
- Each of the 2014-2018 recruitments are assumed to be equivalent to the geometric mean of the time series from 2002-2013. Aside from 2008, this period is representative of the lowest recruitment observed.
- Each of the 2014-2018 recruitments are assumed to be equivalent to the mean of the 5 lowest recruitments from the full time series.

Under the reduced recruitment scenarios, risk to Blim is increased in the initial year of the simulation as a greater proportion of the replicates are below Blim. The effect of suppressing the recruitment for a period of 5 years leads to an initial further increase in risk in the first years of the simulation, during which a catch constraint applies for all replicates and management procedures. However, risk starts to reduce rapidly within 5 years for the most pessimistic scenario falling below 50% in 2023/24 for the baseline/2 and GM 20-14 scenarios and by 2027 for the scenario with the lowest recruitment. It can be concluded that lower recruitment scenarios are expected to lead to longer rebuilding periods, compared to the base case of the 2019 SS3 assessment. This is also confirmed by the exploration of using the 2020 assessment as the basis for the operating model instead of the 2019 assessment.
Under the preferred scenario by PELAC (Double BP with IAV constraint, target F=0.075), the first year of rebuilding in the base case is in 2022, while under the reduced recruitment scenarios it is between 2026 and 2030. Rebuilding to Bpa is expected around 2024 in the base case can would be delayed to 2030-2034 under the reduced recruitment regimes. When the 2020 assessment is used for the OM, rebuilding to Blim would be delayed to 2023 and to Bpa to 2027.

4) Exploration of the sensitivity to the uncertainty and autocorrelation in advice error

The exploration to the sensitivity to the uncertainty and autocorrelataion in advice error has been explored for the SS3 assessment only. Results are summarized in section 2.9. In the focus group report (Pastoors et al., 2020), a review of the methods to generate cv and autocorrelation in advice error was presented. This concluded that the method of estimating advice error has some potential drawbacks. A similar conclusion has recently been reached in the North Sea herring assessment (HAWG 2021, forthcoming).

To explore the sensitivity of the results to the uncertainty (cv) and autocorrelation (phi), additional sensitivity simulations were conducted assuming “perfect knowledge” such that zero error is assumed in the assessment process and also with an arbitrary high correlated error (cv=0.6, phi=0.6). As expected, the removal of assessment error from the simulation leads to a lowering of the risk to Blim. The inflated and highly correlated error assumption (“High Err”) scenario is associated with the highest risk. For both HCRs, risk reduces over the initial years due to the relatively strong recruitment between 2014 and 2018. In the PELAC selected double breakpointrule, the difference between these scenarios is relatively small. A more discernible difference can be seen with the constant F rule where the use of the default values (with a higher autocorrelation in F error) is associated with an increase in risk to Blim.

5) Exploration of the sensitivity to the reference points used

The exploration to the sensitivity to the reference points used has been explored for both the SS3 and SAM assessments. Results are summarized in sections 2.12 and 8.4.5. The biomass reference points in the SAM assessment are lower than the biomass reference points in the SS3 assessment. This is due to the overall lower SSB and higher F in the SAM assessment compared to SS3. As could be expected, the application of the SS3 reference points to the SAM assessment lead to a substantial longer period of rebuilding that is required (because the bar is set higher) and vice versa for the SS3 assessment with SAM reference points.

A table linking the MSE work in this working document the guidelines for MSE from WKGMSE has been included (section 2.4) and similarly for the guidelines on rebuilding plans from WKREBUILD has been included (section 6.2).

1 Introduction

1.1 Background to this working document

This working document has been produced by the group of experts that developed the original evaluation of a rebuilding plan of Western Horse mackerel on behalf of the Pelagic Advisory Council (Pastoors et al., 2020). The evaluation was submitted to ICES for a scientific review and on the basis of that review a number of additional questions for clarification were addressed to the group of experts. These questions for clarification are addressed within this working document.

1.2 Initial review by WKWHMRP

The ToR of WKWHMRP 2021 directed the reviewers and workshop participants to consider a specific report and proposed rebuilding plan. Based on the ToR and on additional discussion with ICES secretariat, it was the understanding of the review panel that it was not under the remit of WKWHMRP, as initially setup, to undertake any major additional analysis.

A conclusion of the reviewers and the workshop was that various elements of the paper and rebuilding plan could not be evaluated due to lack of detail. Therefore, it was not possible to determine from the report under review if all criteria listed under the ToR were met.

Given the main response of the reviewers (Annex 1) in addition to the questions and discussion during the 5 February 2021 meeting it was concluded that additional work to support evaluation of the report and plan against the ToR was warranted. The additional work related not so much to the methodological approach used for the evaluation as a similar approach using the same basic software has been recently used in, for example, the evaluation of Blue whiting (ICES, 2016).

The list of additional analysis and diagnostics that would allow for sufficient evaluation of the proposed plan (ToR Cii) generated by the WKWHMRP at large were identified as being:

1. In general, the documentation of the setup of the main biological input parameters need to adhere more closely to the WKGMSE guidelines (ICES 2013, 2019 and 2021), include justification of the choices made and presentation of diagnostics. Provide comments on the form of rebuilding plan in relation to suggestions by WKREBUILD.
2. Provide descriptions and comparison of major model settings used (both for SS3 and SAM) as well as all a plot of all key output metrics (catch, recruitment, fishing mortality and stock trajectory).
3. Provide plots of all relevant biological and fisheries variables, both in history and in simulations (e.g., weight-at-age, maturity-at-age, gear selectivity, stock and recruitment relationship). Present the probability distribution of future recruits generated in relation to that observed in the past. Specifically, assess how extreme/strong recruitments are handled; determine how often strong recruitments occur for the simulated stock.
4. Provide descriptions of all assumptions for the forward projections including rational. In the case of SAM, specify the process errors that are carried forward in projections and provide estimates of their variance (standard deviation).
5. Provide detailed description of how the assessment error model is applied in the simulations.

6. Add material (figures, tables) that illustrates that the properties of projected biomass are consistent with historical estimates as per ICES guidance on simulations using the short-cut method.
7. Show the results of the actual rebuilding criteria suggested in the plan: minimum of 3 years above the biomass threshold (initially Blim, ultimately Bpa).
8. Describe how the SS3 and SAM implementations parameterize initial conditions in a manner that overcomes the lack of uncertainty in initial conditions in the analysis by Cox et al. (2018).
9. Elaborate on the approach of using historical advice error rather than assessment error based on the most recent assessment. Include robustness testing of the base-case scenario conclusions on the rebuilding plan to the values used the advice error model (Fcv and Fphi). i.e. evaluate the sensitivity of results to the uncertainties used (Fcv and Fphi); e.g. explore what would happen if default values of Fcv and Fphi are used instead of the estimated values based on only 8 years.
10. Check if results are sensitive to the assessment used as the baseline. E.g. use 2020 assessment as baseline vs older version used in current simulations.
11. Evaluate the robustness of results to the perceived strong recruitments in the recent years. For example, is rebuilding performance acceptable if recent recruitments are set to recent average values instead of the current estimates?
12. Evaluate the sensitivity of results to alternate interpretations of reference points, e.g., use reference points estimated with SS3 and apply to SAM simulation and vice versa.
13. Explain the importance of stock structure information in relation to the expected performance of the candidate management procedures. Describe steps that could be taken to reduce the effects of scale mismatch between stock distribution and stock assessment data collection.
14. Provide the equations that indicate how each performance statistic is calculated, and add results for performance criteria that were described but not reported (e.g., criteria for achieving rebuild status).

2 Point-by-point addressing requests from WKWHMRP

2.1 In general, the documentation of the setup of the main biological input parameters need to adhere more closely to the WKGMSE guidelines (ICES 2013, 2019 and 2021), including justification of the choices made and presentation of diagnostics. Provide comments on the form of rebuilding plan in relation to suggestions by WKREBUILD.

See sections 3.1, 3.2 and 4.1 for the documentation of the input parameters while adhering to the WKGMSE guidelines. Details on the setup and basic processes are also included in sections 2.2, 2.3 and 2.8.

The WKGMSE table of MSE specification has been included in section 2.4.

Comments on the form of rebuilding plan (and rebuilding criteria) in relation to suggestions by WKREBUILD are in the discussion section (section 6.2).

2.2 Provide descriptions and comparison of major model settings used (both for SS3 and SAM) as well as all a plot of all key output metrics (catch, recruitment, fishing mortality and stock trajectory).

2.2.1 SS3 – WGWIDE 2019

An implementation of the integrated Stock Synthesis (SS3) assessment model has been used by ICES for stock assessment and the provision of catch advice for Western Horse Mackerel since 2017, following a benchmarking exercise (ICES, 2017). SS3 was preferred to the previous SAD (linked Separable-ADAPT VPA) model which had been specifically developed for this stock, utilising catch at age information and an index of production (egg count) from a triennial survey. The SAD assessment suffered from retrospective issues with the assessment output rescaling upon the introduction of a new egg survey data point every three years. A major advantage of the SS3 implementation lay in its ability to more appropriately incorporate additional fishery-independent data sources available annually.

The SS3 assessment is configured as a single area, single season and single fleet (all countries and gears combined) model. 3 surveys are used; an egg count from the triennial Mackerel and Horse Mackerel Egg Survey(MEGS), an index of recruitment from a delta-lognormal fit to catch rates of 0-group fish from a number of groundfish surveys conducted as part of the International Bottom Trawl Survey (IBTS) in the Bay of Biscay, Celtic Sea and West of Ireland and Scotland and an acoustic index of abundance and length composition data from the Spanish PELACUS survey in ICES division 8c (which covers only a small proportion of the overall stock distribution). Total landed catch from 1982, age disaggregated for years up to 2002 and as conditional age at length from 2003 onwards and length composition from 2000 to present are derived from the sampling of the national commercial fisheries. The complete range of data used in the 2019 stock assessment (ICES 2020) is given in Figure 2.1.

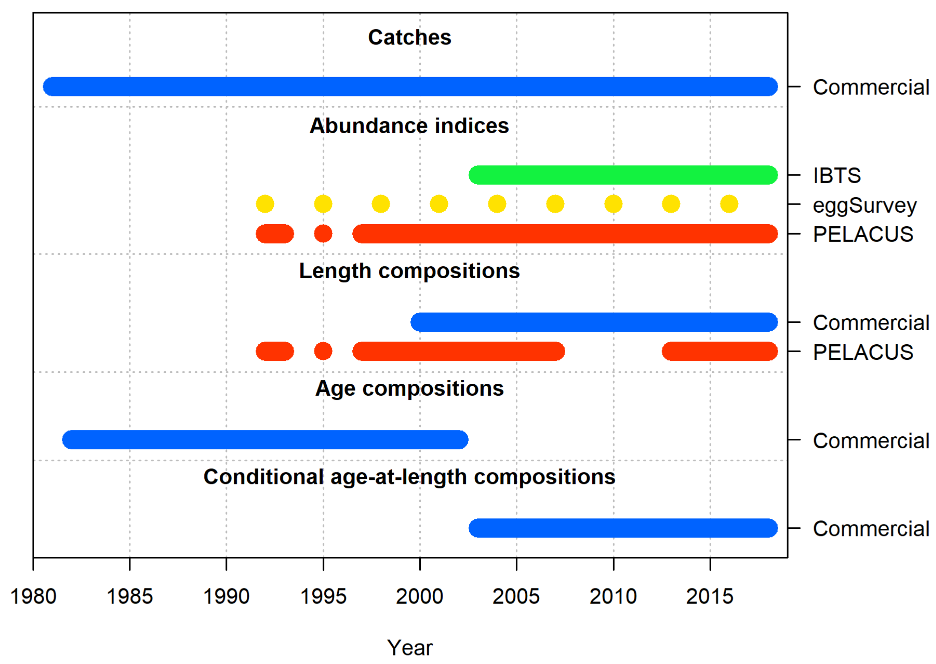


Figure 2.1 WGWIDE 2019 Stock Synthesis Western Horse Mackerel Assessment input data

Catch at age data for Western Horse Mackerel (Figure 2.2) are characterised by infrequent strong year classes that typically persist in the for several years, often contributing significantly to the plus group (which was extended from 12+ to 15+ following the benchmark in 2017). Annual catch at age estimates derived from the sampling of commercial catch by countries with significant quota share are shown in figure 2.2.

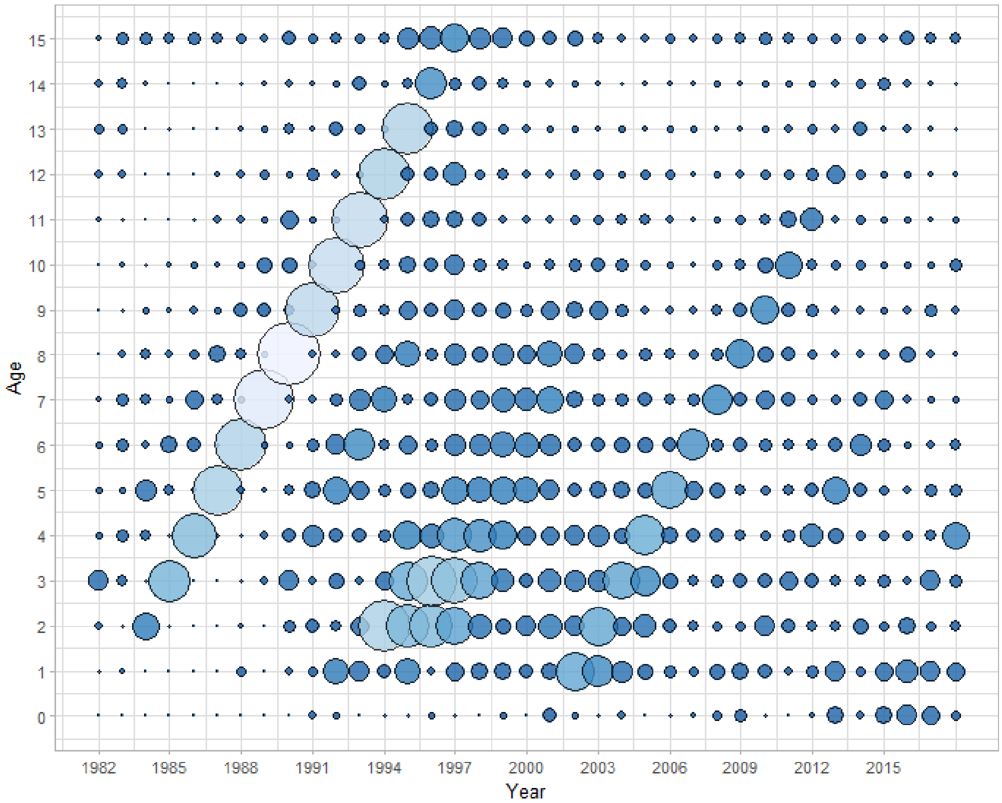


Figure 2.2 Catch at age (numbers- area of bubble proportional to catch number). 15 is a plus group.

The length compositions from the commercial landings are shown in Figure 2.3. Limited information is available on the composition of the discards from some fleets and this is not included in the current assessment configuration. Overall levels of discarding are considered to be negligible.

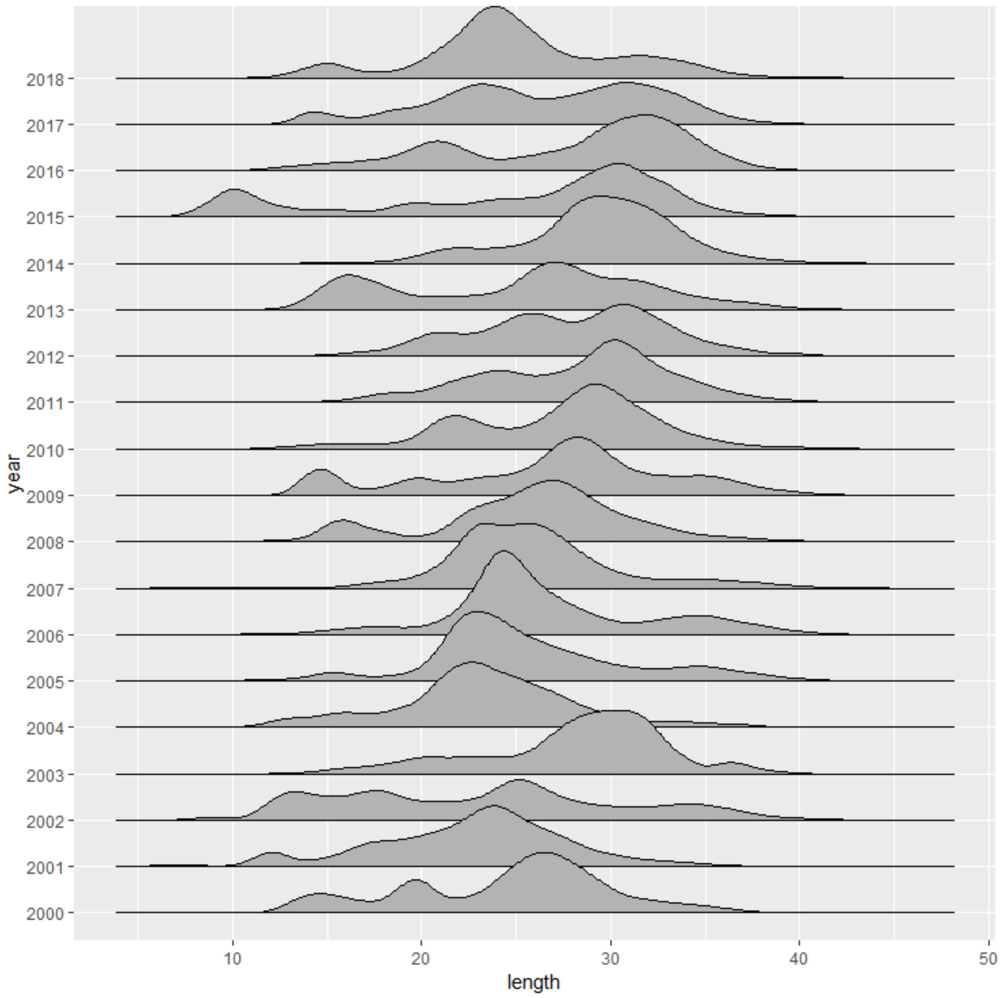


Figure 2.3 Length frequency distribution of the catch.

The time series for each of the three fishery independent indices, along with annual estimates of uncertainty are shown in Figure 2.4.

Figure 2.4 Survey indices as used by the assessment in 2019. IBTS (top) is an index of recruitment derived from annual IBTS surveys, MEGS is the triennial egg survey conducted from Feb-Jul and PELACUS is the biomass index from the Spanish acoustic survey conducted in division 8c.

The main model assumptions and settings are summarised in Table 2.1.

SETTING	VALUE/DESCRIPTION
Starting-Ending Year	1982-2019
Number Areas/Seasons/Fishing Fleets	1/1/1
Surveys	3 (Egg Survey, IBTS, PELACUS acoustic)
Individual Growth	Von Bertalanffy Parameters (initial values) L_at_Amin (5), Linf (40), k (0.205)

Number parameters	71
Length Bins	2-50 (1cm intervals)
Maturity	Time-invariant; 0% ages 0,1, 5% age 2, 25% age 3, 70% age 4, 95% age 5, 100% ages 6 and over
Selectivity	Commercial - Double normal, length-based IBTS – recruitment Egg survey – SSB PELACUS – double normal, length-based
Natural Mortality	0.15
Beverton-Holt Steepness	0.999
LW parameters	0.00000585, 3.087

Table 2.1 Western Horse Mackerel SS3 stock assessment key model assumptions and parameters.

Additional information on the configuration of the assessment can be found in the benchmark workshop report (ICES, 2017) and the stock annex (ICES, 2020a)

The output of the 2019 final WGWIDE assessment is shown in Figure 2.5. This assessment is consistent with that from the previous year with model estimates and residual patterns similar to those presented at the benchmark when this present assessment model was adopted.

Following a peak in 1988 as a result of the very strong 1982 year-class, SSB declined until 2003 before recovering following the moderate 2001 year-class. From 2001-2013 recruitment was generally weak with the exception of 2008. Recruitments since 2014 are estimated to be above the geometric mean (excluding 1982). Fishing mortality has been reducing in recent years as a result of a reduction in catch.

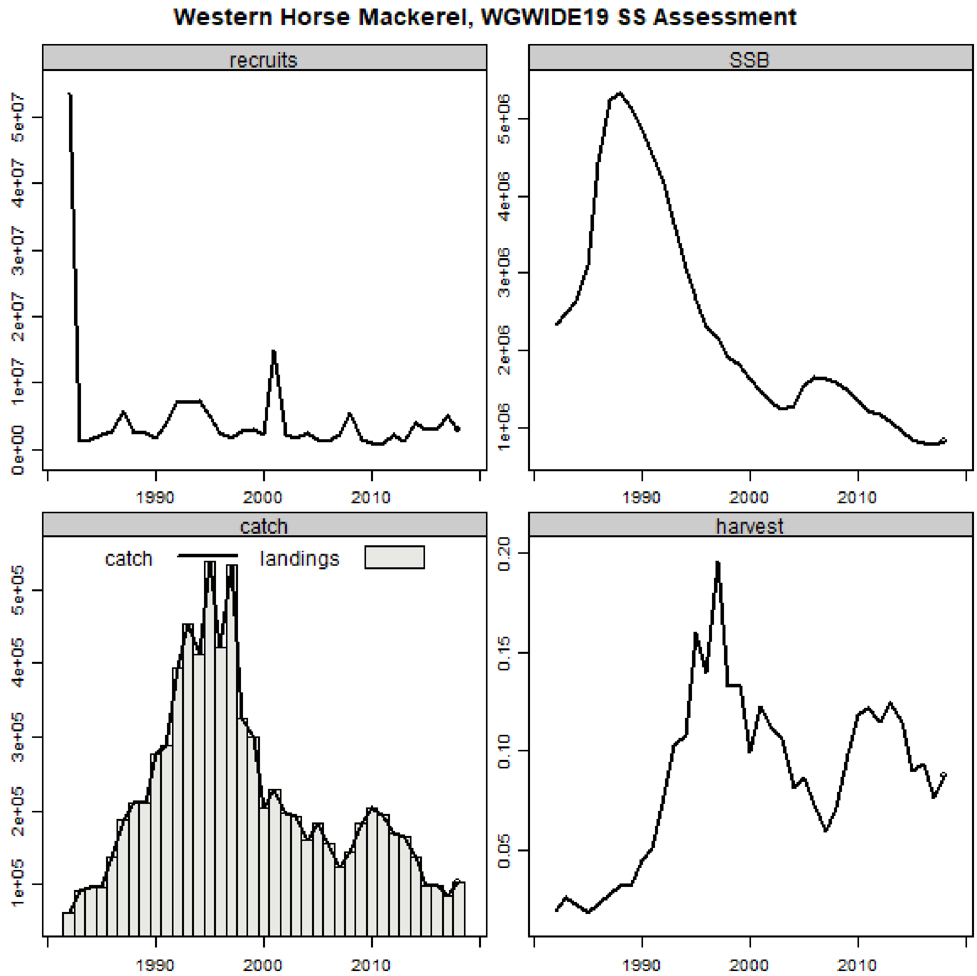


Figure 2.5 WGWIDE 2019 Western Horse Mackerel final assessment stock summary

Despite the introduction of additional (fishery independent) information with the adoption of the stock synthesis model, the assessment continues to suffer from a degree of rescaling, particularly for SSB and fishing mortality although at a reduced level to that typically seen with the previous SAD implementation. The historical results (2015,2016 SAD, 2017-2019 SS3 assessments) are shown in figure 2.6. Note that SSB is estimated on 1 January with the SS3 assessments (2017-2019). The previous assessment model calculated SSB in May.)

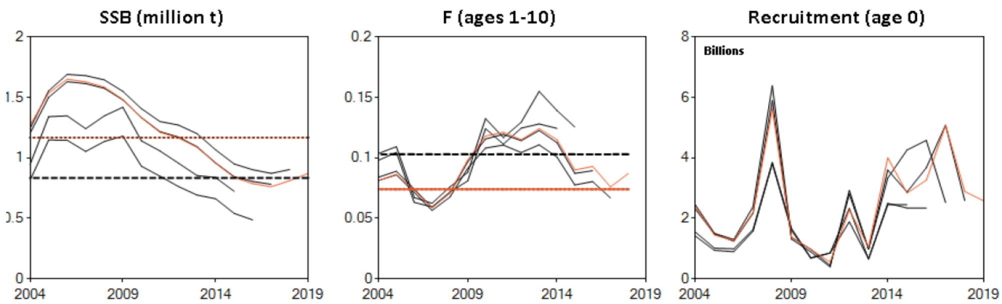


Figure 2.6 Western Horse Mackerel historic assessment results (2015-2019 assessments)

During the 2017 benchmark, simulations were conducted to derive updated MSY reference points (ICES, 2017). The precautionary biomass limit reference point B_{lim} was set to the lowest observed biomass, given no clear indication of impaired recruitment at any observed biomass above this. As outlined above, subsequent assessments led to revisions in the absolute value and timing of the assessment B_{loss} such that the annual changes in catch advice were at odds with the most recent information on stock development.

The reference points were therefore re-examined in 2019 (ICES, 2019). An alternative basis for B_{lim}, robust to the assessment output, was proposed with the lowest SSB estimate from the stable part of the assessment output (2003) considered a suitable proxy for B_{pa} with the corresponding $B_{lim} = B_{pa}/1.4$.

2.2.2 SS3 – WGWIDE 2020

The update assessment conducted at WGWIDE in 2020 resulted in a similar perception of the stock to that in WGWIDE 2019 although the SSB was revised downwards slightly with a corresponding increase in fishing mortality The historic assessments (2016-2020) are shown in figure 2.7.

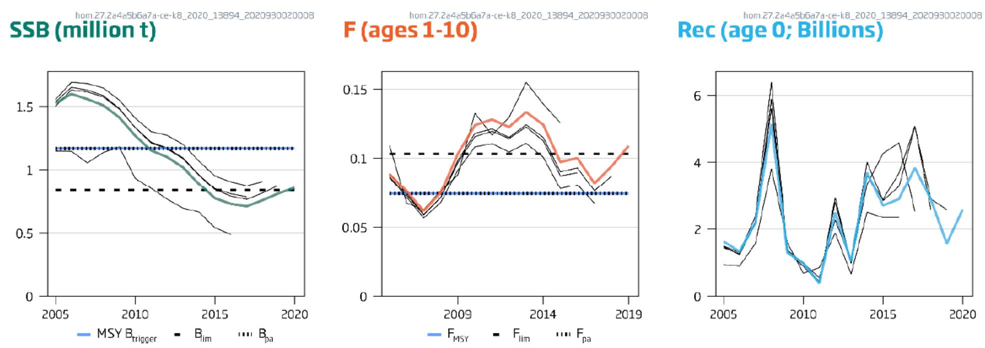


Figure 2.7 Western Horse Mackerel historic assessment results (2016 -2020)

2.2.3 SAM assessment 2019

To be done; see IBPWHM 2019 (ICES, 2019)

2.2.4 SAM assessment 2020

To be done

2.3 Provide plots of all relevant biological and fisheries variables, both in history and in simulations (e.g., weight-at-age, maturity-at-age, gear selectivity, stock and recruitment relationship). Present the probability distribution of future recruits generated in relation to that observed in the past. Specifically, assess how extreme/strong recruitments are handled; determine how often strong recruitments occur for the simulated stock.

Recruitment Modelling

The time series of recruitment and SSB-R data pairs as estimated by the 2019 WGWIDE assessment are shown in Figure 2.8

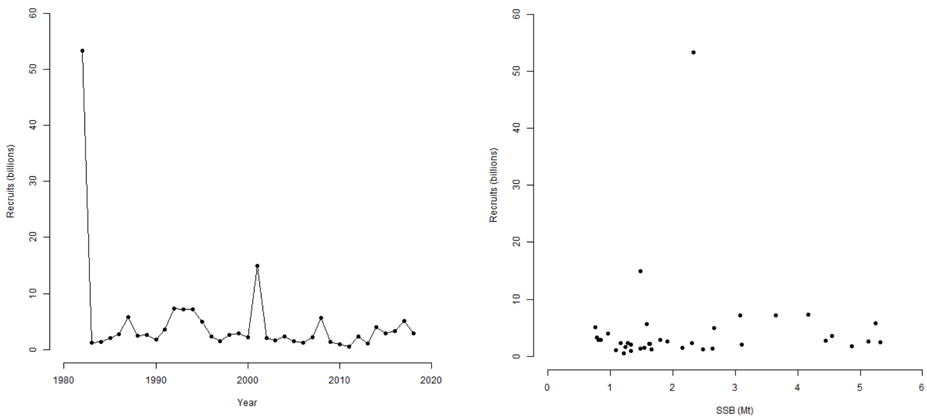


Figure 2.8 WGWIDE 2019 Western Horse Mackerel Assessment: time series of recruitment (left) and stock-recruit pairs (right).

The recruitment time series is dominated by the extremely large event in the first data year, 1982. This single year class permitted the expansion of the fishery and sustained catches for over a decade, accounting for more than 50% of the total annual catch by abundance in some years. A second large recruitment occurred in 2001, although this was approximately 1/3 of the size of that of 1982. There are occasional extended periods (3-5 years) of relatively low recruitment during the time series with indications of autocorrelation. Recruitment in the most recent period has increased following the two lowest on record in 2010 and 2011.

Extensive investigations of the S-R relationship have been carried out during previous MSE exercises, the assessment benchmark of 2018 and 2019 ICES Interbenchmark Protocol on reference points (Campbell *et al*., 2015; ICES, 2017; ICES, 2019). Consideration has been given to both single and mixed model approaches although it was found that this led to some implausible model fits with the relative proportions of the individual models highly sensitive to the 1982 and 2001 data points. Investigations including the fitting to the complete SR dataset and also the excluding the large recruitments with consideration also given to separate modelling the occurrence of a spike.

The most recent examination was conducted during the 2019 Interbenchmark (ICES, 2019). This exercise concluded that the SR dataset should be restricted to data points from 1995 onwards, to exclude not only the very large 1982 recruitment event but also the subsequent years when the contribution of this single cohort to the total SSB (peaking at ~65% by weight) exceeded the next largest from 2001. Nothing comparable to the 1982 recruitment has been observed over the subsequent 40 years and the magnitude of the largest recruitments (*e.g.* 2001,2008) have been much reduced. It was therefore considered appropriate that, for the purposes of modelling future recruitment, an event of the magnitude of 1982 should not be considered likely, hence the restriction to data from 1995 onwards. There is no evident stock-recruit relationship and therefore the IBP characterised recruitment for the *EqSim* long term projections by a segmented regression with a breakpoint set at B_{lim}. For the purposes of these simulations, the same approach is adopted.

The *EqSim* package (function *eqsr_fit*) was used to provide 1000 SRRs for the projection. Each SRR is based on a fit of a segmented regression constrained at B_{lim} to a bootstrap

SSB-recruitment dataset (1995 onwards, excluding the terminal assessment year). The associated cv and estimated autocorrelation (AR1) are used during the projection period to draw stochastic recruitments based on the iteration specific SRR parameters and the operating model SSB. The fit and predicted distribution from the 2019 assessment dataset is shown in **Figure** 2.9. The grey points in **Figure** 2.9 represent a number of stochastic draws over the observed range of SSB.

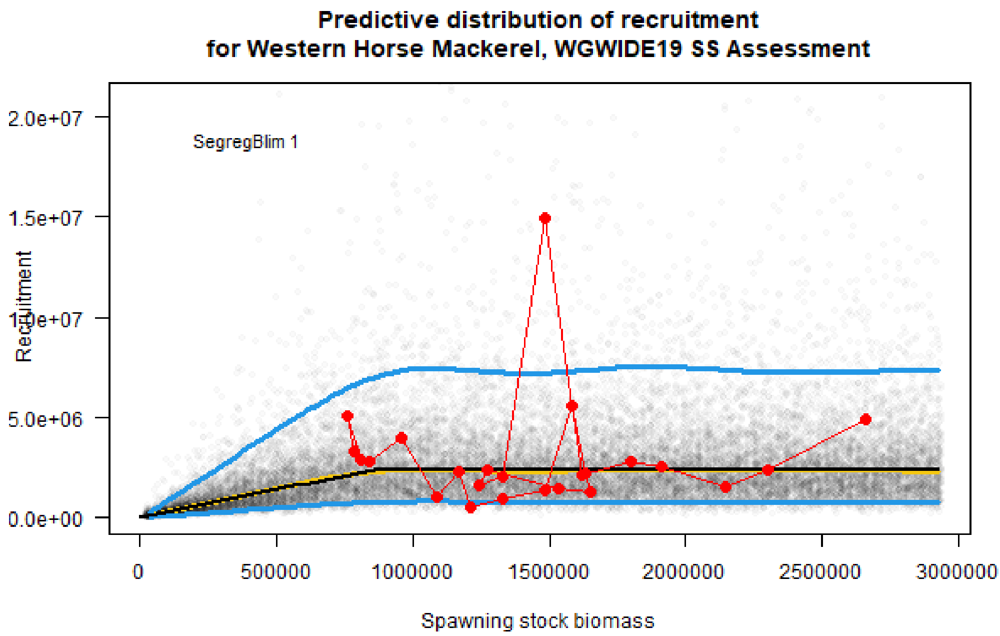


Figure 2.9 EqSim for WGWIDE 2019 based on SSB-Rec pairs from 1995-2017 and a segmented regression constrained at Blim.

Figure 2.10 displays the same data in the form of a cumulative distribution and can be used to examine the ability of the approach to replicate the observed recruitment distribution (indicated by the points).

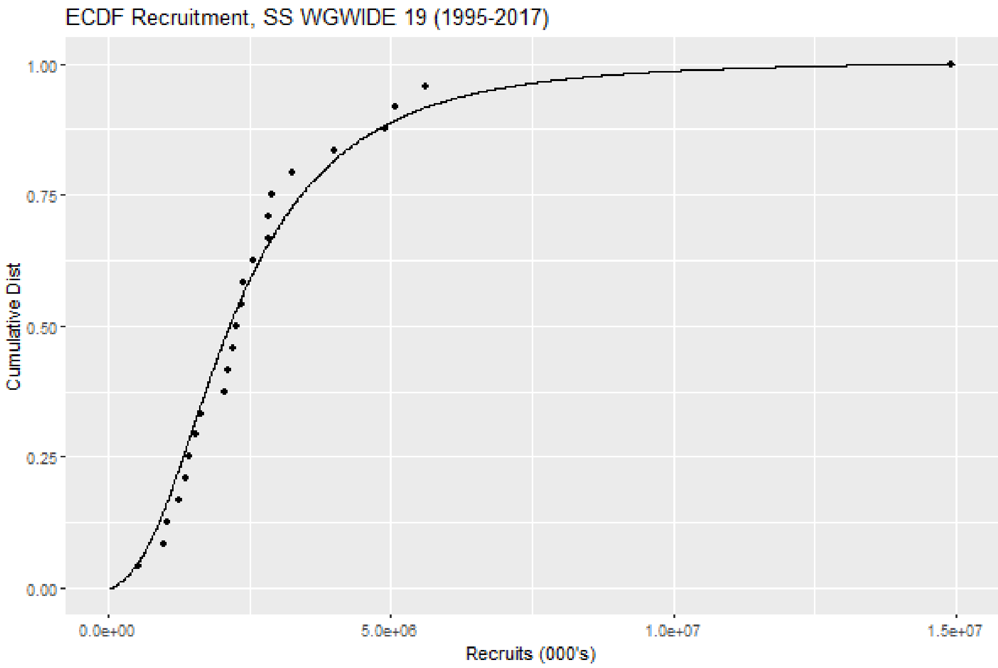


Figure 2.10 ECDF of observed (WGWIDE 2019, 1995-2017) and *EqSim* modelled recruits. Note the highest data point corresponds to the 2001 recruitment event.

2.4 Provide descriptions of all assumptions for the forward projections including rational. In the case of SAM, specify the process errors that are carried forward in projections and provide estimates of their variance (standard deviation).

A summary of the MSE is provided below in the format requested by ICES MSE guidelines (developed by the series of WKGMSE workshops)

Background		
Motive/initiative/background.	There is no agreed management plan currently in place for western horse mackerel. The Pelagic Advisory Council (PELAC) initially proposed a plan in 2007 and this was used for the purposes of catch advice until 2010. Since this time, advice has been based on the MSY approach and the PELAC has continued efforts to develop a new management plan. The evaluation that is presented, has been conducted by a group of experts, on request from the PELAC. On the basis of the evaluation, the PELAC has selected a preferred management procedure.	
Main objectives	Rebuilding of the stock to levels well above B_{lim}. In recent years, the median SSB estimate from the stock assessment has been just above B_{lim} . During periods of lower than average recruitment, management should lead to catches that maintain the stock above B_{lim} with a high probability whilst avoiding large interannual changes in TAC.	
Formal framework	EC request to ICES to evaluate the MSE conducted by the PELAC technical focus group	
Who did the evaluation work	PELAC technical focus group on horse mackerel (Pastoors *et al*., 2020)	
Method		
Software Name, brief outline include ref. or documentation	1. A further adaptation of SimpSIM which was, initially developed from the *EqSim* and FLR R packages for the 2016 blue whiting MSE (ICES, 2016). Code of the adapted version is on github (https://github.com/ices-eg/wk_WKREBUILD; folder EqSimWHM) . The *EqSim* based simulator was conditioned on the SS3 assessment model as carried out during WGWIDE 2019. Additional runs were also conditioned with the WGWIDE 2020 assessment and with the exploratory SAM assessments (2019, 2020) as reported in the relevant expert groups (ICES, 2019; ICES, 2020b). The *EqSim* simulator is based on an age-structured operating model, with catches, numbers and F at age derived from simulated assessments. Assessment/advice error on SSB and F is incorporated following guidelines from WKMSYREF3 (ICES, 2014) 2. SAM HCR forecast based on the exploratory SAM assessment. The WHOM assessment model scripts and results used in this study are available at https://www.stockassessment.org/set-Stock.php?stock=WHOM2019. The HCR forecast in SAM that is used in this study was first developed and presented in Trijoulet et al. (Trijoulet et al., 2021). The code is available at https://github.com/vtrijoulet/SAM/tree/master2 and the main function is forecast2(). The forecast is a stochastic forecast that allows HCRs in the management procedure. It was set up to replicate as closely as possible the settings in the *EqSim* simulator.	
Type of stock	Medium life span (>20 years), pelagic	
Knowledge base *	Analytic assessment (SS3) and exploratory assessment (SAM)	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Segmented regression with a breakpoint at B_{lim} . Four scenarios considered (base case and three alternative (more pessimistic) scenarios whereby recent increases in recruitment (since 2014) is reduced).	Log-normal, CV from residuals

Weight at age	Random selection from stochastic time series of catch/stock weight at age.	Weight at age estimates from each of the 1000 model replicates were used to generate a time series stochastic weights via application of an error term derived from variability in the historic weight at age datasets to the replicate mean. The error term is considered age specific and autocorrelated.
Maturity	Maturity during the simulations is time invariant (as in the assessment)	No stochasiticity implemented
Natural mortality	0.15 – all ages and years	No stochasiticity implemented
Selectivity	From assessment	The MSE draws at random from the 1000 individual selection profiles available from the replicates for each year of the projection period.
Initial stock numbers	From assessment	Using 1000 model iterations to derive starting values.
Reference points	OM assessment-specific reference points are estimated (because of different partitioning between SSB and F for the exploratory SAM and WGWIDE SS3 assessments)	
	WGWIDE/SS3 reference points B_{lim} 834,480t B_{pa} 1,168,272t F_{MSY} 0.074	SAM reference points B_{lim} 611,814t B_{pa} 856,540t F_{MSY} 0.115
Decision basis **	SSB in the TAC year	
Number of iterations	1000	
Projection time	23 years	
Observation and implementation models (no assessment in the loop)		
Type of noise	*EqSim*: historic advice error on F and SSB SAMHCR: uncertainty directly taken from SAM assessment (function: simulate)	
Comparison with ordinary assessment?	Advice uncertainty is larger than model parametric uncertainty (SS3 estimates relative low uncertainties). The historic advice error for the last ten years was derived according to the methodology detailed in WKMSYREF3 (ICES, 2014) SAMHCR is based on uncertainty estimated within the SAM model.	
Projection: If yes - how?	*EqSim*: no forecast included. TAC derived directly from SSB estimate taking into account the advice error that was estimated for the forecast. SAMHCR: is essentially a multi-annual forecast using different HCR types.	
Projection: Deviations from WG practice?	Yes, WG runs a deterministic projection based on the SS3 model	
Implementation	Catches in numbers at age from projection according to the rule. No implementation error assumed.	
Harvest rule		
Harvest rule design	Proposed HCR: Double breakpoint on SSB: If SSB < Blim, Flow = 0.2 * Ftarget If Blim<SSB<Bpa: linear slope between Flow and Ftarget If SSB > Bpa: F = Ftarget Fixed F and ICES advice rule (with breakpoint at Bpa) HCRs were also evaluated.	
Stabilizers	Optionally, 20% TAC change limits (only applied above $B_{trigger}$ (B_{pa}) Scenarios with fixed minimum catch (50 kt) were run.	
Duration of decisions	Annual	
Revision clause	LTMP specifies a revision in no more than 5 years from first year of implementation.	
Presentation of results		
Interest parameters	The probability of stock size exceeding a biomass threshold *e.g.* B_{lim} or B_{pa}. For this MSE, the stock is considered to have achieved rebuilding when the probability that biomass exceeds B_{pa} is at least 50% for three consecutive years. Other parameters of interest are yield and inter-annual variation in yield.	
Risk type and time interval	Type 3	

Precautionary risk level	5%
Experiences and comments	
Review, acceptance:	Under review (WKWHMRP 2021)
Experiences and comments	Running two different operating models (SS3 and SAM) was useful to explore the sensitivity of the rebuilding probability to the different model formulations. Running two different evaluation methods (EqSim and SAMHCR) was useful to explore how the results of the management procedure are dependent on the evaluation method. Model results were comparable but not identical. Main difference is due to the different perception of the stock relative to reference points.

Table 2.2 Western Horse Mackerel MSE: WKGMSE Summary template. template

2.5 Provide detailed description of how the assessment error model is applied in the simulations.

No assessment or forecast is conducted during the simulation *i.e.* this is a 'short-cut' MSE approach. The parameterisation of uncertainty in observed SSB and F are incorporated into the simulations following the methodology developed for *EqSim* whereby SSB and FBar outputs from historic assessments and short term forecasts are compared to the most recent assessment and described via a pair of parameters (CV and autocorrelation) for each of SSB and FBar. These parameters are then used to generate an error and derived "observed" values of SSB and Fbar for the management procedure.

The SSB used in the application of a harvest rule in year y ($SSBobs_y$) is derived from the underlying population in the same year ($SSBtrue_y$) via

$$SSBobs_y = SSBtrue_y * \exp\left(SSBerr_y\right)$$

where

$$SSBerr_y = SSB_{phi} * SSBerr_{y-1} + rnorm(0, SSB_{cv})$$

A similar approach is adopted for the calculation of the fishing mortality that will be applied to the stock ($Ferr_y$), modifying the value initially supplied by the harvest rule (F_{HCRy}):

$$Ferr_y = F_{HCRy} * \exp\left(Ferr_y\right)$$

where

$$Ferr_y = F_{phi} * Ferr_{y-1} + rnorm(0, F_{cv})$$

The base line values adopted for this exercise are

FCV	FPHI	SSBCV	SSBPHI
0.2193	0.0212	0.2927	0.5776

The ratio of perceived to true SSB for the double breakpoint rule with an F_{target} value of 0.075 is shown in Figure 2.11. The grey lines and coloured dots represent 3 individual replicates. The black line during the simulation period represents the median annual value and that from the historic period the ratio of the annual SSB estimates from the assessment upon which the operating model is conditioned (WGWIDE 2019) with the estimate of SSB from the historic assessments[2].

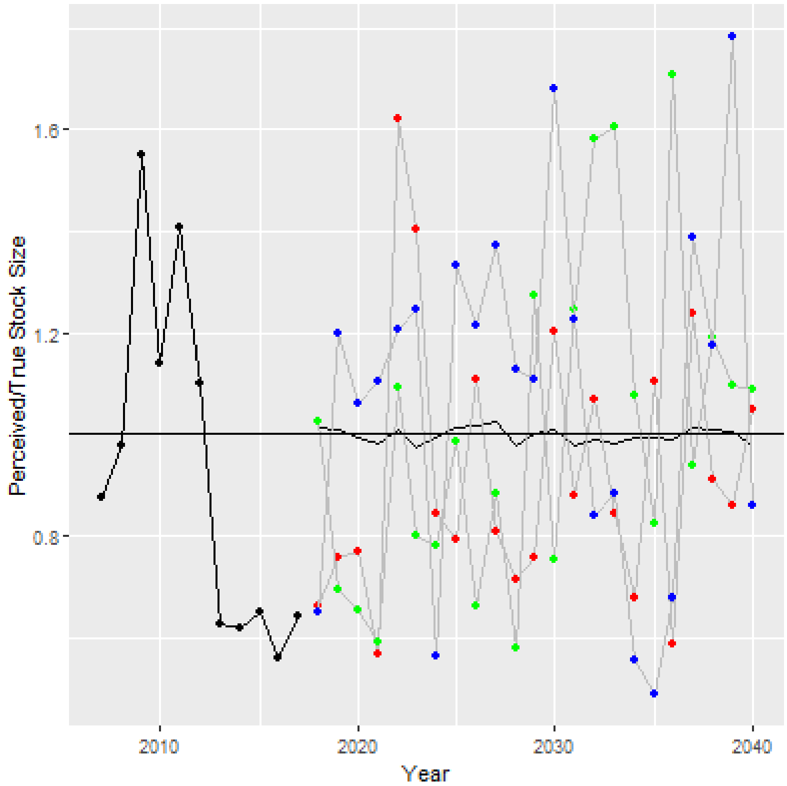


Figure 2.11 Ratio of perceived stock biomass versus true stock biomass. Historic period is pre-2018

Assessment error in SAM HCR forecast

The SAM HCR forecast does not account for assessment error as explicitly as the EqSim simulator, but the forecast starts with 1000 starting values for the stock population and fishing mortality, this allows accounting for current stock perception uncertainty and error in the assessment. In addition, the projections are run with propagating the process errors on fish survival, which could somehow relate to the error in SSB added in the *EqSim* simulator.

2.6 Add material (figures, tables) that illustrates that the properties of projected biomass are consistent with historical estimates as per ICES guidance on simulations using the short-cut method.

The presentation of the temporal evolution of key stock metrics (SSB, Fbar, Recruitment, Yield, Risk etc.) now include the historic period (since 2000) and are presented in matrix format to facilitate comparisons to be drawn between the various management strategies and HCR parameter values.

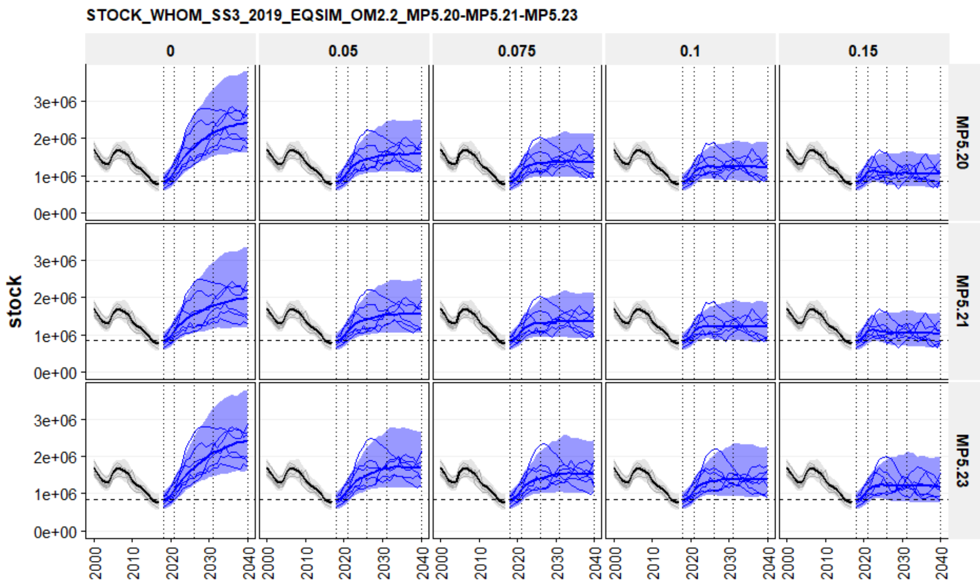
[2] NB: prior to 2017 the SSB reported by the assessment was at spawning time (May) whereas the assessments from 2017 – 2019 report SSB on Jan 1).

As an example, the plots below show the development of stock size and fishing mortality from 2000 to the end of the simulation period. These plots compare outputs for 3 management strategies with each run for five target fishing mortalities, as indicated im the column title bar. The management strategy code is shown to the right of each row with

MP5.20 – The double breakpoint HCR

MP5.01 – The double breakpoint HCR with a 50kt minimum TAC

MP5.23 – The double breakpoint HCR with a 20% change limitation in TAC when above $B_{trigger}$.



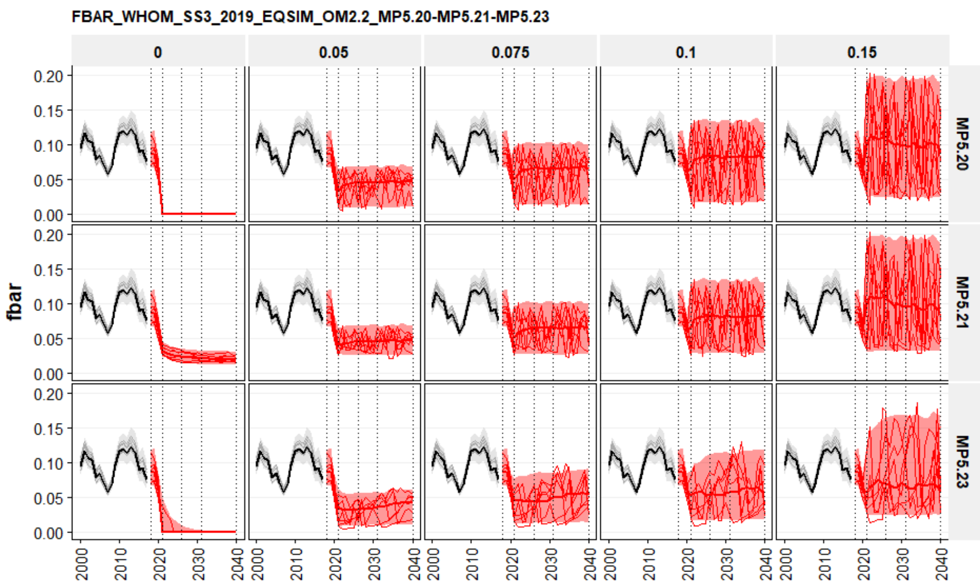


Figure 2.12 Example of simulation summary based on the 2019 SS3 assessment for three different MPs: MP5.20 Double BP; MP5.21 Double BP with minimum TAC of

50kT; MP5.23 Double BP with 20% IAV constraint above Btrigger. More results in section 8

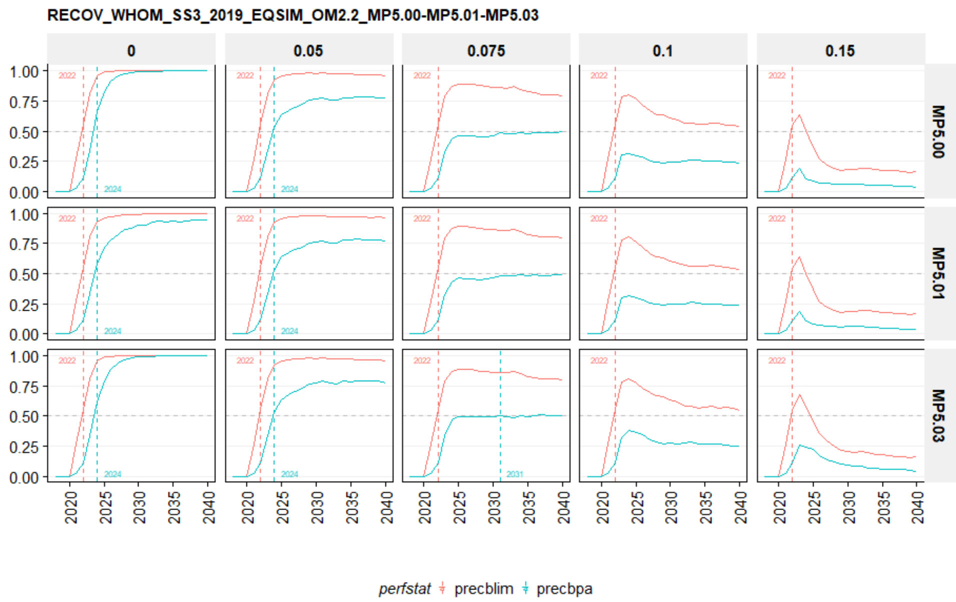
2.7 Show the results of the actual rebuilding criteria suggested in the plan: minimum of 3 years above the biomass threshold (initially Blim, ultimately Bpa).

The status of the stock with respect to rebuilding above a biomass threshold is determined by the number of replicates with SSB greater than the threshold value for a period of 3 or more years following a period of 1 or more years below the threshold. Recognising that some replicates may never fall below the biomass threshold throughout the simulation period and that some may start above before falling below (e.g. as a result of the catch constraints imposed in the initial years), the annual proportion of replicates that meet the criteria is calculated based on the total number of replicates that have fallen below the threshold in at least one of the years between the start of the simulation period and the year of calculation.

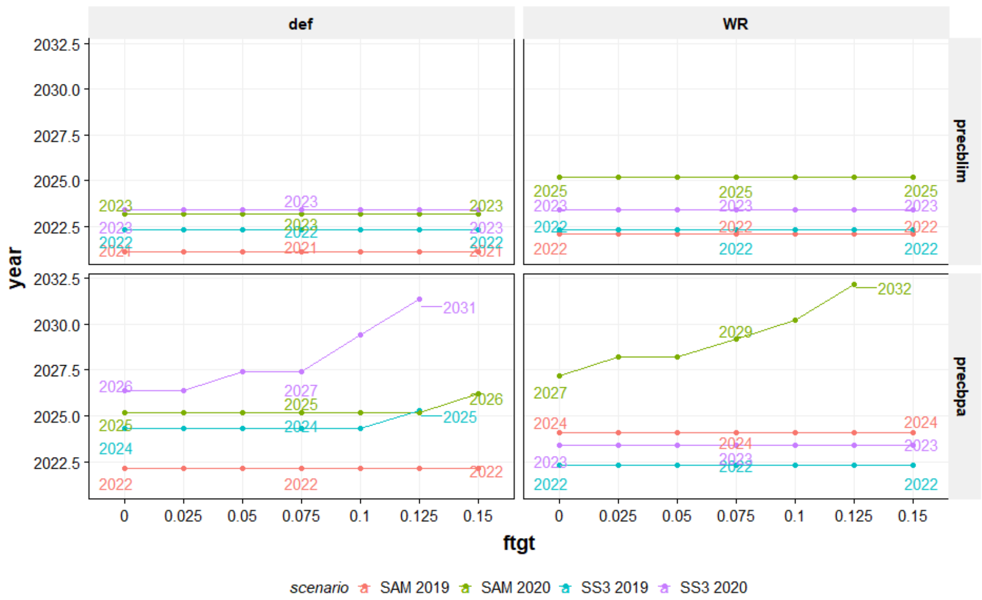
For this evaluation, the metric used to monitor the rebuilding status of the stock is the simulation year in which the SSB has exceeded B_{pa} for at least 3 years following a period of 1 or more years below B_{pa} for at least 50% the replicates. The corresponding year for the B_{lim} threshold is also reported.

The year of rebuilding to B_{lim} and B_{pa} has been visualized for all scenarios in the Annex (section 8), using two types of plots:

1. Annual probability of being rebuilt (3 consecutive years, precblim/precbpa) above a biomass threshold. The first year that the rebuilding probability is more than 50% is indicated by the vertical line



2. Comparison of the first year to achieve the rebuilding criteria for different sensitivity exploration (using the double breakpoint HCR with IAV constraint) and different target fishing mortality (in the specific example shown, it deals with the comparison between the appropriate assessment model specific reference points, and using the wrong reference points, *i.e.* estimated for the other assessment model).



2.8 Describe how the SS3 and SAM implementations parameterize initial conditions in a manner that overcomes the lack of uncertainty in initial conditions in the analysis by Cox et al. (2018).

2.8.1 SS3: incorporating Uncertainty into Initial Conditions and Parameterising Variability in Catch and Stock Weights at Age and Fishery Selection.

The simulation start date coincides with the terminal year of the assessment upon which the operating model is conditioned. To incorporate uncertainty in the various population vectors (abundance, stock and catch weight and fishery selectivity), the variance-covariance matrix from the appropriate assessment was used to compile a set of 1000 stock replicates for the purposes of projection.

Initially, 10,000 parameter sets were randomly drawn from the variance-covariance matrix using the R multivariate normal *mvrnorm* function from the MASS package. Any draws with parameter values outside of the bounds defined in the assessment configuration were rejected. A subset of 1000 replicates were then selected such that the median and 95% confidence intervals of the terminal year SSB match those from the stock assessment. The stock development for 10 random iterations and the final assessment are shown in Figure 2.13 while Figure 2.14 compares the distribution of SSB in the terminal assessment year from the 1000 replicates.

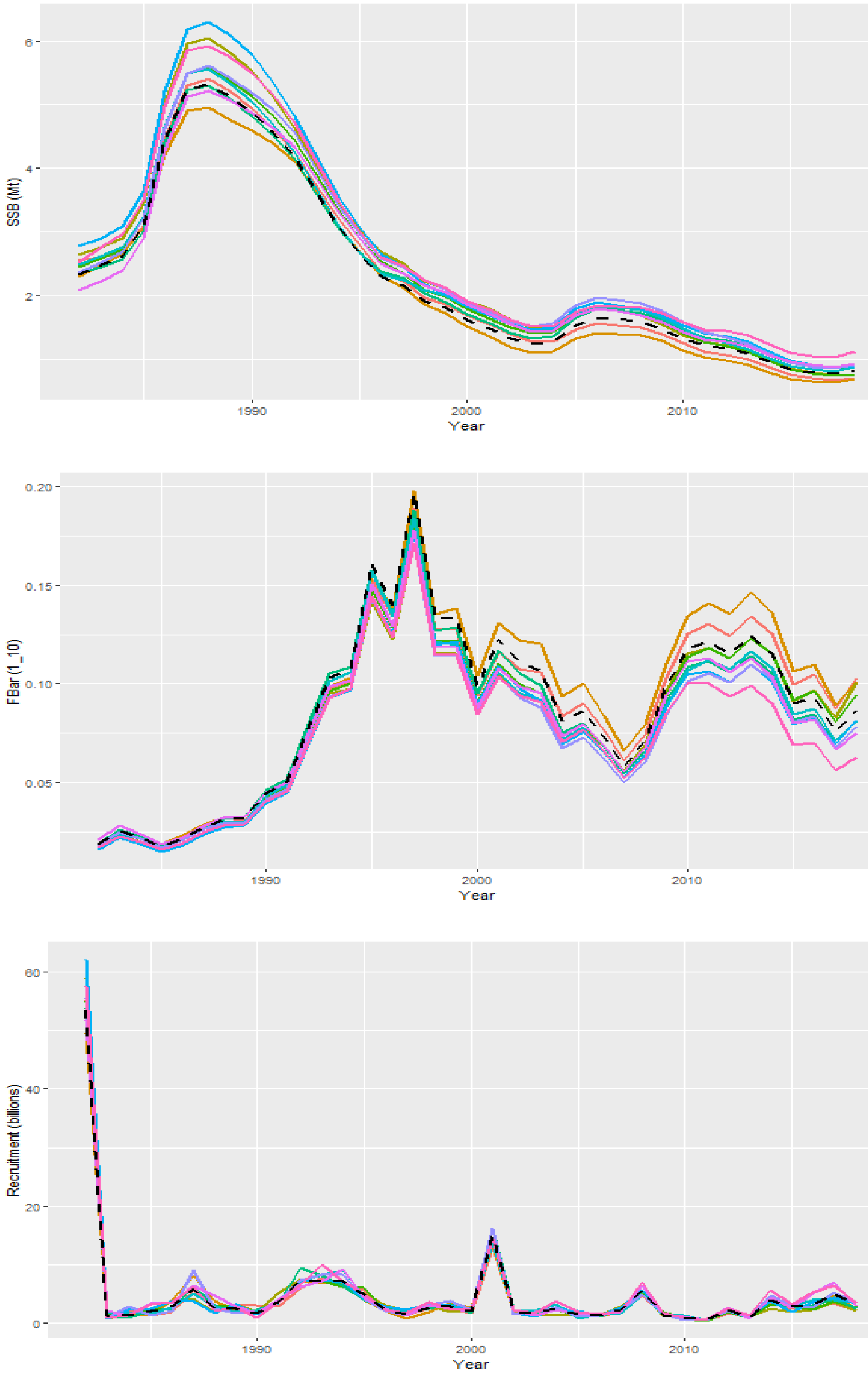


Figure 2.13 Ten randomly selected iterations (assessment in black), WGWIDE SS 2019

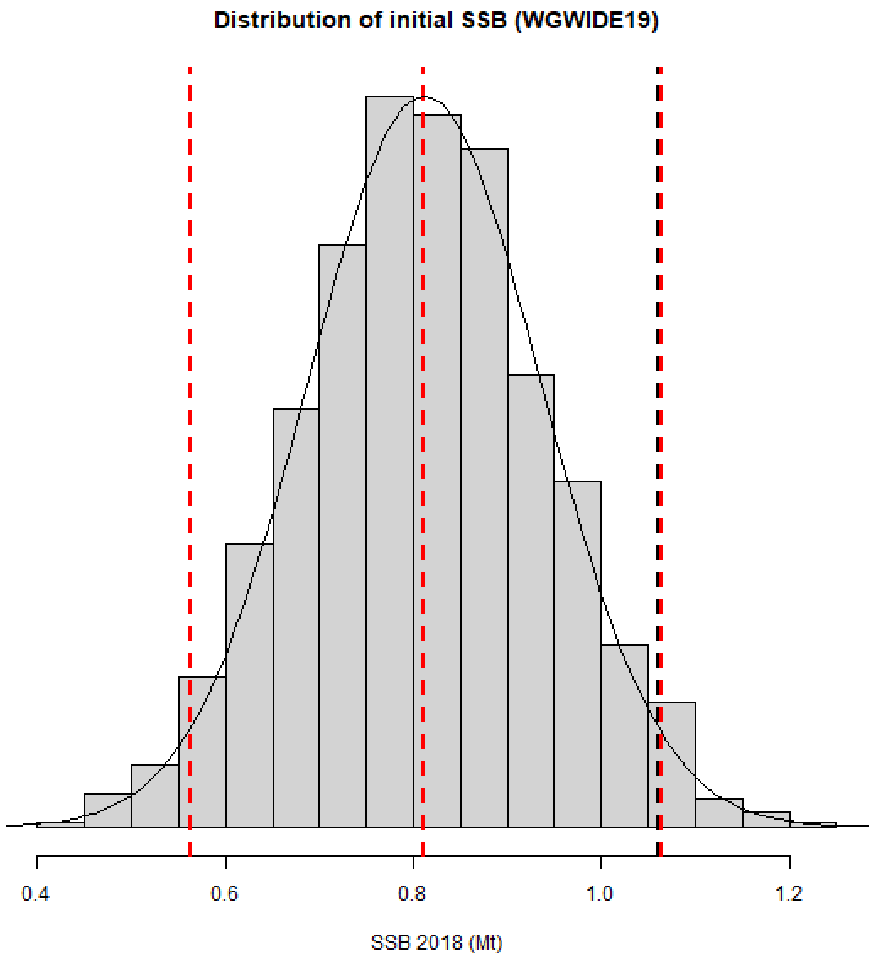


Figure 2.14 Distribution of SSB, 2018

The distribution of abundance at age for the simulation start year from the 1000 replicates is shown in Figure 2.15. The red line indicates the assessment point estimate.

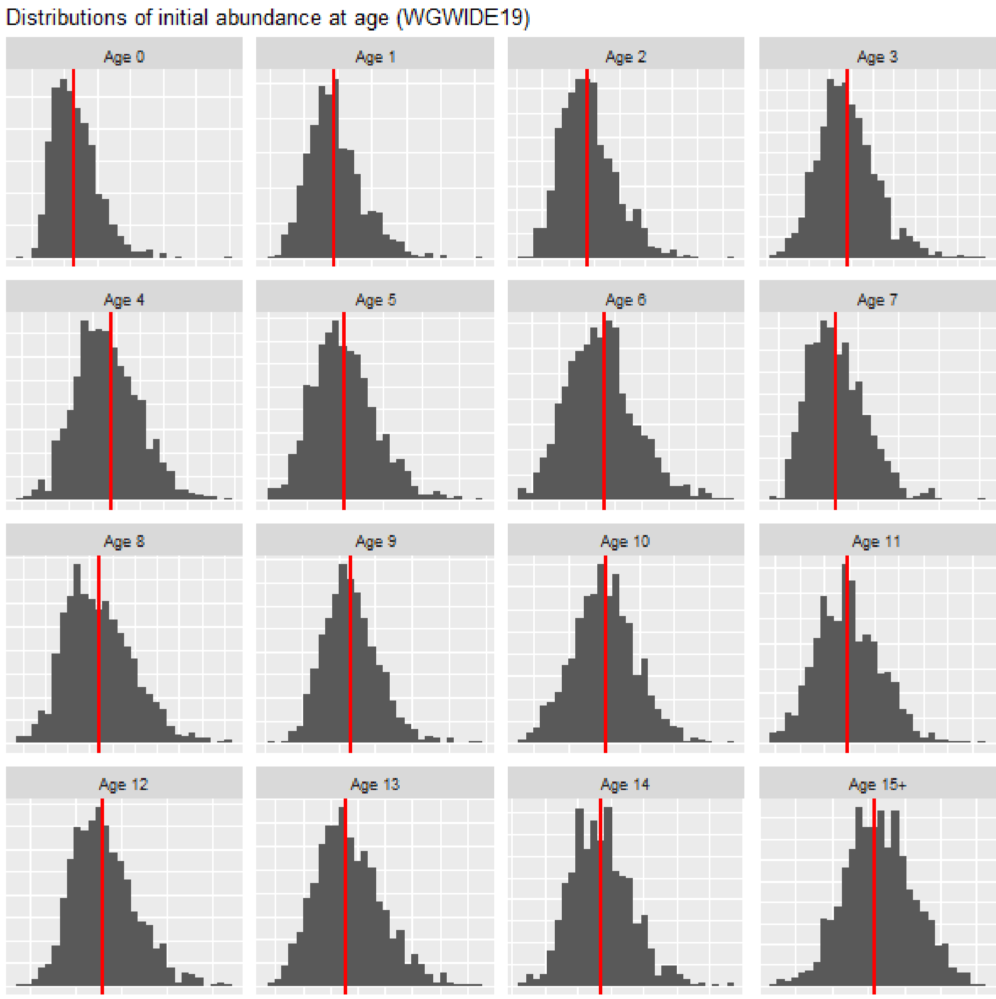


Figure 2.15 Distribution of abundance at age, 2018. Red line WGWIDE19 final SS assessment.

Stock and Catch Weights

As currently configured, the SS3 assessment upon which the OM is configured assumes a time invariant von Bertalanffy growth model with weight at age derived from modelled length via a static length-weight relationship. The input data for the assessment includes length composition data and conditional age at length (see table 2.1 for further detail). As a consequence of this configuration, the assessment estimates a fixed weight at age.

There are however, clear indications of fluctuations in weight at age as revealed by the sampling of commercial catch and it was considered appropriate to include an appropriate level of variability in weight at age for the MSE simulation period. A log-normally distributed, auto-correlated (AR1) age-specific error term was randomly drawn and applied to the replicate-specific estimate of weight at age for each year of the projection. The (age specific) CV and autocorrelation values for each age were derived from the historic weight at age data from 2000 onwards, excluding the period of rapid change prior to this. Within each year of the projection, the same error term is applied to both the stock and catch weight data series as historic stock and catch weight at age is autocorrelated, to be expected since the stock weights are simply derived from a subset of the catch weight information. For some ages, there is evidence of an increasing trend in weight at age (e.g. ages 7-10) while some (generally the younger ages) appear more constant. The simulation assumes no trend in weight at age during the projection. Continuation of the trend observed into and through the simulation period would likely lead to unrealistic and previously unobserved weights for certain age classes. Figure 2.16 (stock weights at age) and Figure 2.17 (catch weights at age) show the historic data (black), the 95% range of values from the 1000 assessment iterations (red), 95% range of future weights (blue) and 3 randomly selected iterations (grey).

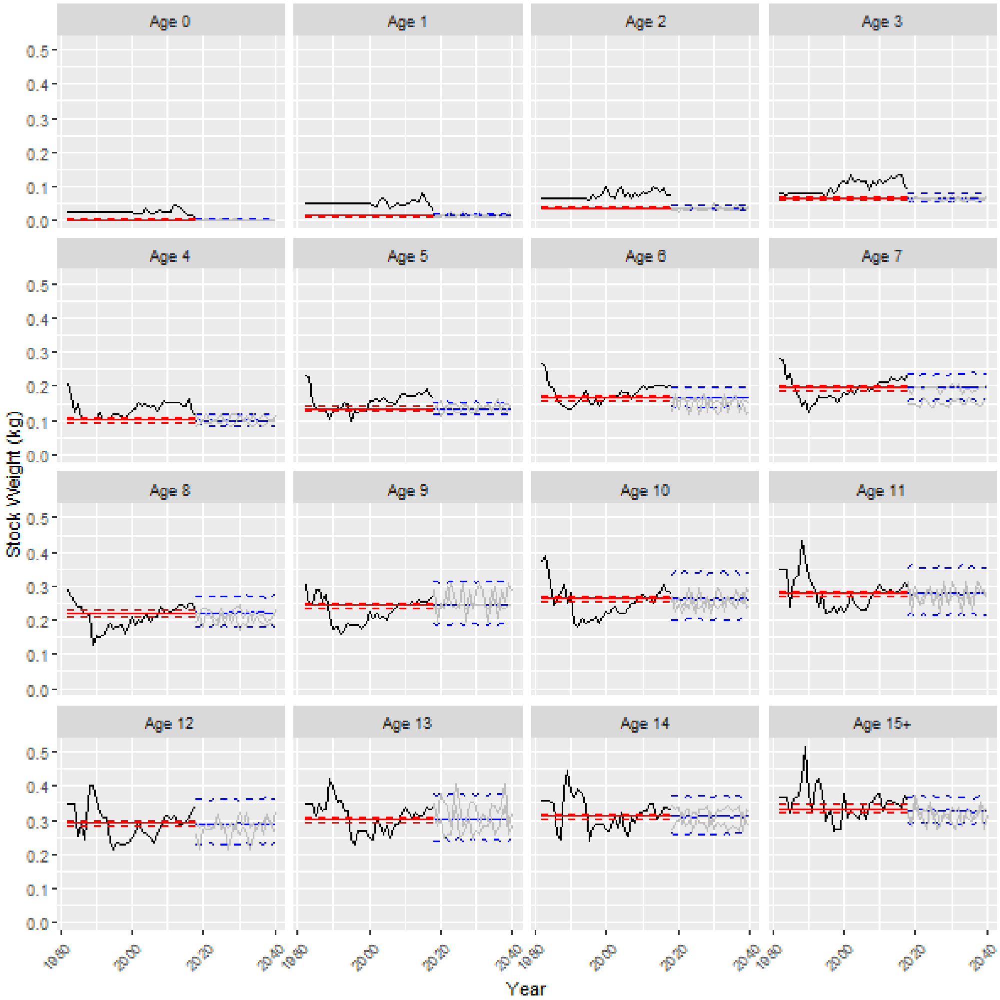


Figure 2.16 Stock weight at age, historic data (black line), SS3 estimates from the 1000 replicates (red lines), simulated future weights from a random selection of assessment estimates with stochasticity (blue lines) and 3 random iterations (grey lines).

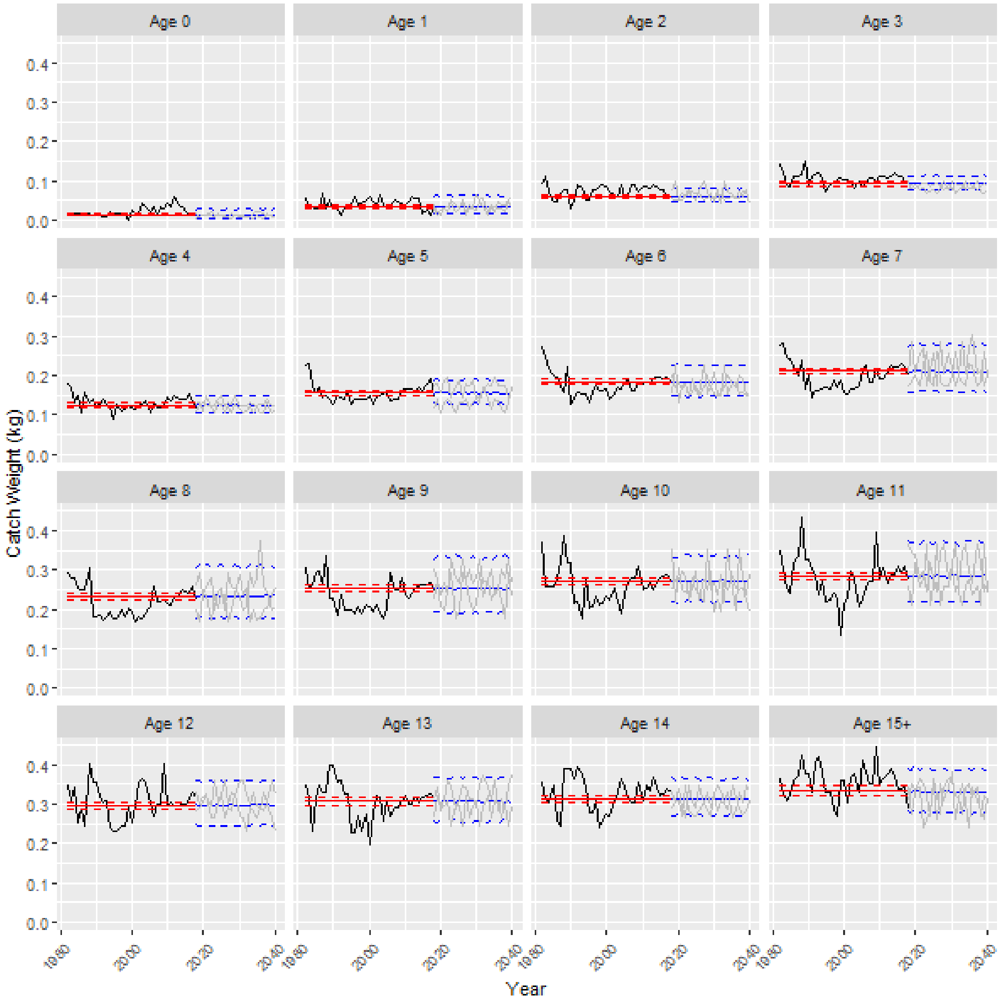


Figure 2.17 Catch weight at age, historic data (black line), SS3 estimates from the 1000 replicates (red lines), simulated future weights from a random selection of assessment estimates with Stochasticity (blue lines) and 3 random iterations (grey lines).

Fishery Selectivity

SS3 is configured with a single sex, single fishery fleet with selectivity set up as a double normal, forcing the selectivity to be asymptotic. In the current assessment configuration, the selectivity is assumed time invariant. However, it was considered appropriate to introduce variability in selectivity over the projection period. This was achieved by drawing a selectivity pattern at random from the collection of 1000 selection profiles available from the 1000 stock replicates derived from the variance-covariance matrix of the assessment upon which the OM is configured. A new selectivity pattern was selected for each year of the simulation for each replicate. To visualise the variability in selection random selection of 10 individual selection patterns are shown in Figure 2.18, along with the profile from the WG assessment.

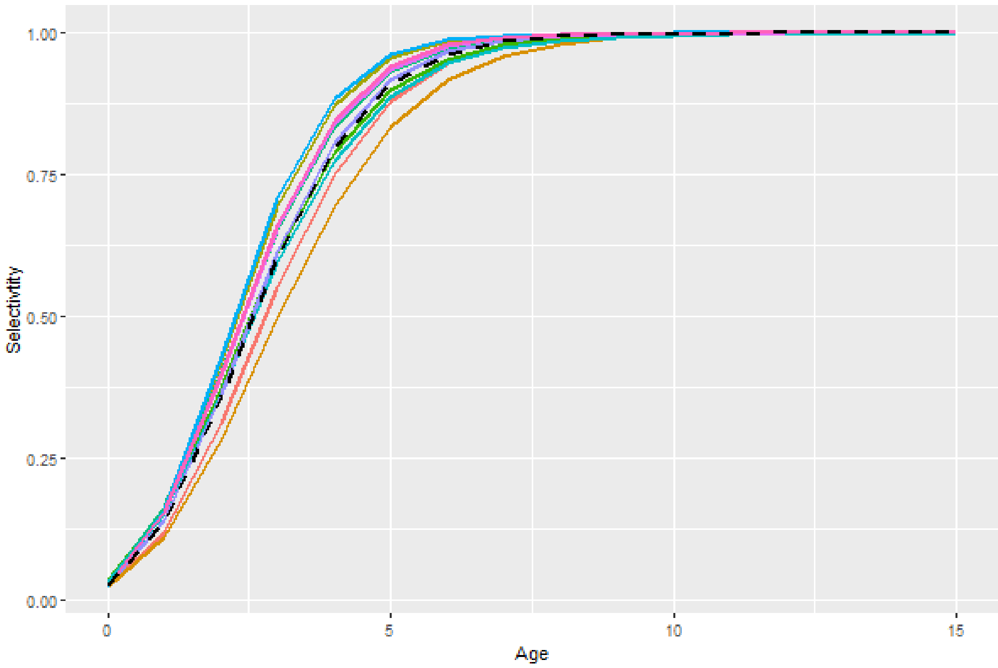


Figure 2.18 WGWIDE 2019 selection pattern (black dashed line) and 10 randomly selected selection patterns from the 1000 replicates derived from the assessment parameter variance-covariance matrix

Selection is most variable for the younger ages (1-6). The youngest age class (age 0) has a very low selection whilst ages 7 and over are fully selected. An indication of the range of selectivity by age class can be seen in Figure 2.19.

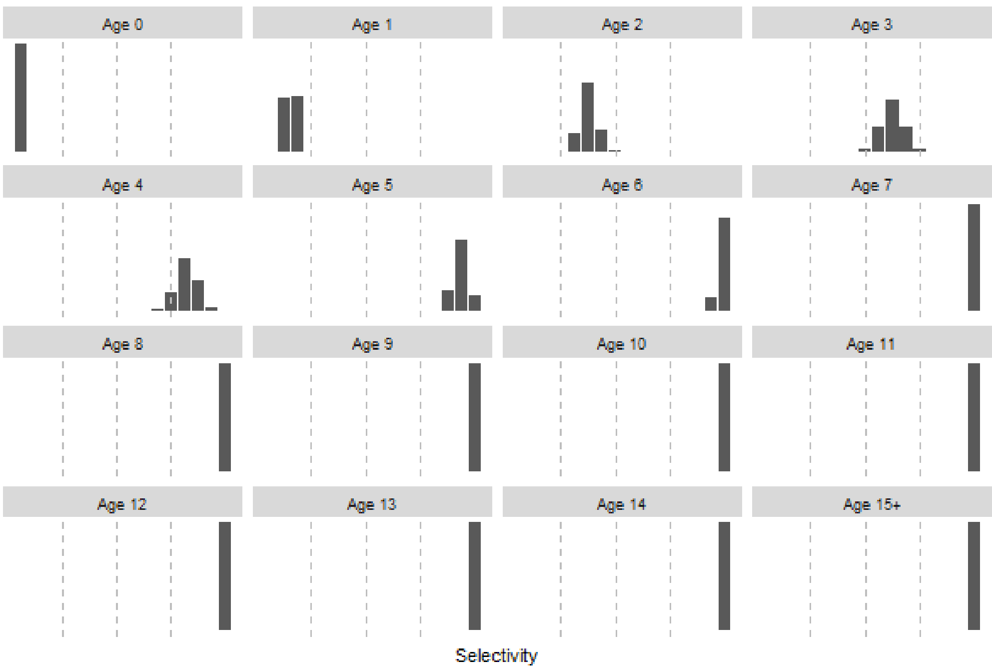


Figure 2.19 Histograms of selectivity at age from the 1000 replicates. The vertical dashed lines correspond to 25%, 50% and 75% selectivity.

Historic (assessment, time-invariant) and future selectivity for ages 0-5 are shown in Figure 2.20.

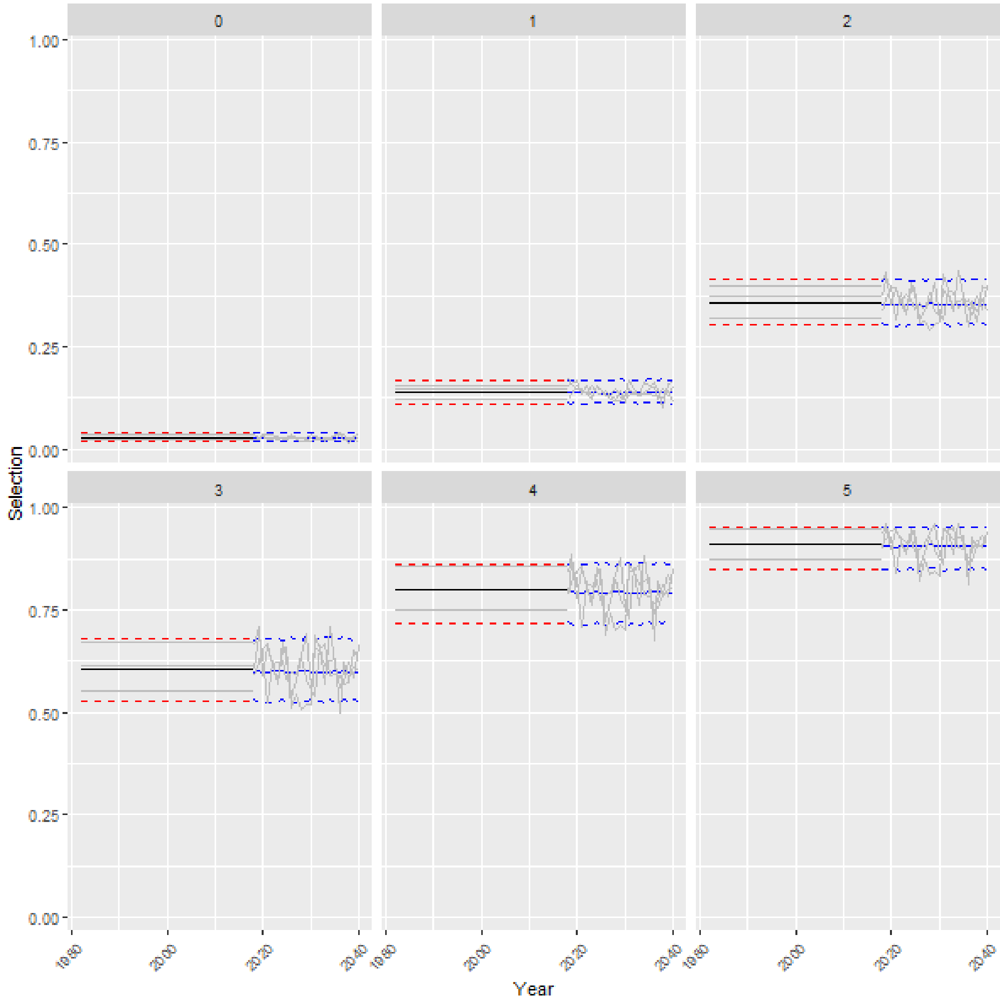


Figure 2.20 Selectivity for ages 0-5. 2019 SS3 assessment estimate (black line), SS3 estimates from the 1000 replicates (red lines), 3 randomly selected iterations (grey lines), simulation period constructed by randomly selecting from 1000 replicates.

Maturity and Natural Mortality

Maturity and natural mortality during the simulation are assumed time invariant. In the assessment, natural mortality is assumed equal to 0.15 for all ages and this is maintained for the simulation. The maturity ogive has been unchanged since 1998 and is shown in Figure 2.21.

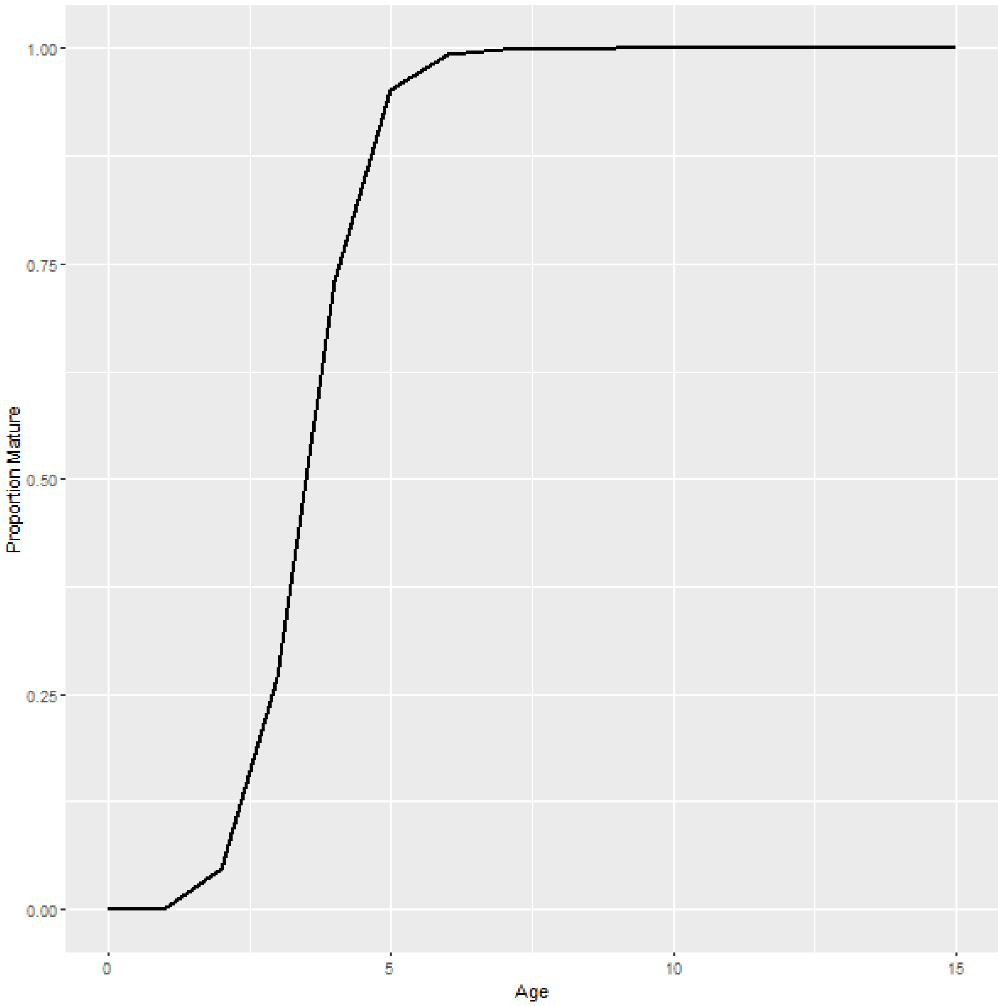


Figure 2.21 WGWIDE 2019 Western Horse Mackerel maturity ogive.

2.8.2 SAM: generating

2.9 Elaborate on the approach of using historical advice error rather than assessment error based on the most recent assessment. Include robustness testing of the basecase scenario conclusions on the rebuilding plan to the values used the advice error model (Fcv and Fphi). i.e. evaluate the sensitivity of results to the uncertainties used (Fcv and Fphi); e.g. explore what would happen if default values of Fcv and Fphi are used instead of the estimated values based on only 8 years.

The inclusion of assessment/advice error (as described in §2.5) in the baseline simulations is based on error function parameters calculated by comparing the output of the 2019 WGWIDE assessment with the assessments and short term forecasts conducted between 2010 and 2018. The methodology adopted is the same as that used to parameterise assessment/advice error for the purposes of deriving an estimate of F_{MSY} using ICES MSY R package (*EqSim*) and is detailed in the WKMSYREF3 report (ICES, 2014). The values adopted for the baseline simulations are:

FCV	FPHI	SSBCV	SSBPHI

0.2193	0.0212	0.2927	0.5776

During the simulation, the fishing morality actually applied to the stock is based on the HCR target value, modified via an autocorrelated error term calculated from Fcv and Fphi. During the decision sub-model of the MSE, future harvesting decisions are based on an observed, perceived SSB which is derived from the underlying, true SSB modified by an error characterised by SSBcv and SSBphi.

Due to the relatively short time series of assessment output available upon which to base a calculation of the error parameters and to investigate the sensitivity of the MSE results to the levels of assessment error, additional simulations were conducted. These simulations adopted the default parameter values used by *EqSim* analyses lacking stock specific values *e.g.* stocks with no or limited assessment history. Based on median values from a suite of stocks analysed during the development of the MSY framework, default values for Fcv and Fphi are 0.24 and 0.42 respectively. For this exercise, the same values were selected for the SSB equivalents. The default CV values are relatively similar to those calculated from the historic Western Horse Mackerel assessments. However, the AR1 autocorrelation phi parameter differs markedly, particularly for the fishing mortality error.

Additional sensitivity simulations were also conducted assuming “perfect knowledge” such that zero error is assumed in the assessment process and also with an arbitrary high correlated error (cv=0.6, phi=0.6). Although zero error is unfeasible in reality, this permits an insight into the relative contribution of this component of uncertainty compared with the other sources included in the MSE.

A comparison of the risk to B_{lim} for the constant F and double breakpoint harvest rules, each with a target fishing mortality of 0.075 is shown in Figure 2.22.

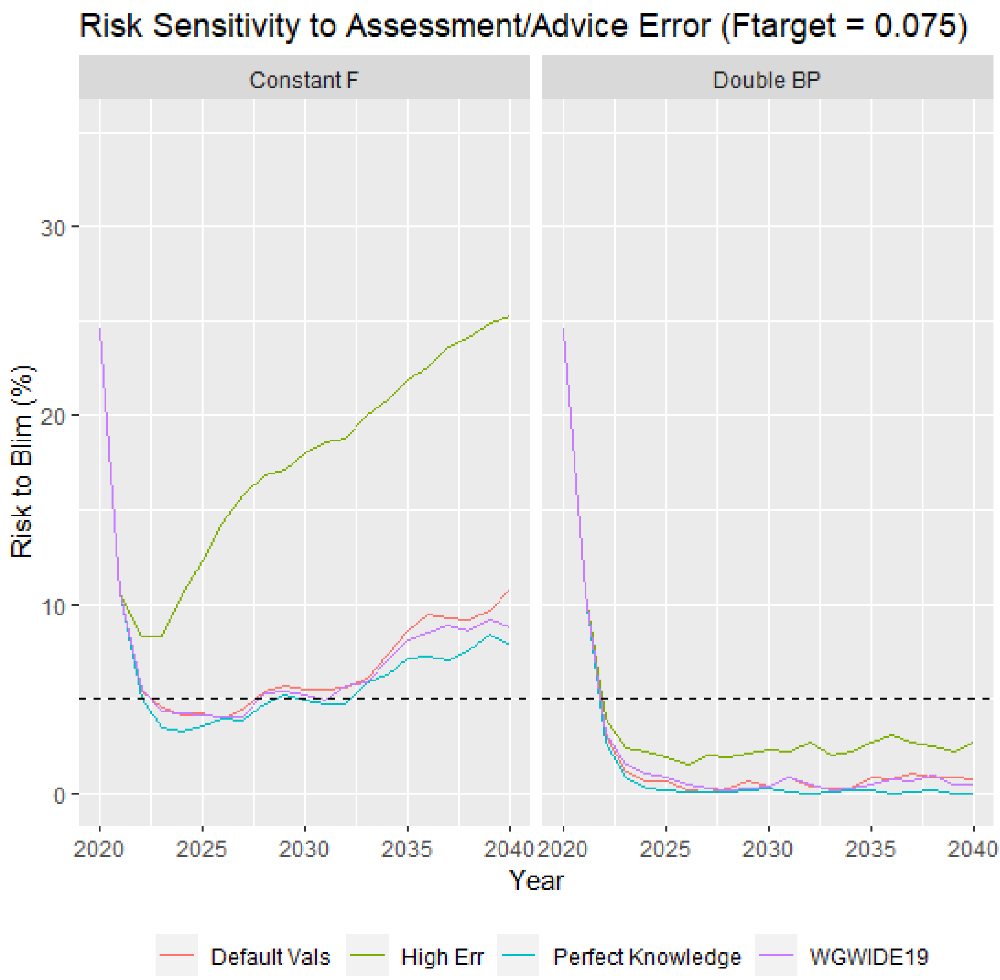


Figure 2.22 Comparison of the effect of varying assessment error parameters on the temporal development of risk to Blim for the constant F and double breakpoint HCR types (Ftarget = 0.075, TAC change limits of 20% when above Btrigger).

As expected, the removal of assessment error from the simulation leads to a lowering of the risk to B_{lim}. While remaining consistently low throughout the latter part of the projection period for the double breakpoint rule, the constant F strategy remains non precautionary in the absence of assessment error with risk increasing over time.

The inflated and highly correlated error assumption ("High Err") scenario is associated with the highest risk. For both HCRs, risk reduces over the initial years due to the relatively strong recruitment between 2014 and 2018 before rising, rapidly in the case of the constant F strategy.

In terms of the error parameter values based on historic performance ("WGWIDE19") and the default values, there is little to choose between the simulations with the double breakpoint rule in terms of risk. The difference between these scenarios is lies in the value of the autocorrelation parameter for the error applied to the fishing mortality. A more discernible difference can be seen with the constant F rule where the use of the default values (with a higher autocorrelation in F error) is associated with a slightly elevated risk.

2.10 Check if results are sensitive to the assessment used as the baseline. E.g. use 2020 assessment as baseline vs older version used in current simulations.

During WGWIDE 2020 an error had been detected in the 2019 SS assessment of horse mackerel. The error had to do with the length frequencies of catches. However, on closer inspections, it was found that those changes did not have a noticeable effect on the assessment outcomes. Results of the comparisons are shown in Figure 2.23.

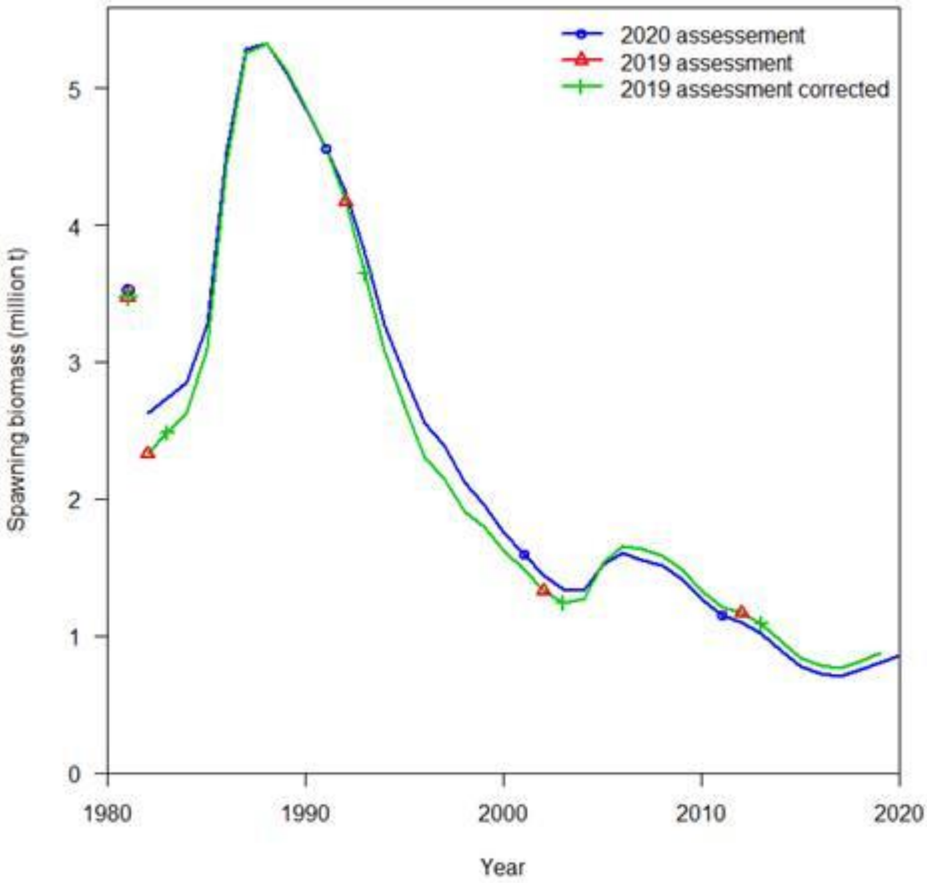


Figure 2.23 Comparison of SSB trajectories in the SS3 assessments 2019, 2019 corrected and 2020.

Simulations were carried out with the Stock synthesis assessments of WGWIDE 2019 and 2020 and with the exploratory SAM assessment of 2019 and 2020. Figure 2.24 illustrates the distribution of the 2018 SSB from the simulated populations relative to the appropriate Blim. This indicates that the 2020 assessments for both SS and SAM give a lower estimate of SSB.

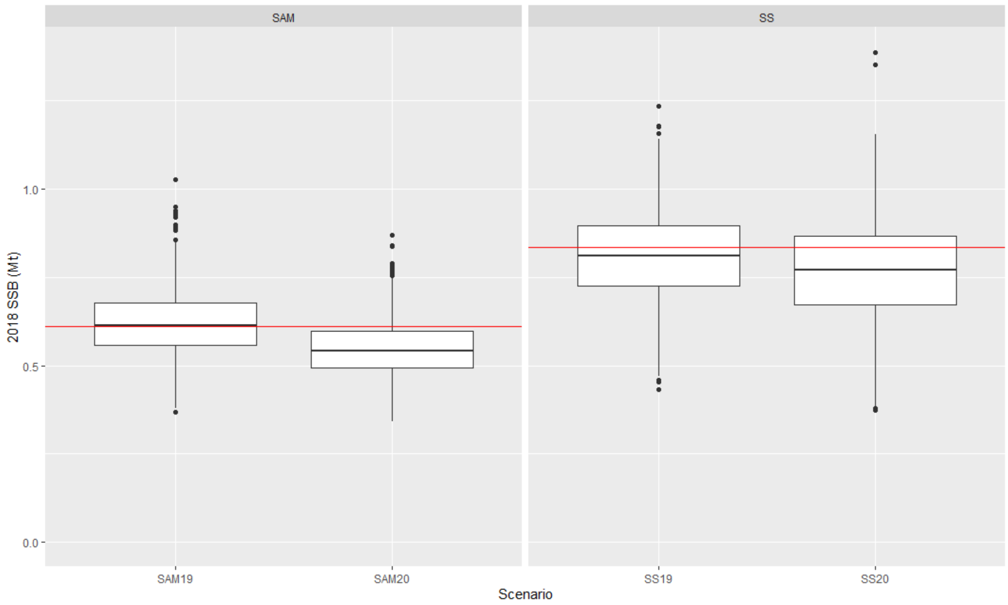
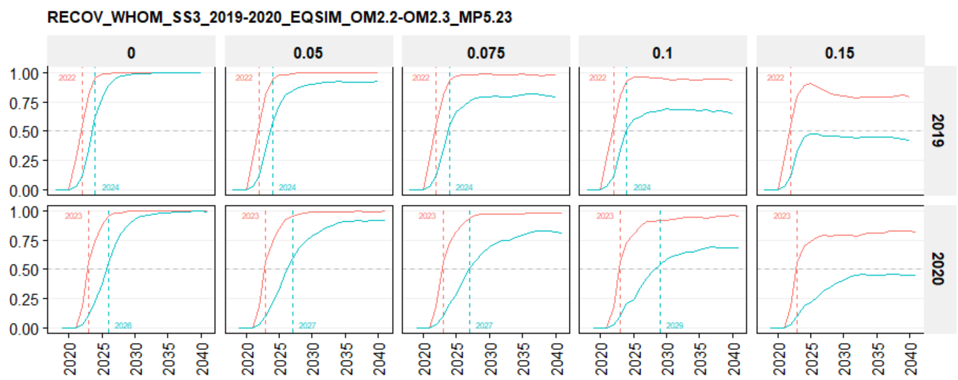


Figure 2.24 Comparison of 2018 distributions of SSB from SAM assessments 2019 and 2020 and SS assessments 2019 and 2020. The red line indicates the Blim for the relevant assessment method.

Comparison of detailed results for the SS assessments are in section 8.4.1. Comparison of detailed results for the SAM assessments are in section 8.4.2. Main results in terms of rebuilding potential to Blim and Bpa are summarized below (for the preferred double breakpoint scenario 5.23 with IAV). Using the 2020 assessment as the OM for the simulations leads to a delay in achieving rebuilding to both Blim and Bpa. For the target F of 0.075 (~Fmsy) with SS3, the rebuilding to Blim would be delayed to from 2022 to 2023 and the rebuilding to Bpa from 2024 to 2027. The main factor contributing to the delay is the lower estimates of recruitment in the most recent years.



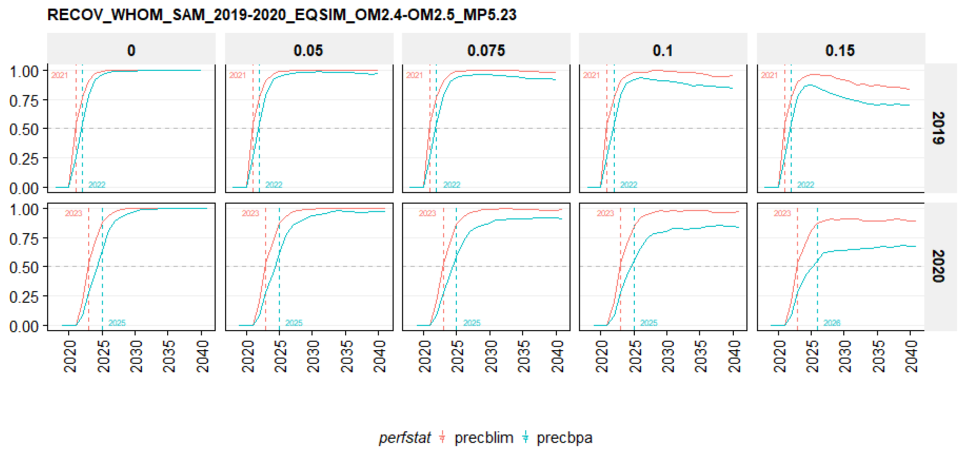


Figure 2.25 Comparison of the rebuilding potential of horse mackerel, using the 2019 and 2020 SS assessments (above) and the 2019 and 2020 SAM assessments (below). The red line indicates the probability of rebuilding to Blim and the green-blue line the probability of rebuilding to Bpa.

2.11 Evaluate the robustness of results to the perceived strong recruitments in the recent years. For example, is rebuilding performance acceptable if recent recruitments are set to recent average values instead of the current estimates?

Recent annual SS3 assessments estimate an increase in recruitment since 2014. With regard to early information on the size of incoming year classes, the assessment is principally informed by the IBTS recruitment index as they have yet to substantially appear in the catch data. The IBTS recruitment time series included in the assessment starts in 2003 and there is broad agreement between the index and subsequent catch-at-age with regard to the relative size of earlier year classes (2008 & 2012 strong, 2010 & 2011 weak). Since the introduction of the SS3 assessment in 2017, revisions in historic recruitment estimates have been relatively minor, save for the most recent years. However, to investigate the robustness of the management strategies to the magnitude of the recent recruitment, alternative scenarios with regard to recruitments from the 5 year classes between 2014 and 2018 were explored. This was achieved by adjusting the abundance at age in the initial year of the simulation for each replicate. The appropriate abundance was calculated on the basis of the alternative recruitment assumption and the replicate estimates of fishing mortality at age. A base case and three alternative recruitment assumptions were tested (see also Table 2.3):

1) Base case: use recruitment estimates directly from the assessment(s)
2) Annual recruitments for 2014-2018 are equivalent to one half of the baseline estimate (i.e. that from the final assessment)
3) Each of the 2014-2018 recruitments are assumed to be equivalent to the geometric mean of the time series from 2002-2013. Aside from 2008, this period is representative of the lowest recruitment observed.
4) Each of the 2014-2018 recruitments are assumed to be equivalent to the mean of the 5 lowest recruitments from the full time series.

Scenario/ Year class	Baseline	Scenario 1 (1/2 baseline)	Scenario 2 (GM 2002-2013)	Scenario 3 (Mean 5 lowest)
2014	4,004,110	2,002,055	1,620,516	1,001,051
2015	2,837,020	1,418,510	1,620,516	1,001,051
2016	3,263,620	1,631,810	1,620,516	1,001,051
2017	5,070,720	2,535,360	1,620,516	1,001,051
2018	2,887,740	1,443,870	1,620,516	1,001,051

Table 2.3 SS3 Alternative recruitment scenarios by year class (in thousands). Geometric mean of the full time series (ex 1982) is 2,572,649.

The risk to B_{lim} for each of these scenarios for the double-breakpoint HCR (including 20% TAC limitation when above the trigger) is shown in Figure 2.26 for 4 target fishing mortalities.

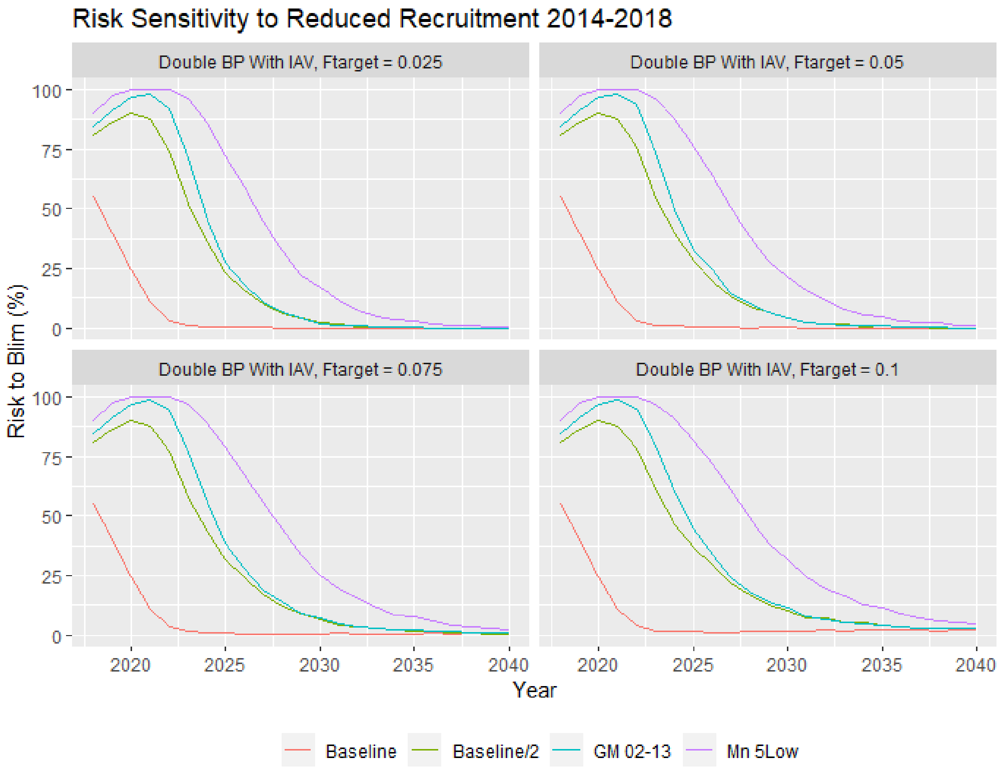


Figure 2.26 Risk to Blim over the projection period for the baseline and alternative recruitment scenarios.

The clear conclusion for the reduced recruitment regimes is that in the initial years of the projection, the risk to B_{lim} is substantially increased. The reduced recruitment assumptions have a significant effect on the initial SSB, as shown in Figure 2.27. The majority of replicates for each of the alternative scenarios are below B_{lim} (indicated by the red line)

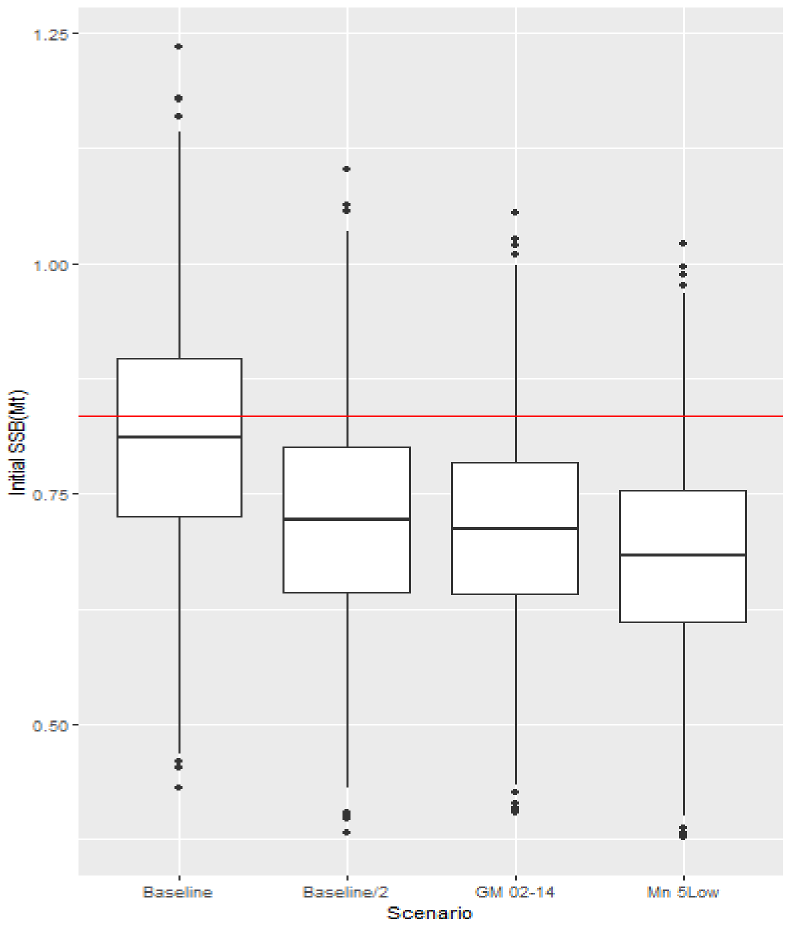


Figure 2.27 Range of SSB values for starting year for the 1000 replicates by recruitment scenario.

Risk to B_{lim} is increased in the initial year of the simulation as a greater proportion of the replicates are below B_{lim}. The effect of suppressing the recruitment for a period of 5 years leads to an initial further increase in risk in the first years of the simulation, during which a catch constraint applies for all replicates and management procedures (2021 is the first year for application of the HCR). However, risk starts to reduce rapidly within 5 years for the most pessimistic scenario falling below 50% in 2023/24 for the baseline/2 and GM 20-14 scenarios and by 2027 for the scenario with the lowest recruitment.

2.12 Evaluate the sensitivity of results to alternate interpretations of reference points, e.g., use reference points estimated with SS3 and apply to SAM simulation and vice versa.

The sensitivity of the analysis to the reference points used, was carried out by crossing reference points between SS3 and SAM, *i.e.* the SAM reference points were applied to the SS simulation and vice versa. Since the biomass reference for the SAM assessment are substantially lower than the biomass reference points for the SS assessment, it can be foreseen that the rebuilding capacity of the HCR will be mostly affected for the SAM assessment with the SS3 assessment based reference points. Detailed results are in section 8.4.5 and summarized in Figure 2.25. Using the appropriate reference point values or those derived from the alternative assessment does not affect the rebuilding year to

B_{lim} (although it is affected by assessment and assessment year). With regards to rebuilding to B_{pa}, the SS3 assessment with the appropriate reference points shows an increase in years when the stock may be rebuilt. The opposite (and logical) result is seen for the SAM assessment with longer rebuilding times associated with the alternative (SS3 based).

Figure 2.28 Rebuilding time to B_{pa} for double breakpoint HCR (including 20% interannual limits on TAC) for range of F_{target} values. The plot matrix row corresponds to the OM assessment model (SAM top, SS3 bottom) and columns to the reference point values employed in the calculation of the recovery metric (left column - default values *i.e.* those consistent with the underling OM assessment, right column – values from the alternative assessment).

2.13 Explain the importance of stock structure information in relation to the expected performance of the candidate management procedures. Describe steps that could be taken to reduce the effects of scale mismatch between stock distribution and stock assessment data collection.

For all assessments and management strategy evaluations a good understanding of the structure of the stock and the certainty in the origin of commercial catches are important. For a highly migratory species as horse mackerel, which are divided into three separate stocks in the North East Atlantic, this can potential be a problem for the validity and interpretation of the assessment and management strategy evaluation results.

The stock structure of horse mackerel has been research subject for several projects. During the period 2000-2003, HOMSIR, an EU funded project, focused on stock discrimination via genetic markers, life history parameters and morphometrics. This project was later followed up by a whole-genome sequencing project in collaboration between PELAC, Uppsala university and Edward Farrell from EDF Scientific Limited. In the latter project concerns about the appropriateness of the 9.a (Southern horse mackerel)/ 8.c (Western horse mackerel) border was raised as indications of Western horse mackerel distributed into the 9.a. division was found.

The uncertainty of splitting Western and southern horse mackerel along the 9.a/8.c border is not new as but has also led to a provision for inter-area flexibility where it is, for Portugal and Spain, allowed to fish up to 5% of their Western Horse mackerel quota (Subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a–c, and 7.e–k) in Division 9.a which is the distribution area of Southern Horse mackerel. Similar it is also allowed to fish 5% of the Southern Horse mackerel quota Division 9.a quota in the Western Horse mackerel area, Subarea 8.c.

In 2018, ICES was requested to evaluate the area flexibility and one conclusion reached was, that it was expected to be riskier for the southern horse mackerel stock than the western (ICES, 2018). This is because this stock since 1992 has been with a smaller TAC (19 000–41 000 tonnes) than the Wester horse mackerel stock (83 000-535 000 tonnes. Similarly, the reference point B_{lim} is also lower in the Southern stock (103 000 tonnes) compared to the Western stock (661 917 tonnes).

It is difficult to judge if conclusion from present management strategy evaluation would be changed if the border between the two stocks is erroneous or as a consequence of the 5% interarea flexibility. In general, the fact that the quota sizes for the southern horse mackerel are smaller than for the western horse mackerel might indicate that the concern for this specific management strategy evaluation should not be higher than for management strategy evaluation in general.

2.14 Provide the equations that indicate how each performance statistic is calculated, and add results for performance criteria that were described but not reported (e.g., criteria for achieving rebuild status).

Statistical Periods

Evaluation of a recovery/rebuilding plan has a necessary focus on the near term. For this reason, the short and medium periods are constrained to 5 years. The current period covers the time from the start of the simulation up to the present day. During the "current" statistical period, which is in the past, information is available on the catch taken and it would not be appropriate to apply a management rule. Catch constraints are therefore imposed based on the information available at the time of the appropriate assessment (WG catch estimates, short term forecast assumptions, published advice).

The 23-year simulation period is subdivided into 4 separate periods. For simulations based on an OM derived from the WGWIDE2020 assessment, years should be incremented by 1.

Statistical Period	Simulation years	period duration
Current	2018-2020	3 years
Short Term (ST)	2021-2025	5 years
Medium Term (MT)	2026-2030	5 years
Long Term (LT)	2031-2040	10 years

The catch constraints applied (and their basis) are described in the table below:

Simulation	WGWIDE19	WGWIDE20

YEAR	(START 2018)	(START 2019)
2018	101,682 (WG estimate)	NA
2019	110,381 (STF assumption)	124,947 (WG estimate)
2020	83,954 (advice)	69,527 (STF assumption)
2021	NA	81,376 (advice)

Risk to B_{pa}, B_{lim}

The annual risk of SSB being below a threshold value (such as B_{lim} or B_{pa}) is simply the proportion of the iterations with SSB below the threshold value for the year in question. When considering any particular statistical period, the associated risk is the maximum of the annual risks for that period (type 3 risk, ICES 2019).

Interannual variability in SSB, Yield

The annual variability in SSB or Yield is calculated as the mean percentage change over the previous year. When applied to a period, the median of the mean change over the period is reported.

Time to recovery

For each year of the projection period, the proportion of iterations that have been above a threshold for 3 consecutive years following a drop below the threshold for a period of one year or longer are reported. Below is an example for the calculation for rebuilding to B_{pa}. A similar procedure is implemented for B_{lim}. The threshold for rebuilding above a candidate biobass is considered to have been achieved if the proportion exceeds 50%.

```
#mark all those values below the threshold
t[t<OM$refPts$Bpa] <- 0
runningTot <- rep(NA,length(simYears))
names(runningTot) <- simYears
runningTot[simYears[1]] <- 1

#not very efficient but cycle through by simulation year
for (y in as.character(simYears[2]:simYears[length(simYears)])){
  runningTot[y] <- sum(apply(t[1:which(simYears==y),],MARGIN=2,function(x) (any(x==0))))
  t[y,!t[y,] == 0 & t[as.character(as.integer(y)-1),] == 0] <- 1
  t[y,!t[y,] == 0 & t[as.character(as.integer(y)-1),] == 1] <- 2
  t[y,!t[y,] == 0 & t[as.character(as.integer(y)-1),] == 2] <- 3
  t[y,!t[y,] == 0 & t[as.character(as.integer(y)-1),] == 3] <- 3
}

#proportion recovered is the proportion of iterations with value 3
propRec <- apply(t,MARGIN=1,function(x){sum(x==3)})/runningTot

Stats[["precBpa"]][["val"]] <- propRec

# Calculate first year rebuilt to Bpa
yearsRebuiltToBpa      <- propRec[propRec > rebuiltThreshold]
firstYearRebuiltToBpa <- min(an(attributes(yearsRebuiltToBpa)$names))

Stats[["firstYearRebuiltToBpa"]][["val"]] <- firstYearRebuiltToBpa
```

3 EqSim simulations

3.1 SS3-based operating model and EqSim simulations

3.1.1 Operating model (true biology)

3.1.1.1 Initial population vector

See section 2.8.1.

The simulation start date coincides with the terminal year of the assessment upon which the operating model is conditioned. To incorporate uncertainty in the various population vectors (abundance, stock and catch weight and fishery selectivity), the variance-covariance matrix from the appropriate assessment was used to compile a set of 1000 stock replicates for the purposes of projection.

Initially, 10,000 parameter sets were randomly drawn from the variance-covariance matrix using the R multivariate normal *mvrnorm* function from the MASS package. Any draws with parameter values outside of the bounds defined in the assessment configuration were rejected. A subset of 1000 replicates were then selected such that the median and 95% confidence intervals of the terminal year SSB match those from the stock assessment. The stock development for 10 random iterations and the final assessment are shown in Figure 2.13 while Figure 2.14 compares the distribution of SSB in the terminal assessment year from the 1000 replicates.

3.1.1.2 Recruitment

See section 2.2.

The recruitment time series is dominated by the extremely large event in the first data year, 1982. This single year class permitted the expansion of the fishery and sustained catches for over a decade, accounting for more than 50% of the total annual catch by abundance in some years. A second large recruitment occurred in 2001, although this was approximately 1/3 of the size of that of 1982. There are occasional extended periods (3-5 years) of relatively low recruitment during the time series with indications of autocorrelation. Recruitment in the most recent period has increased following the two lowest on record in 2010 and 2011.

Extensive investigations of the S-R relationship have been carried out during previous MSE exercises, the assessment benchmark of 2018 and 2019 ICES Interbenchmark Protocol on reference points (Campbell et al., 2015; ICES, 2017; ICES, 2019). Consideration has been given to both single and mixed model approaches although it was found that this led to some implausible model fits with the relative proportions of the individual models highly sensitive to the 1982 and 2001 data points. Investigations including the fitting to the complete SR dataset and also the excluding the large recruitments with consideration also given to separate modelling the occurrence of a spike.

The most recent examination was conducted during the 2019 Interbenchmark (ICES, 2019). This exercise concluded that the SR dataset should be restricted to data points from 1995 onwards, to exclude not only the very large 1982 recruitment event but also the subsequent years when the contribution of this single cohort to the total SSB (peaking at ~65% by weight) exceeded the next largest from 2001. Nothing comparable to the 1982 recruitment has been observed over the subsequent 40 years and the magnitude of the largest recruitments (*e.g.* 2001,2008) have been much reduced. There is no evident stock-recruit relationship and therefore the IBP characterised recruitment for the *EqSim* long term projections by a segmented regression with a breakpoint set at B_{lim}. For the purposes of these simulations, the same approach is adopted.

3.1.1.3 Weight at age

See section 2.2.

The *EqSim* simulation tool incorporates future variability in catch and stock weight via a random selection of historic weight at age data from a predefined period (usually the most recent 10 years but typically less if a trend is evident). As currently configured, the SS3 assessment upon which the OM is conditioned does not directly use historic catch and stock weight at age data. Length composition data and conditional age at length is used and a time invariant von Bertalanffy growth model is fit and the growth parameters estimated. Weight at age is derived from modelled length via a static length-weight relationship.

To ensure consistency between the historic and simulation periods, the *EqSim* approach of randomly selecting from the historic weight at age data has been modified. Instead, weight at age estimates from each of the 1000 model iterations upon which the simulation is conditioned are used to generate a time series of future weights from which the simulation can randomly select. Variability is incorporated within the time series by applying an error term derived from an analysis of the variability in the historic weight at age datasets. This error term is considered age specific (generally higher for the younger ages) and autocorrelated. The same error term is applied to both the stock and catch weight data series as historic stock and catch weight at age is autocorrelated, to be expected since the stock weights are simply derived from a subset of the catch weight information. The analysis is based on data from 2000 onwards to exclude the period of rapid change prior to this.

3.1.1.4 Natural mortality

See section 2.2.

Maturity and natural mortality during the assessment and simulation are assumed time invariant. In the assessment, natural mortality is assumed equal to 0.15 for all ages and this is maintained for the simulation. There is no additional data available to characterize the trends or uncertainty in natural mortality for horse mackerel.

3.1.1.5 Maturity

See section 2.2.

The maturity ogive has been unchanged since 1998.

3.1.1.6 Fishery selection at age

See section 2.2.

SS3 is configured with a single sex, single fishery fleet with selectivity set up as a double normal, forcing the selectivity to be asymptotic. In the current assessment configuration, the selectivity is assumed time invariant. However, in order to ensure that appropriate variability in selectivity in incorporated within the simulations, the MSE draws at random from the 1000 individual selection profiles available from the replicates for each year of the projection period.

Selection is most variable for the younger ages (1-6). The youngest age class has a very low selection, ages 7 and over are fully selected. Historic (assessment) and future selectivity for ages 0-5 are shown in Figure 2.20.

3.1.1.7 Confounding between variables / correlated processes

No specific arrangements have been included to account for correlated processes apart from using similar random draws for the same type of variables (e.g. catch weights and stock weights.

Auto-correlation in advice error, and thereby the error in perception of stock size, is included via the Fphi and SSBphi parameters estimated through the process indicated by WKMSYREF3 (ICES, 2014).

3.1.1.8 Ecosystem, biological and technical interactions

Ecosystem, biological and technical interactions are not taken into account (and are largely unknown)

3.1.2 Observation and estimation models

3.1.2.1 Shortcut approach

See section 2.5.

No assessment or forecast is conducted during the simulation *i.e.* this is a 'short-cut' MSE approach. The parameterisation of uncertainty in observed SSB and F are incorporated into the simulations following the methodology developed for *EqSim* whereby SSB and FBar outputs from historic assessments and short term forecasts are compared to the most recent assessment and described via a pair of parameters (CV and autocorrelation) for each of SSB and FBar. These parameters are then used to generate an error and derived "observed" values of SSB and Fbar for the management procedure. All simulation code is available on github

https://github.com/ices-eg/wk_WKREBUILD; folder EqSimWHM

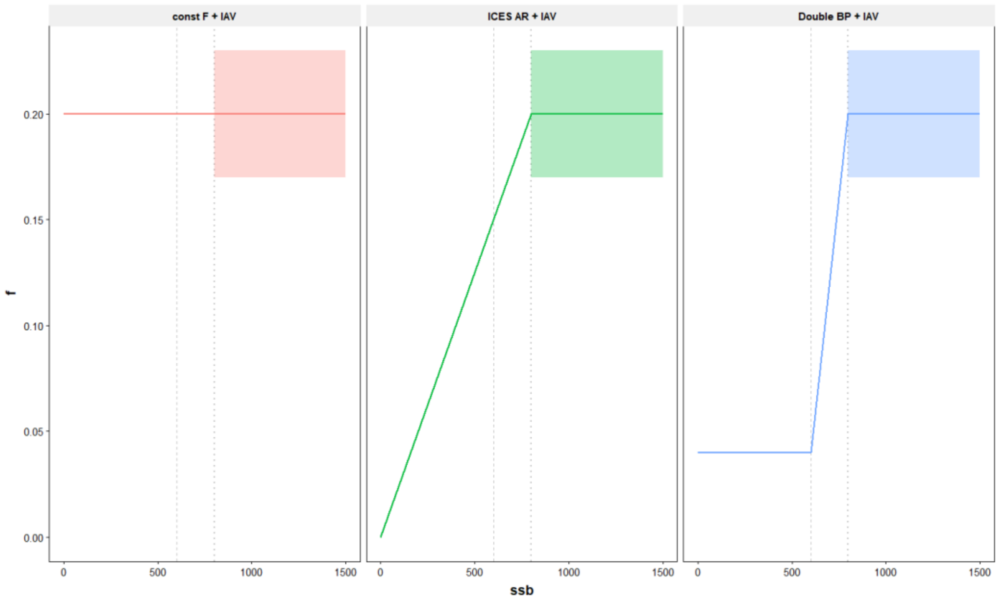
3.1.2.2 Harvest rules

Three different types of harvest control rules were evaluated:

- Constant F strategy: fixed Ftarget independent of biomass level
- ICES Advice Rule: breakpoint at Btrigger and straight decline in F to zero below Btrigger.
- Double Breakpoint rule: breakpoint at Btrigger and straight decline in F to 20% of Ftarget at Blim. Below Blim continued fishing at F = 0.2 * Ftarget.

For each of the HCRs, a number of different target fishing mortalities were explored (0.0, 0.05, 0.075, 0.1, 0.125, 0.15). No evaluation of different Btrigger values was carried out, so that all evaluations used MSY Btrigger as the trigger point. All HCRs where evaluated with three variants:

- Without any additional constraints
- With a minimum TAC of 50 kT
- With a maximum 20% inter-annual variation (IAV) in TAC, but only when the stock is above Btrigger)



3.1.2.3 Implementation model

No explicit implementation model has been implemented in the simulator. All TACs calculated via the HCR are assumed to be fully utilized. There is currently no information to underpin any scenarios on alternative implementation.

3.1.2.4 Validation

See section 2.9.

Trials were carried out with different values for Fcv and Fphi to explore how the results were sensitive to the variability and auto-correlation in advice error. A comparison was made between perfect knowledge and knowledge that has been influenced by different levels of uncertainty and autocorrelation.

3.1.2.5 Results

See section 2 for a collection of results relevant to the ToRs of this working document.

The collection of all simulation results are in the Annex to this working document (Section 8.

A data frame with all simulation results is available on the WKWHMRP sharepoint site:

https://community.ices.dk/ExpertGroups/wkwhmrp/2021%20Meeting%20Docs/06.%20Data/Results/WHOM_simulations_df.RData

3.1.2.6 Sensitivity analysis

3.1.2.6.1 Sensitivity to CV and Autocorrelation

See section 2.9

Due to the relatively short time series of assessment output available upon which to base a calculation of the error parameters and to investigate the sensitivity of the MSE results to the levels of assessment error, additional simulations were conducted. These simulations adopted the default parameter values used by *EqSim* analyses lacking stock specific values *e.g.* stocks with no or limited assessment history. Based on median values from a suite of stocks analysed during the development of the MSY framework, default values for Fcv and Fphi are 0.24 and 0.42 respectively. For this exercise, the same values were selected for the SSB equivalents. The default CV values are relatively similar to those calculated from the historic Western Horse Mackerel assessments. However, the AR1 autocorrelation phi parameter differs markedly, particularly for the fishing mortality error.

Additional sensitivity simulations were also conducted assuming "perfect knowledge" such that zero error is assumed in the assessment process and also with an arbitrary high correlated error (cv=0.6, phi=0.6). Although zero error is unfeasible in reality, this permits an insight into the relative contribution of this component of uncertainty compared with the other sources included in the MSE.

As expected, the removal of assessment error from the simulation leads to a lowering of the risk to B_{lim}. While remaining consistently low throughout the latter part of the projection period for the double breakpoint rule, the constant F strategy remains non precautionary in the absence of assessment error with risk increasing over time.

The inflated and highly correlated error assumption ("High Err") scenario is associated with the highest risk. For both HCRs, risk reduces over the initial years due to the relatively strong recruitment between 2014 and 2018 before rising, rapidly in the case of the constant F strategy.

In terms of the error parameter values based on historic performance ("WGWIDE19") and the default values, there is little to choose between the simulations with the double breakpoint rule in terms of risk. The difference between these scenarios is lies in the value of the autocorrelation parameter for the error applied to the fishing mortality. A more discernible difference can be seen with the constant F rule where the use of the default values (with a higher autocorrelation in F error) is associated with a slightly elevated risk.

3.1.2.6.2 Sensitivity to assessment used (2019, 2020)

See section 2.10

Simulations were carried out with the Stock synthesis assessments of WGWIDE 2019 and 2020 and with the exploratory SAM assessment of 2019 and 2020. Figure 2.24 illustrates the distribution of the 2018 SSB from the simulated populations relative to the appropriate Blim. This indicates that the 2020 assessments for both SS and SAM give a lower estimate of SSB.

Comparison of detailed results for the SS assessments are in section 8.4.1. Comparison of detailed results for the SAM assessments are in section 8.4.2. Main results in terms of rebuilding potential to Blim and Bpa for the preferred double breakpoint scenario 5.23 with IAV suggest that the 2020 assessment leads to a delay in achieving rebuilding to both Blim and Bpa. For the target F of 0.075 (~Fmsy) with SS3, the rebuilding to Blim would be delayed to from 2022 to 2023 and the rebuilding to Bpa from 2024 to 2027. The main factor contributing to the delay is the lower estimates of recruitment in the most recent years.

3.1.2.6.3 Sensitivity to recent strong(er) recruitment

See sections 2.11 and 8.4.3.

Recent annual SS3 assessments estimate an increase in recruitment since 2014. The increase in recruitment is also visible in the catch at age matrix (Figure 2.2).

With regard to early information on the size of incoming year classes, the assessment is principally informed by the IBTS recruitment index as they have yet to substantially appear in the catch data. The IBTS recruitment time series included in the assessment starts in 2003 and there is broad agreement between the index and subsequent catch-at-age with regard to the relative size of earlier year classes (2008 & 2012 strong, 2010 & 2011 weak). Since the introduction of the SS3 assessment in 2017, revisions in historic recruitment estimates have been relatively minor, save for the most recent years. However, to investigate the robustness of the management strategies to the magnitude of the recent recruitment, alternative scenarios with regard to recruitments from the 5 year classes between 2014 and 2018 were explored. This was achieved by adjusting the abundance at age in the initial year of the simulation for each replicate. The appropriate abundance was calculated on the basis of the alternative recruitment assumption and the replicate estimates of fishing mortality at age. A base case and three alternative recruitment assumptions were tested (see also Table 2.3):

1) Base case: use recruitment estimates directly from the assessment(s)
2) Annual recruitments for 2014-2018 are equivalent to one half of the baseline estimate (i.e. that from the final assessment)
3) Each of the 2014-2018 recruitments are assumed to be equivalent to the geometric mean of the time series from 2002-2013. Aside from 2008, this period is representative of the lowest recruitment observed.
4) Each of the 2014-2018 recruitments are assumed to be equivalent to the mean of the 5 lowest recruitments from the full time series.

The risk to B_{lim} for each of these scenarios for the double-breakpoint HCR (including 20% TAC limitation when above the trigger) is shown in Figure 2.26 for 4 target fishing mortalities. The clear conclusion for the reduced recruitment regimes is that in the initial years of the projection, the risk to B_{lim} is substantially increased. The reduced recruitment assumptions also have a significant effect on the initial SSB. The majority of replicates for each of the alternative scenarios are below B_{lim}.

Risk to B_{lim} is increased in the initial year of the simulation as a greater proportion of the replicates are below B_{lim}. The effect of suppressing the recruitment for a period of 5 years leads to an initial further increase in risk in the first years of the simulation, during which a catch constraint applies for all replicates and management procedures (2021 is the first year for application of the HCR). However, risk starts to reduce rapidly within 5 years for the most pessimistic scenario falling below 50% in 2023/24 for the baseline/2 and GM 20-14 scenarios and by 2027 for the scenario with the lowest recruitment.

3.1.2.6.4 Sensitivity to reference points used

See sections 2.12 and 8.4.5.

The sensitivity of the analysis to the reference points used, was carried out by crossing reference points between SS3 and SAM, i.e. the SAM reference points were applied to the SS simulation and vice versa. Since the biomass reference for the SAM assessment are substantially lower than the biomass reference points for the SS assessment, it can be foreseen that the rebuilding capacity of the HCR will be mostly affected for the SAM assessment with the SS reference points. Detailed results are in section 8.4.5 and summarized in Figure 2.25. Using the appropriate or wrong reference points does not affect the rebuilding year to Blim (although it is affected by assessement and assessment year). With regards to rebuilding to Bpa, the SS3 assessment with the appropriate reference points shows an increase in years when the stock may be rebuilt to Bpa. When using the wrong reference points, this then – logically – affecting the SAM 2020 assessment mostly.

3.2 SAM-based operating model and EqSim simulations

Unfortunately this part of the simulation has not been fully documented at this stage of the submission of the working document. However, the main assumptions and processes are similar to what was described for the SS3-based operating model.

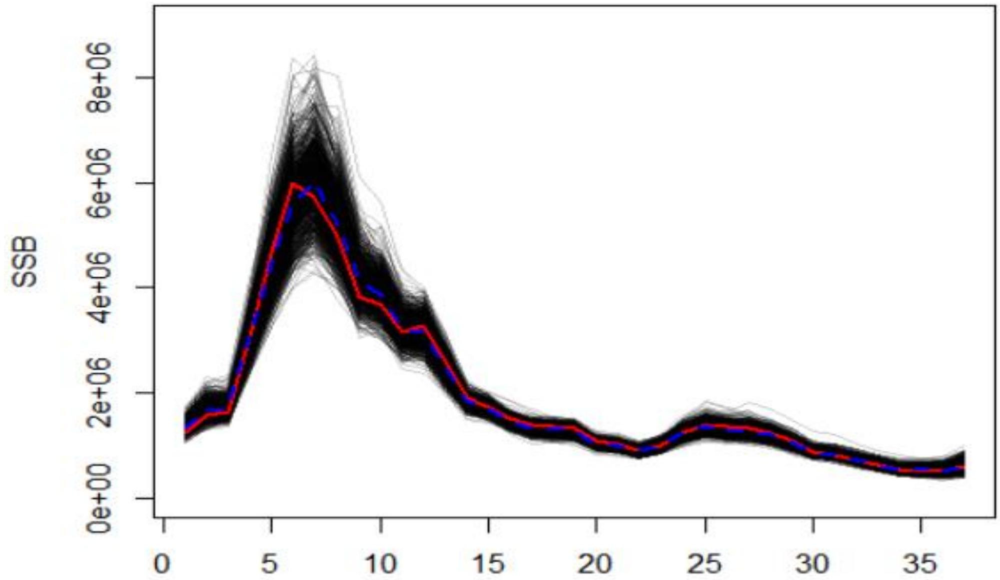
3.2.1 Operating model (true biology)

3.2.1.1 Initial population vector

Initial population vectors and other relevant variables (fishing mortality, recruitment) were derived by application of the stockassessment::simulate function and then reestimating the SAM model on the simulated data.

Because not all simulated data resulted in a converged assessment, a larger number of simulations were carried out (~1700) so that 1000 converged iterations could be used for the Eqsim simulations.

A comparison between the summary metrics SSB and Fbar for the individual iterations, the assessment (red) and the median (bluew) is shown in Figure 3.1.



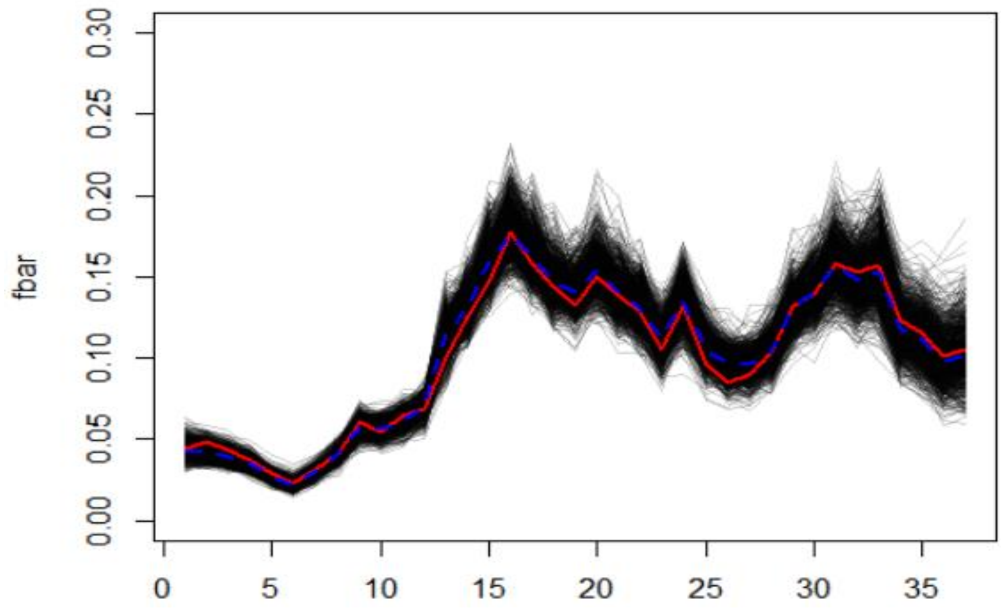


Figure 3.1 Comparison of the 1000 iterations of the SAM 2019 model. Red is the final assessment, blue is the median from iterations.

3.2.1.2 Recruitment

Same assumptions as in SS3 approach.

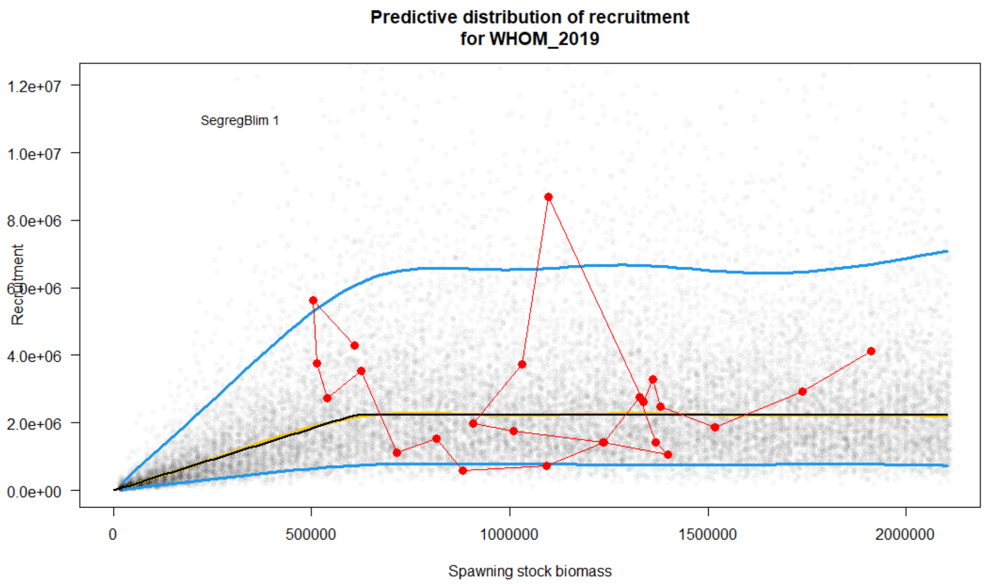


Figure 3.2 Stock and recruitment modelling for SAM 2019 assessment using segment regression through Blim.

3.2.1.3 Weight-at-length or age

3.2.1.4 Natural mortality

3.2.1.5 Maturity

3.2.1.6 Fishery selection at length or age

3.2.1.7 Confounding between variables / correlated processes

3.2.1.8 Ecosystem, biological and technical interactions

3.2.2 Observation and estimation models

3.2.2.1 Shortcut approach

3.2.2.2 Harvest rules

3.2.2.3 Implementation model

3.2.2.4 Validation

3.2.2.5 Results

3.2.2.6 Sensitivity analysis

3.2.2.6.1 Sensitivity to CV and Autocorrelation

3.2.2.6.2 Sensitivity to assessment used (2019, 2020)

3.2.2.6.3 Sensitivity to recent strong(er) recruitment

3.2.2.6.4 Sensitivity to reference points used

3.3 Comparison of SS3 and SAM based results

No additional exploration have been carried out into comparison between SS3 and SAM simulations.

4 SAM HCR forecasts

4.1 SAM based assessment and HCR forecasts

The SAM assessment is well documented (see Nielsen and Berg 2014). The WHOM assessment model scripts and results used in this study are publicly available at https://www.stockassessment.org/set-Stock.php?stock=WHOM2019. The HCR forecast in SAM that is used in this study was first developed and presented in Trijoulet et al. (Trijoulet et al., 2021). The code is available at https://github.com/vtrijoulet/SAM/tree/master2 and the main function is *forecast2()*. The forecast is a stochastic forecast that allows HCRs in the management procedure. It was set up to replicate as closely as possible the settings in the EqSim simulator.

4.1.1 Forecast settings

4.1.1.1 Initial population vector

Simulated 1000 replicates from the last year assessment estimates are used as starting values for the forecast. The replicates follow the same assumptions than the assessment model.

4.1.1.2 Recruitment

A segmented regression with inflection point fixed at Blim was fitted to the SAM 1995-2017 SSB and recruitment pairs to reproduce as closely as possible the method used in the EqSim simulator.

4.1.1.3 Weight-at-length or age

The average weight-at-age from the last 10 years of assessment are used in the forecast.

4.1.1.4 Natural mortality

The average natural mortality from the last 10 years of assessment are used in the forecast.

4.1.1.5 Maturity

The average maturity from the last 10 years of assessment are used in the forecast.

4.1.1.6 Fishery selection at length or age

The forecast starts with 1000 replicates of fishing selectivity. Over time in the forecast, the selection follows the one of the previous year so that if selectivity is changed by the TAC forcing in the intermediate year, the corresponding selectivity will follow.

4.1.2 Forecast model

4.1.2.1 Description of the approach

The stochastic forecast is run as a continuity of the assessment model so the process error assumptions from the assessment model are used in the forecast, e.g. process errors in fish numbers. Uncertainty in the perception of the current stock level is accounted for by simulating 1000 replicates of stock numbers at age and fishing mortality at age. These replicates are then used independently as starting values for the forecast. In the intermediate year, catch is fixed to the agreed TAC and a management strategy is applied thereafter.

4.1.2.2 Harvest rules

Similar harvest control rules than the ones for the EqSim simulator were used for the SAM forecast.

4.1.2.3 Implementation model

The HCR SAM forecast does not allow for implementation error in the HCR.

4.1.2.4 Sensitivity analysis

4.1.2.4.1 Sensitivity to assessment used (2019, 2020)

4.1.2.4.2 Sensitivity to recent strong(er) recruitment

4.1.2.4.3 Sensitivity to reference points used

5 Comparison of EqSim and SAM HCR forecast results

No additional comparison of EqSim and SAM HCR forecasts have been carried out.

6 Discussion

This working document has been produced by the group of experts that developed the original evaluation of a rebuilding plan of Western Horse mackerel on behalf of the Pelagic Advisory Council (Pastoors et al., 2020). The evaluation was submitted to ICES for a scientific review and on the basis of that review a number of additional questions for clarification were addressed to the group of experts. These questions for clarification are addressed within this working document. After the review of the PELAC report by WKWHMRP in February 2021, the reviewers and workshop participants concluded that various elements of the paper and rebuilding plan could not be evaluated due to lack of detail. Therefore, it was not possible to determine from the report under review if all criteria listed under the ToR of WKWHMRP were met. A list of additional questions were formulated that should be addressed. The expert group has subsequently worked on addressing the additional questions and they are presented in the current working document to WKWHMRP. Detailed results of the simulations are futher available on the WKHMRP sharepoint site.

6.1 General findings in this working document

Overall, the additional work consists of the following key elements:

6) Better documentation on the simulation approaches, the parameterization of the models, the results achieved and the metrics being used (specifically the rebuilding metrics)

   The expert group involved in this work acknowledges that the documentation of the simulation work on horse mackerel could have been better and more complete in the first place. On reworking the material, we also carried out a number of checks on the simulation code and improved the standard documentation from the model. We have now included the standard WKGMSE table for MSE documentation in section 2.4.

   We also streamlined the code on github[3], separating it into code for processing of input data, code for general settings and loading of relevant functions, code for simulations and code for presentation of the results.

   Presentation of input data has been included in sections 2.2, 2.3 and 2.8. Presentation of the results has been included in the Annex (section 8). In the presentation of the results, we now report all the relevant metrics and we connected the recent past (from 2000 onwards) to the simulated futures. New figures have been included to present the first year of rebuilding to certain biomass reference points under different assumptions or conditions.

7) Exploration of the sensitivity of the results to the assessment year (2019 or 2020)

   The exploration to the sensitivity to the assessment year has been explored both for the SS3 and SAM assessments. Results are summarized in section 2.10, 8.4.1 and 8.4.2. Using the 2020 assessment as the OM for the simulations leads to a delay in achieving rebuilding to both Blim and Bpa. For the target F of 0.075 (~Fmsy) with

[3] https://github.com/ices-eg/wk_WKREBUILD/tree/master/EqSimWHM

SS3, the rebuilding to Blim would be delayed to from 2022 to 2023 and the rebuilding to Bpa from 2024 to 2027. The main factor contributing to the delay is the lower estimates of recruitment in the most recent years.

8) Exploration of the sensitivity to the recent (relatively stronger) recruitment

The exploration to the sensitivity to the recent (stronger) recruitment has been explored both for the SS3 and SAM assessments. Results are summarized in sections 2.11, 8.4.3 and 8.4.4. A base case and three alternative recruitment assumptions were tested:

- Base case: use recruitment estimates directly from the assessment(s)
- Annual recruitments for 2014-2018 are equivalent to one half of the baseline estimate (i.e. that from the final assessment)
- Each of the 2014-2018 recruitments are assumed to be equivalent to the geometric mean of the time series from 2002-2013. Aside from 2008, this period is representative of the lowest recruitment observed.
- Each of the 2014-2018 recruitments are assumed to be equivalent to the mean of the 5 lowest recruitments from the full time series.

Under the reduced recruitment scenarios, risk to Blim is increased in the initial year of the simulation as a greater proportion of the replicates are below Blim. The effect of suppressing the recruitment for a period of 5 years leads to an initial further increase in risk in the first years of the simulation, during which a catch constraint applies for all replicates and management procedures. However, risk starts to reduce rapidly within 5 years for the most pessimistic scenario falling below 50% in 2023/24 for the baseline/2 and GM 20-14 scenarios and by 2027 for the scenario with the lowest recruitment. It can be concluded that lower recruitment scenarios are expected to lead to longer rebuilding periods, compared to the base case of the 2019 SS3 assessment. This is also confirmed by the exploration of using the 2020 assessment as the basis for the operating model instead of the 2019 assessment.

Under the preferred scenario by PELAC (Double BP with IAV constraint, target F=0.075), the first year of rebuilding in the base case is in 2022, while under the reduced recruitment scenarios it is between 2026 and 2030. Rebuilding to Bpa is expected around 2024 in the base case can would be delayed to 2030-2034 under the reduced recruitment regimes. When the 2020 assessment is used for the OM, rebuilding to Blim would be delayed to 2023 and to Bpa to 2027.

9) Exploration of the sensitivity to the uncertainty and autocorrelation in advice error

The exploration to the sensitivity to the uncertainty and autocorrelataion in advice error has been explored for the SS3 assessment only. Results are summarized in section 2.9. In the focus group report (Pastoors et al., 2020), a review of the methods to generate cv and autocorrelation in advice error was presented. This concluded that the method of estimating advice error has some potential drawbacks. A similar conclusion has recently been reached in the North Sea herring assessment (HAWG 2021, forthcoming).

To explore the sensitivity of the results to the uncertainty (cv) and autocorrelation (phi), additional sensitivity simulations were conducted assuming "perfect knowledge" such that zero error is assumed in the assessment process and also with an arbitrary high correlated error (cv=0.6, phi=0.6). As expected, the removal of assessment error from the simulation leads to a lowering of the risk to Blim. The inflated and highly correlated error assumption ("High Err") scenario is associated with the highest risk. For both HCRs, risk reduces over the initial years due to the relatively strong recruitment between 2014 and 2018. In the PELAC select double BP rule, the difference between these scenarios is relatively small. A more discernible difference can be seen with the constant F rule where the use of the default values (with a higher autocorrelation in F error) is associated with an increase in risk to Blim.

10) Exploration of the sensitivity to the reference points used.

The exploration to the sensitivity to the reference points used has been explored for both the SS3 and SAM assessments. Results are summarized in sections 2.12 and 8.4.5. The biomass reference points in the SAM assessment are lower than the biomass reference points in the SS3 assessment. This is due to the overall lower SSB and higher F in the SAM assessment compared to SS3. As could be expected, the application of the SS3 reference points to the SAM assessment lead to a substantial longer period of rebuilding that is required (because the bar is set higher) and vice versa for the SS3 assessment with SAM reference points.

6.2 Form of rebuilding plan in the light of WKREBUILD

WKREBUILD (2020) made a list of recommendations for rebuilding plan guidelines that will be further refined in the future as cases studies are developed. In this context, we have noted below the recommendation made by WKREBUILD in 2020 and added comments on how this work fits into the recommendations or can help refining them.

Table 6.1. Recommendations for guidelines from WKREBUILD (2020) and comments from the rebuilding plan evaluation carried out for WHOM.

Criteria	Recommendations from WKREBUILD	Comments from work in WKWHMRP
Rebuilding targets	Defining rebuilding biomass and fishing targets is critical to the evaluation of rebuilding plans and should be clearly defined at the beginning of the evaluation. While ICES is currently using Blim as the limit reference point, MSY Btrigger as the trigger reference point, and FMSY as the target fishing mortality; other targets could also be considered in a rebuilding plan if relevant.	A number of different target fishing mortalities were explored (0.0, 0.05, 0.075, 0.1, 0.125, 0.15). Simulations used Blim as the limit reference point and MSY Btrigger as the trigger reference point. Evaluations of different values for the biomass reference points were not evaluated. Risk was evaluated with Blim as target biomass.
Reference points	WKREBUILD raised some concerns regarding the estimation of reference points in ICES. If used in the evaluation of the rebuilding plan, reference points must be suitable, i.e. consistent with the cur-rent productivity of the stock at low SSB and current environmental conditions. Evaluation of rebuilding targets that may differ from current reference points, may be necessary for depleted stocks when those were estimated including periods of high productivity and optimistic stock-recruitment relationships.	FMSY was estimated for both simulation methods (SAM and EqSim) using recruitment pairs from 1995-2017. The reference points were therefore consistent with the simulation method used and taking recruitment from 1995 excluded the early years of very high productivity. FMSY is therefore consistent with the current productivity of the stock.

Time frame leading to a rebuilding plan	Development of rebuilding plan options should be initiated, when not already in the management plan, as soon as the median SSB of a stock is estimated below MSY Btrigger at the beginning of the advice year and the forecast based on the ICES rule (F from the slope) does not reverse the decline in SSB at the end of the forecast year. The effect of retrospective patterns on the possible future forecast of stock biomass should be taken into account when this is deter-mined. Implementation of a rebuilding plan, when not already in the management plan, should be advised if the median SSB of a stock is estimated below Blim at the beginning of the advice year and the forecast based on the ICES rule (F from the slope) does not allow the stock to get above Blim at the end of the forecast year.	Work on the rebuilding plan evaluation was initiated in 2019 when the stock was estimated between MSY Btrigger and Blim (2018 advice). After re-evaluation of reference point at the 2019 inter-benchmark, the stock was estimated below MSY Btrigger and close to Blim. Since then (2020 assessment), the stock is still above but close to Blim. Therefore, the development of the rebuilding plan was initiated when the stock was below MSY Btrigger at the beginning of the advice year and was still below it at the end of the forecast year (2019 advice) as recommended by WKREBUILD. The expert group advise on the implementation of the rebuilding plan despite the fact that the stock is currently close to but not below Blim, because there are concerns regarding retrospective patterns in the model that might bring the stock below Blim in recent future.
Time frame leading out from a rebuilding plan	The exit strategy should be embedded in the rebuilding plan. Leading out from the rebuilding plan too early or too late should be avoided. The exit strategy should preferably contain element on how to ensure a “smooth” transition between the rebuilding phase and the post-rebuilding phase (i.e., ICES advice rule or a LTMP) to reduce the risk of inversion of positive trends. The exit from a rebuilding plan should be robust to uncertainty in the estimation of the stock status to reduce the risk of falling back to a rebuilding phase soon after the exit. Robustness to uncertainty could include setting a certain probability of SSB being above re-building reference points, being above rebuilding targets for a number of consecutive years, a consistent positive trend in SSB, evidences of a strong year class confirmed by independent observations (i.e., survey and commercial fishery) and through time. Maintaining F below FMSY for a sufficient time (at least one generation) then smoothly transitioning to FMSY could also be a possible strategy to exit a rebuilding plan.	The rebuilding plan may be superseded by a long-term strategy for the stock when, according to ICES, the spawning stock biomass is assessed to have been above Bpa for three consecutive years.
Time frame for the evaluation of a rebuilding plan	The evaluation period represents the time window between TMIN and TMAX which is used to assess the level of rebuilding achieved by alternative rebuilding strategies. TMIN is defined as the time taken for the stock to rebuild with zero fishing to above Blim, or the agreed rebuilding target with 95% probability, or other level of probability depending on the state of depletion of the stock.	Tmin could be derived from the zero catch options in the simulations. Tmin may be dependend on the assumptions used during the simulations, e.g. the base year for the simulations and the recruitment assumptions.

	TMAX, defined as the maximum amount of time for rebuilding the stock, is usually specified by managers/requesters but could be ex-pressed as x* TMIN with x > 1. WKREBUILD was not able to conclude on a value for x in the estimation of TMAX. x=2 is often used in other jurisdic-tions.	
Checking the pro-gress of the rebuilding plan	Re-evaluation of the rebuilding plan may be necessary if the stock trajectory is out-side the range of expected performance relative to timelines of the rebuilding plan or if other exceptional circumstances arise such as unexpected data or a new under-standing of the stock. The new rebuilding plan evaluation will need to adapt to the new data or findings. A re-evaluation of the rebuilding targets or objectives may also be necessary.	Checks on the stock response to the re-building plan if accepted will be done an-nually with the update assessment. If the stock patterns deviate from the predictions, the rebuilding plan will be re-evaluated. If changes in data occur when the rebuilding plan is in place, the HCR will be also re-evaluated given the assessment results given the new data.
Probability of achiev-ing rebuild-ing	The default probability for rebuilding above the target is 95% but for certain stocks a lower probability may be more relevant in the short- to medium-term de-pending on the nature of the fishery and socio-economic considerations. This would be notably relevant for short-lived stocks with high recruitment variability that are estimated to be below Blim with a probability larger than 5% even if un-fished	The probability to fall below Blim was esti-mated every year of the forecast. Rebuilding to Blim or Bpa was assumed to have occurred when three consecutive years above the reference points were achieved with a probability of at least 50%.
Harvest rules in re-building phase	Several harvest rules should be evaluated during a rebuilding plan. These should be compared against the zero catch scenario and the ICES advice rule.	These was respected. A total of 3 HCRs were compared and all runs included a F=0 scenario.
Evaluation tools	Rebuilding plans necessitate a prompt management response. Evaluation tools should be available when the evaluation starts. Multiple tools already exist to evaluate re-building plans. Rebuilding plan evalua-tion should use tools that have been re-viewed or validated.	The evaluation used two existing tools (EqSim simulator and SAM HCR forecast) that were further developed for the exer-cise.
Uncertainty considera-tions	Alternative operating models should be evaluated to account for stock specific un-certainties. Typical uncertainties to con-sider in the rebuilding plan context are uncertainties in stock productivity (e.g. recruitment), in the assessment model (e.g. stock perception, bias such as retro-spective patterns) and in implementation error.	The main method (EqSim) considered un-certainty in: - Initial population size - Stock & catch weights - Selectivity at age - Recruitment - Assessment Model The SAM HCR forecast considered uncer-tainty in: - Current perception of the stock (simu-lated starting values for stock and fishing mortality taken from the as-sessment results) - Uncertainty in recruitment (seg-mented regression with fixed inflec-tion point at Blim and variance in stock-recruitment relationship) - Process errors in stock survival (same as in the assessment)

Special considera-tions	The context of the rebuilding plan may be framed based on mixed stocks, mixed fisheries and socio-economic objectives.	[more to come on stock structure implica-tions]

7 References

Campbell, A., De Oliveira, J. A. A., Kelly, C., and Roel, B. 2015. Report on Western Horse mackerel Management Strategy Evaluations (MSE). Report for Pelagic Fishing Industry. .

ICES. 2014. Report of the Joint ICES-MYFISH Workshop to consider the basis for FMSY ranges for all stocks (WKMSYREF3), 17–21 November 2014, Charlottenlund, Denmark. ICES C.M. 2014 / ACOM:64.

ICES. 2016. Report of the Workshop on Blue Whiting Long Term Management Strategy Evaluation (WKBWMS), 30 August 2016, Copenhagen. ICES C.M. 2016 / ACOM: 53.

ICES. 2017. Report of the Benchmark Workshop on Widely Distributed Stocks (WKWIDE). ICES C.M. 2017 / ACOM:36.

ICES. 2018. EU request on an interarea flexibility between divisions 8.c and 9.a for horse mackerel. ICES Special Request Advice, Bay of Biscay ecoregion, Published 17 December 2018, sr.2018.36.

ICES. 2019. Interbenchmark Protocol on reference points for western horse mackerel (Trachurus trachurus) in subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a-c,e-k (the Northeast Atlantic) (IBPWHM). ICES VOLUME 2 | ISSUE 95.

ICES. 2020a. Stock Annex: Horse mackerel (Trachurus trachurus) in Subarea 8 and divi-sions 2.a, 4.a, 5.b, 6.a, 7.a–c,e–k (the Northeast Atlantic).

ICES. 2020b. Working Group on Widely Distributed Stocks (WGWIDE 2020) Volume 2, Issue 82.

Pastoors, M. A., Campbell, A., Trijoulet, V., Skagen, D., Gras, M., Lambert, G., Sparrevohn, C. R., et al. 2020. Western Horse Mackerel Technical Focus Group on Harvest Control Rule Evaluations 2020. 43 pp pp.

Trijoulet, V., Berg, C. W., Sparrevohn, C. R., Nielsen, A., Pastoors, M. A., and Mosegaard, H. 2021. Can we fish on stocks that need rebuilding? Illustrating the trade-offs between stock conservation and fisheries considerations. bioRxiv preprint, 2021.2002.2025.432880.

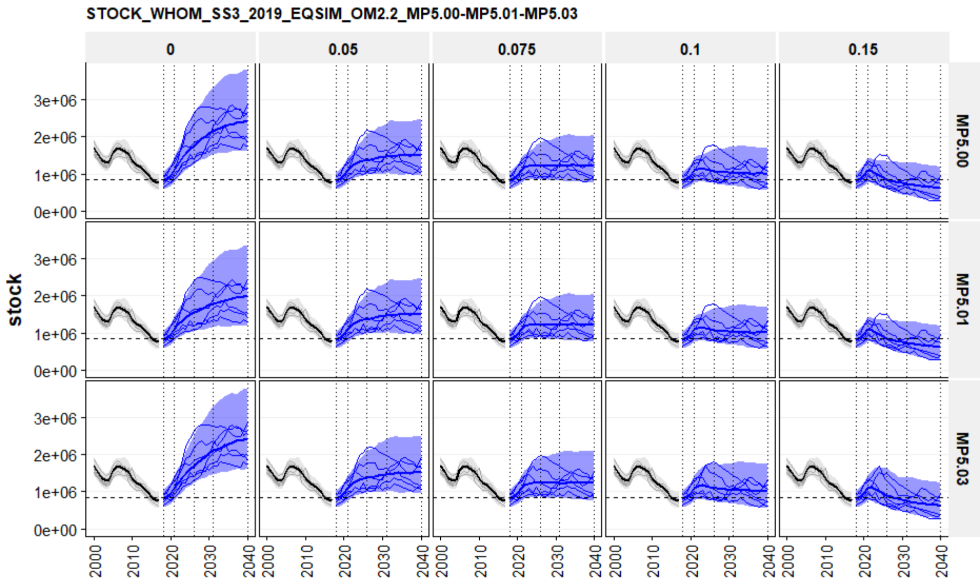
8 Annex: detailed simulation results

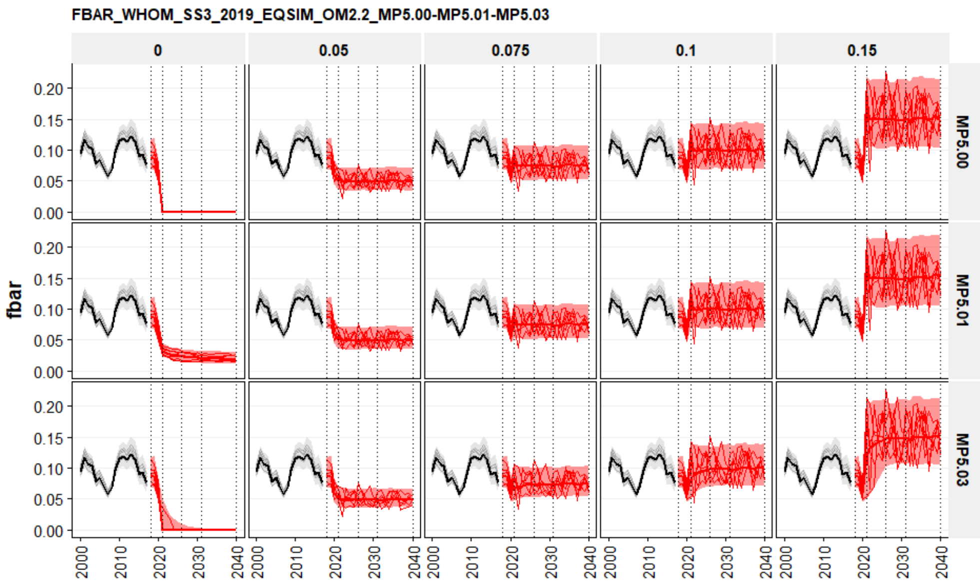
8.1 SS3 and EqSim

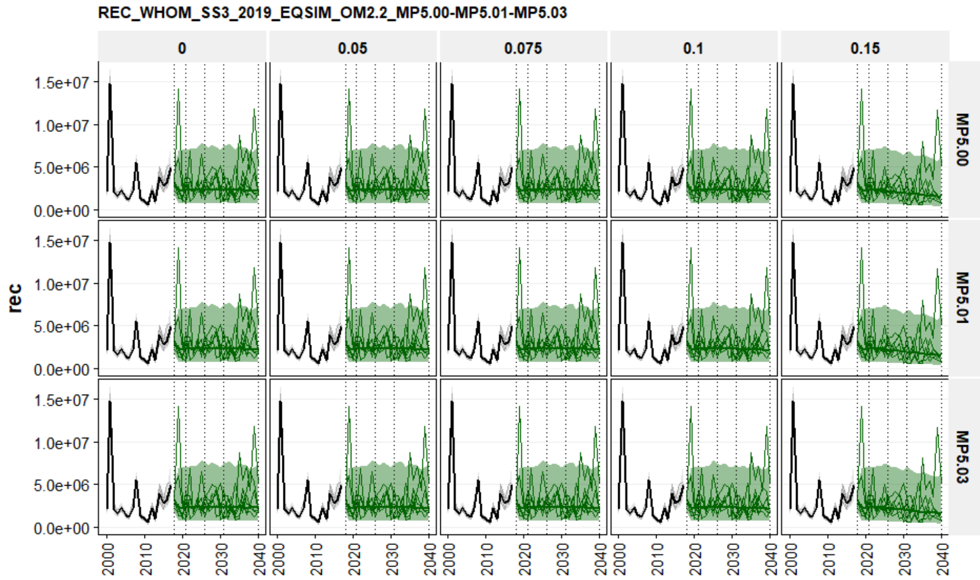
8.1.1 Constant F strategy

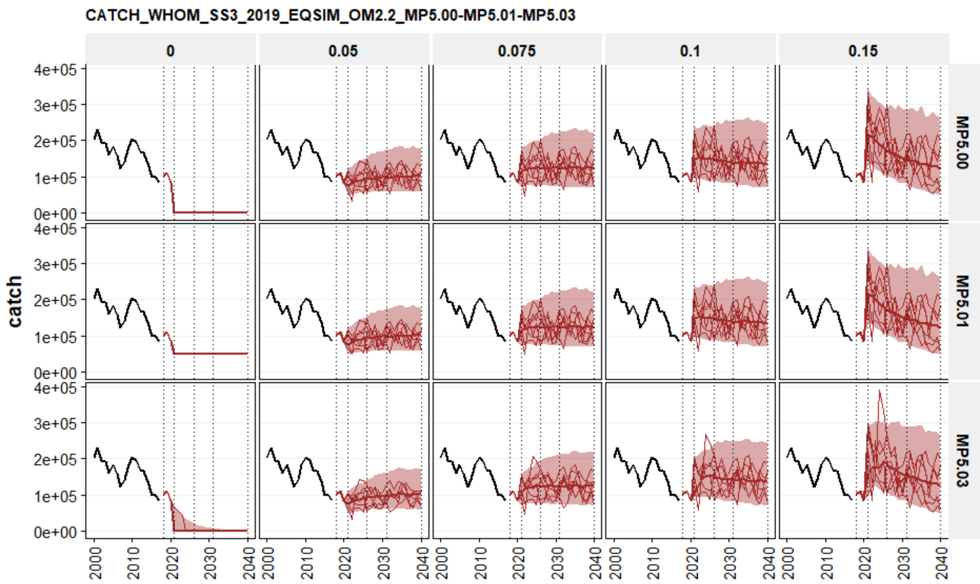
Simulation summary 2019 assessment, SS3, EqSim.

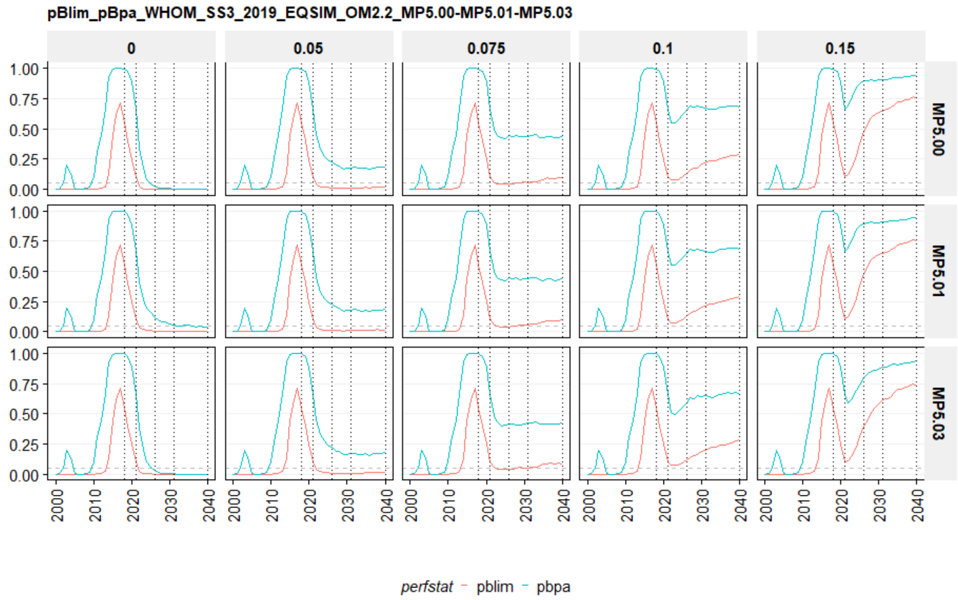
MP5.00 constant F; MP5.01 constant F with minimum TAC of 50kT; MP5.03 constant F with 20% IAV constraint above Btrigger.






REC_WHOM_SS3_2019_EQSIM_OM2.2_MP5.00-MP5.01-MP5.03
0
0.05
0.075
0.1
0.15
MP5.00
MP5.01
MP5.03
rec
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2000
2010
2020
2030
2040


CATCH_WHOM_SS3_2019_EQSIM_OM2.2_MP5.00-MP5.01-MP5.03
0
0.05
0.075
0.1
0.15
MP5.00
MP5.01
MP5.03
catch
4e+05
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040



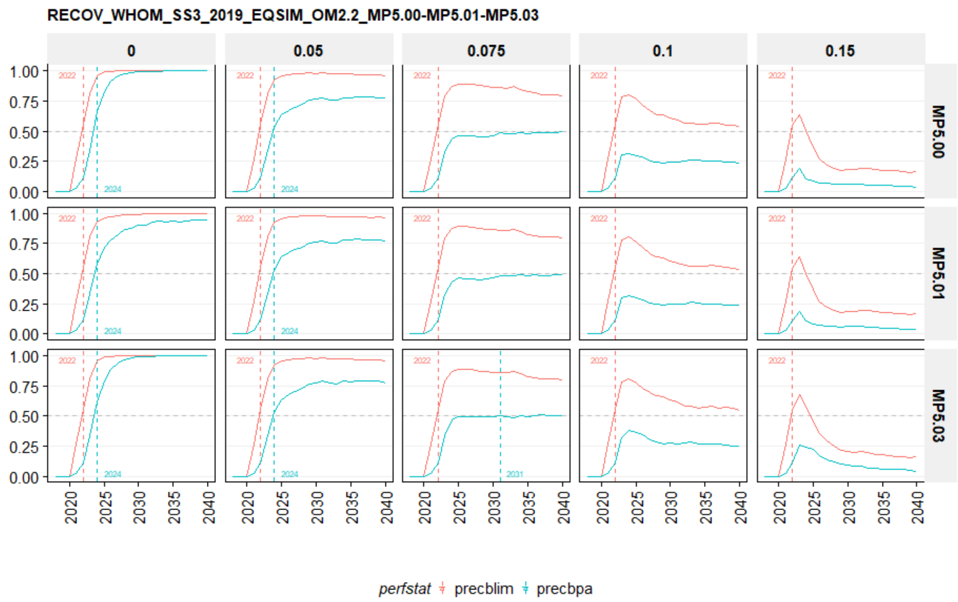
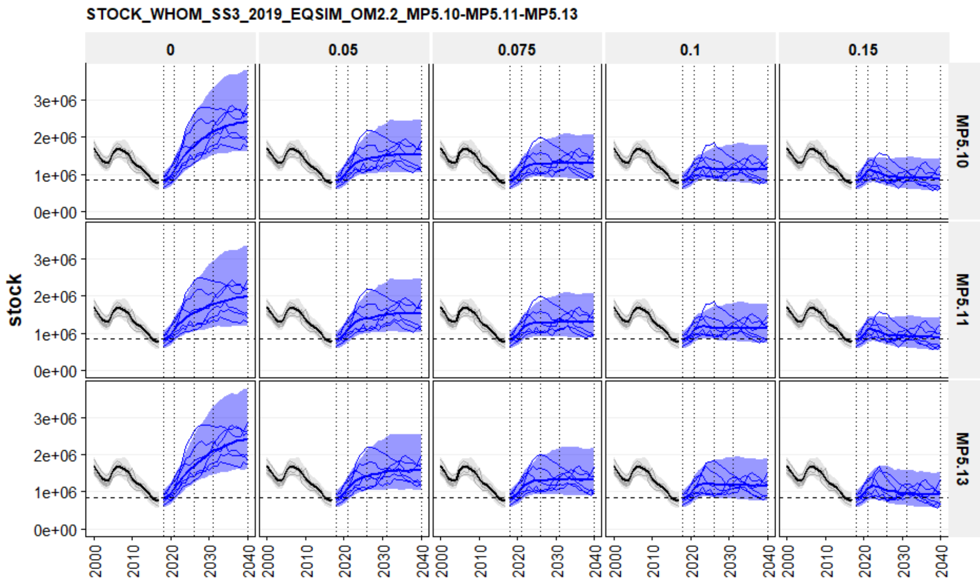


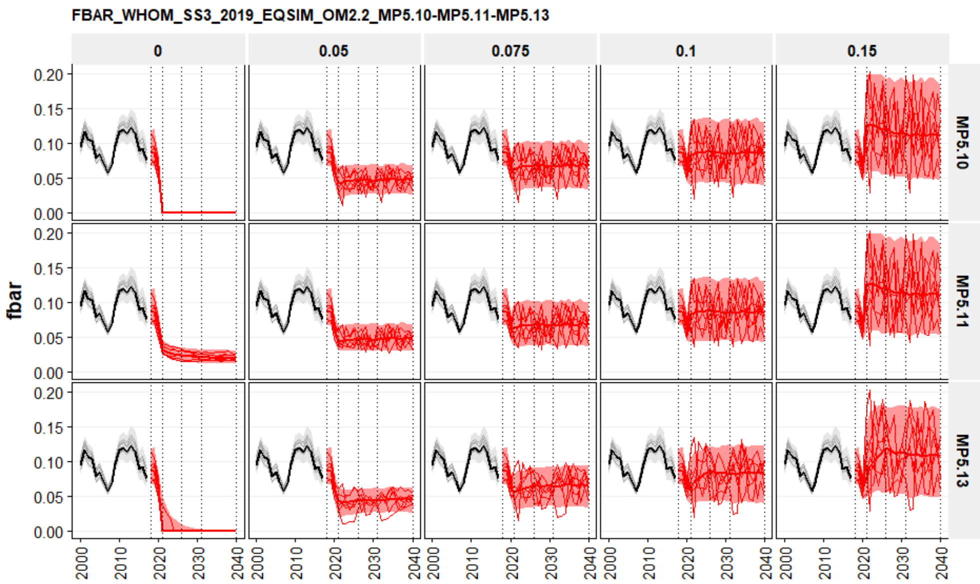
Figure 8.1 Simulation summary 2019 assessment, SS3, EqSim. MP5.00 constant F; MP5.01 constant F with minimum TAC of 50kT; MP5.03 constant F with 20% IAV constraint above Btrigger.

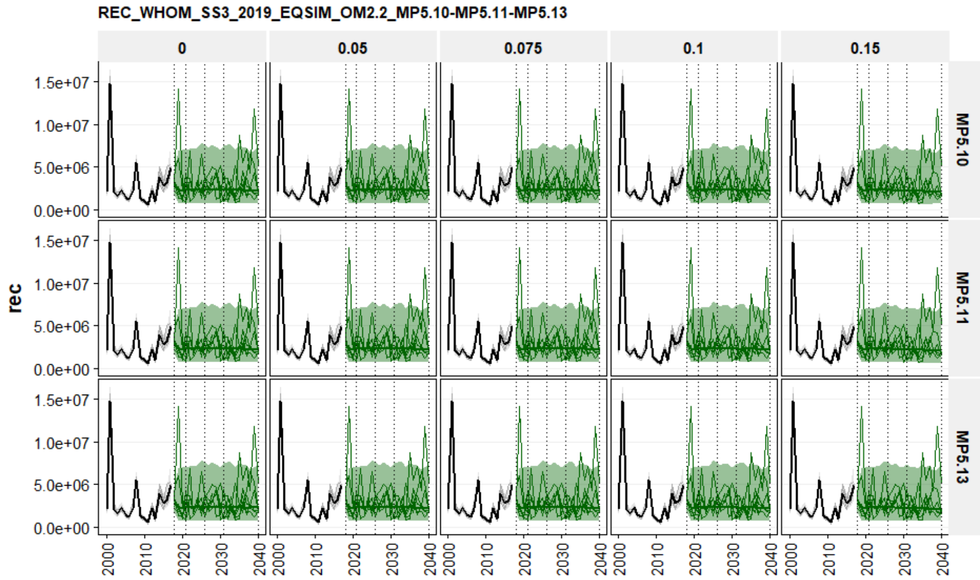
8.1.2 ICES advice rule

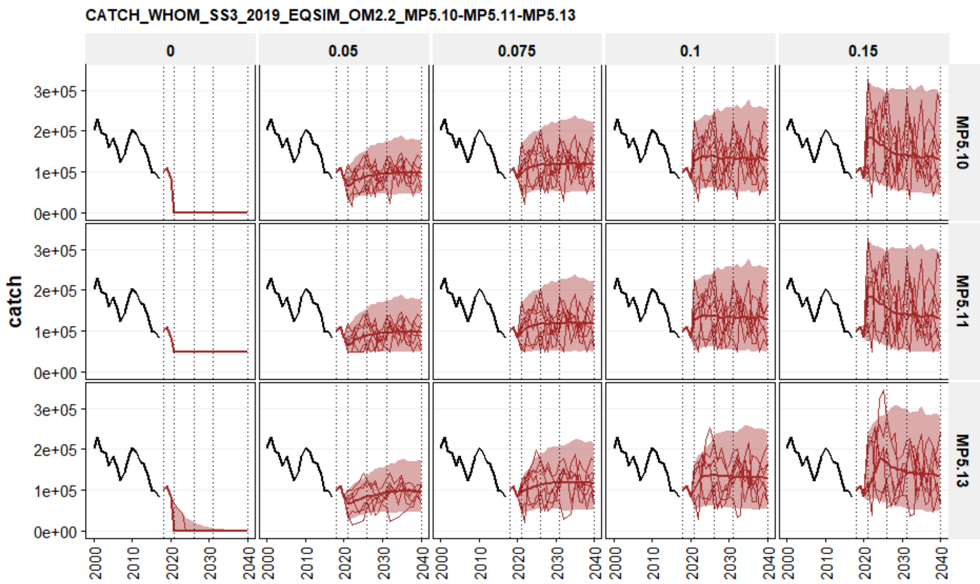
Simulation summary 2019 assessment, SS3, EqSim.

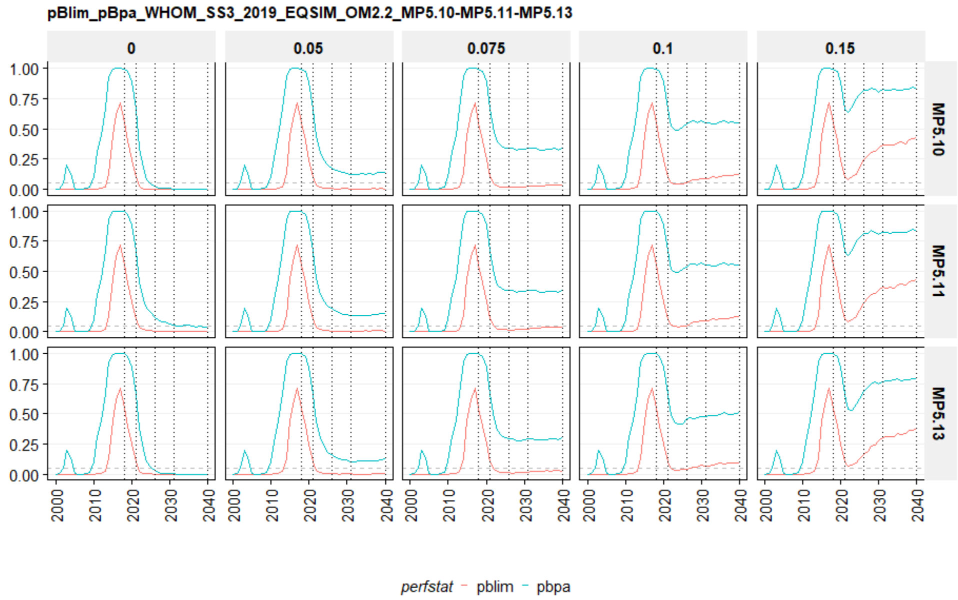
MP5.10 ICES AR; MP5.11 ICES AR with minimum TAC of 50kT; MP5.13 ICES AR with 20% IAV constraint above Btrigger.






REC_WHOM_SS3_2019_EQSIM_OM2.2_MP5.10-MP5.11-MP5.13
0
0.05
0.075
0.1
0.15
MP5.10
MP5.11
MP5.13
rec
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2000
2010
2020
2030
2040


CATCH_WHOM_SS3_2019_EQSIM_OM2.2_MP5.10-MP5.11-MP5.13
0
0.05
0.075
0.1
0.15
MP5.10
MP5.11
MP5.13
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040



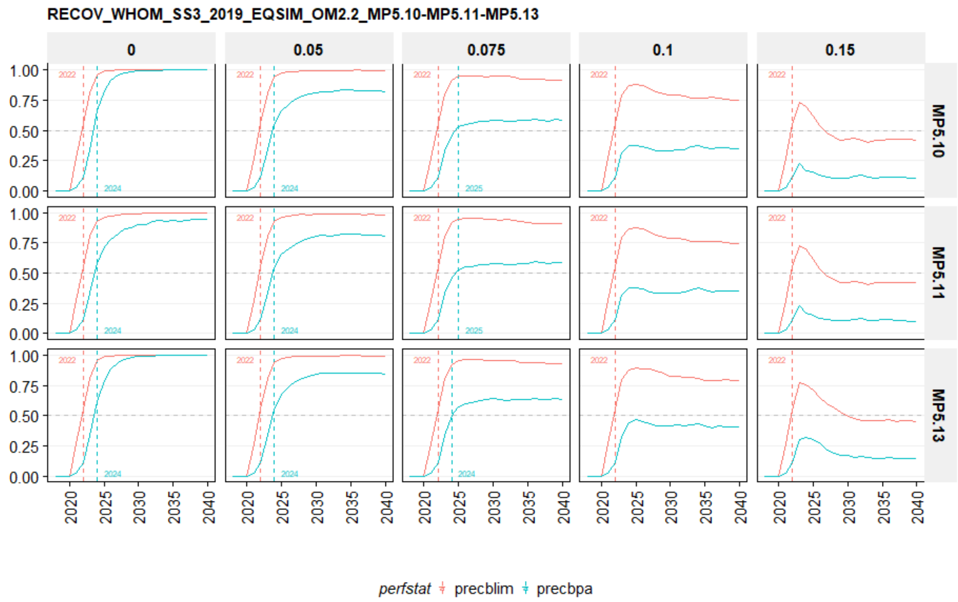
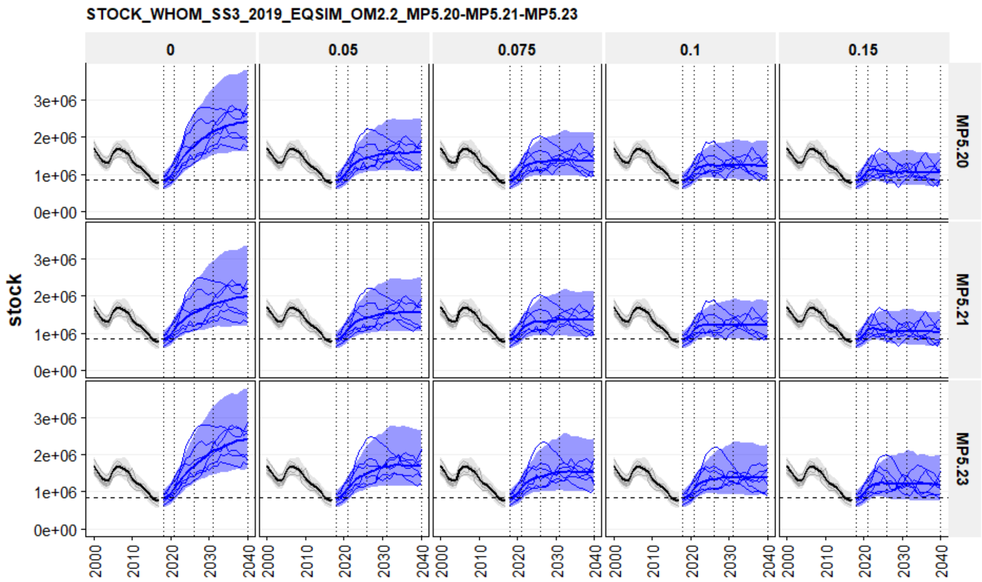


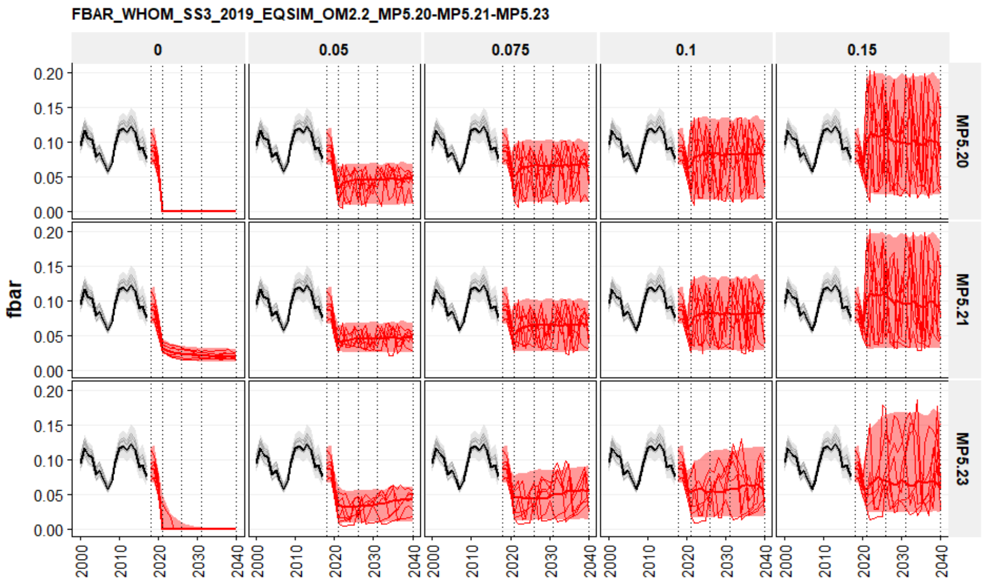
Figure 8.2 Simulation summary 2019 assessment, SS3, EqSim. MP5.10 ICES AR; MP5.11 ICES AR with minimum TAC of 50kT; MP5.13 ICES AR with 20% IAV constraint above Btrigger.

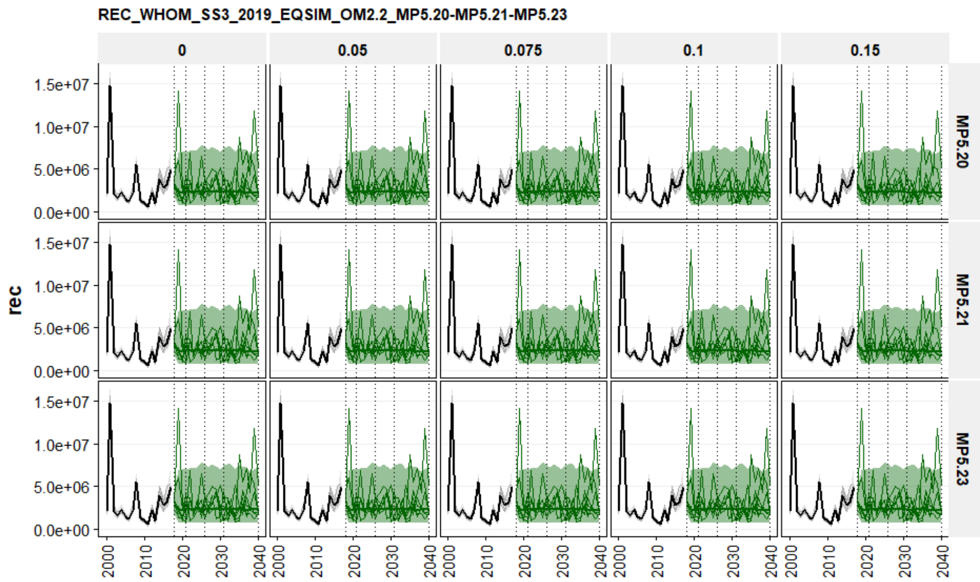
8.1.3 Double breakpoint rule

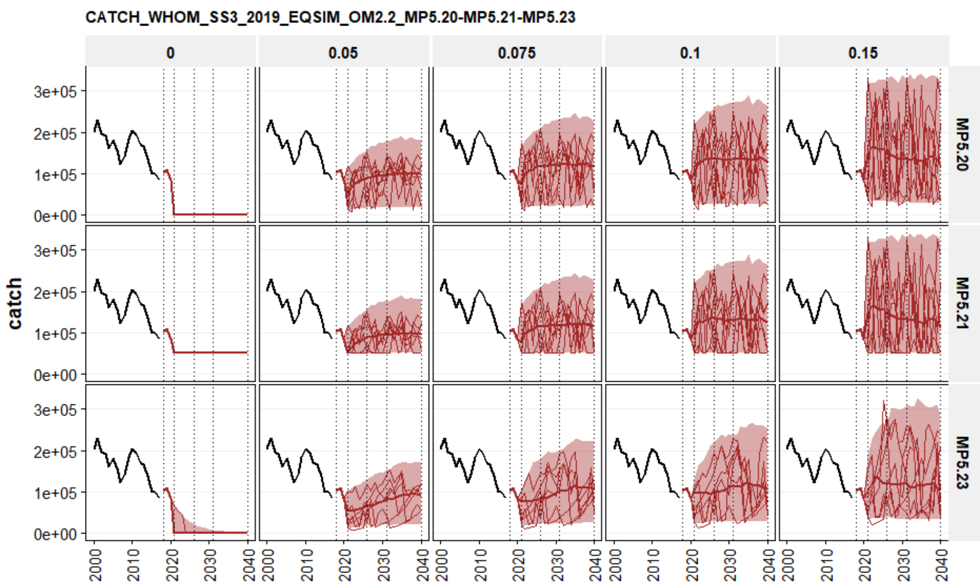
Simulation summary 2019 assessment, SS3, EqSim.

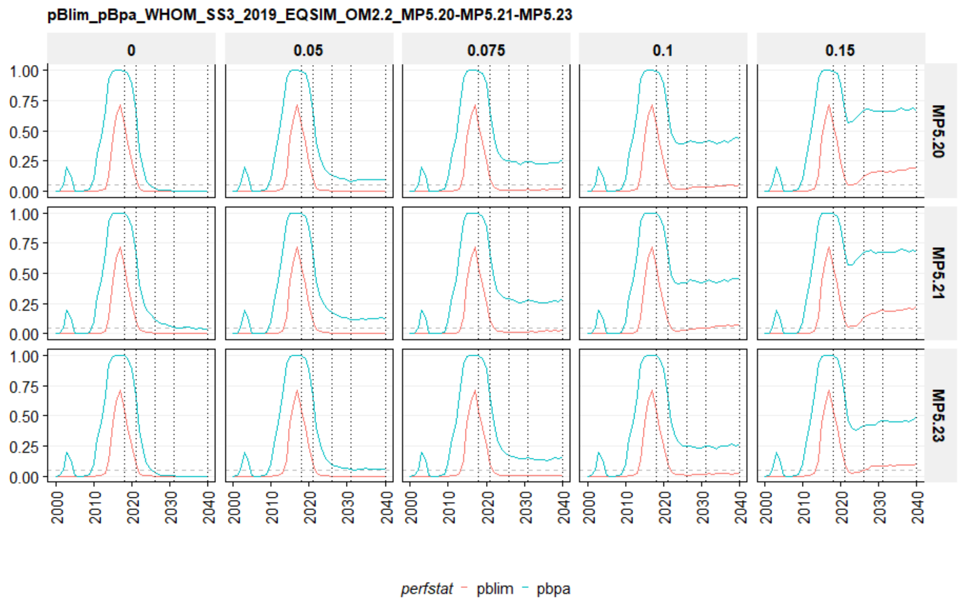
MP5.20 Double BP; MP5.11 Double BP with minimum TAC of 50kT; MP5.13 Doubl BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.






REC_WHOM_SS3_2019_EQSIM_OM2.2_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
rec
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2000
2010
2020
2030
2040


CATCH_WHOM_SS3_2019_EQSIM_OM2.2_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040



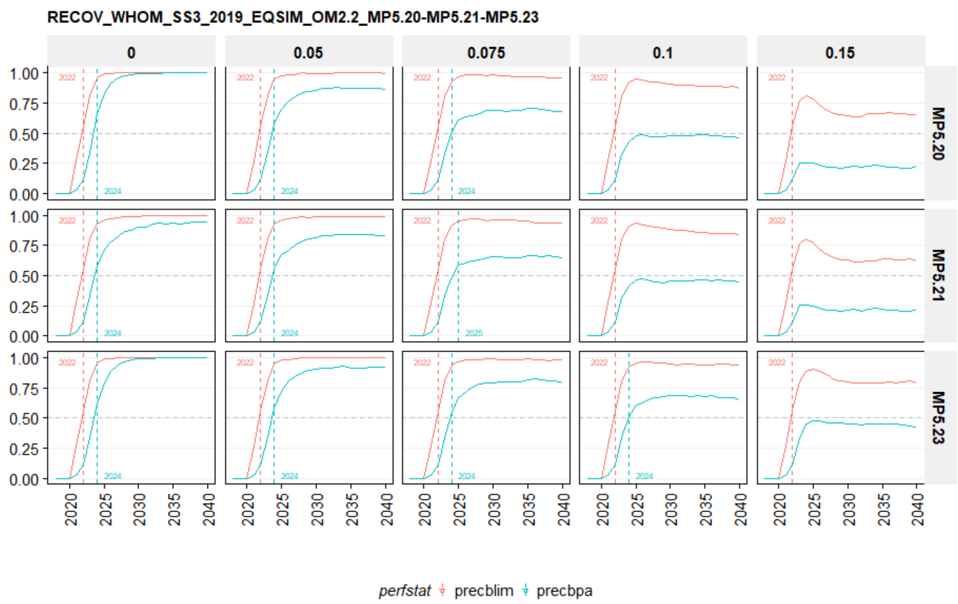


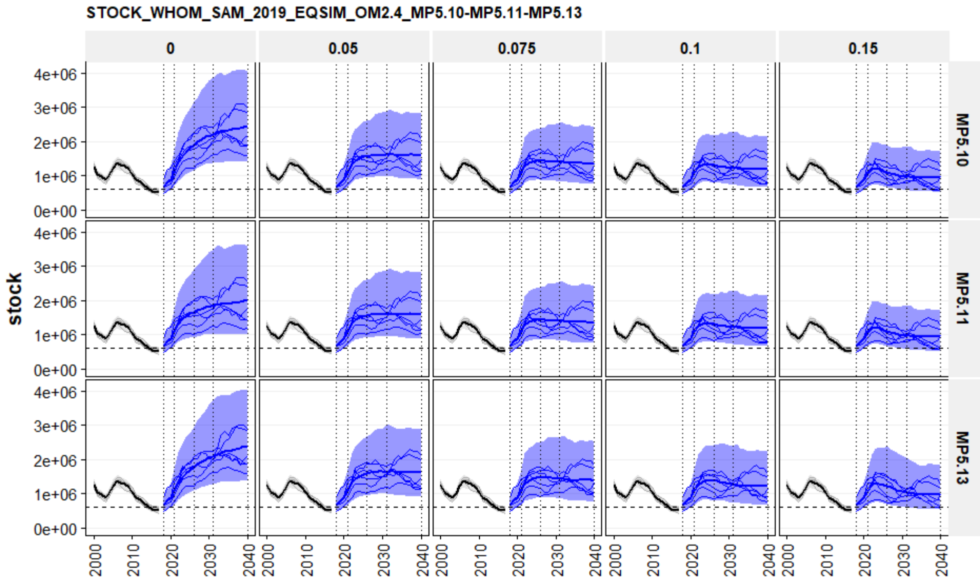
Figure 8.3 Simulation summary 2019 assessment, SS3, EqSim. MP5.20 Double BP; MP5.21 Double BP with minimum TAC of 50kT; MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

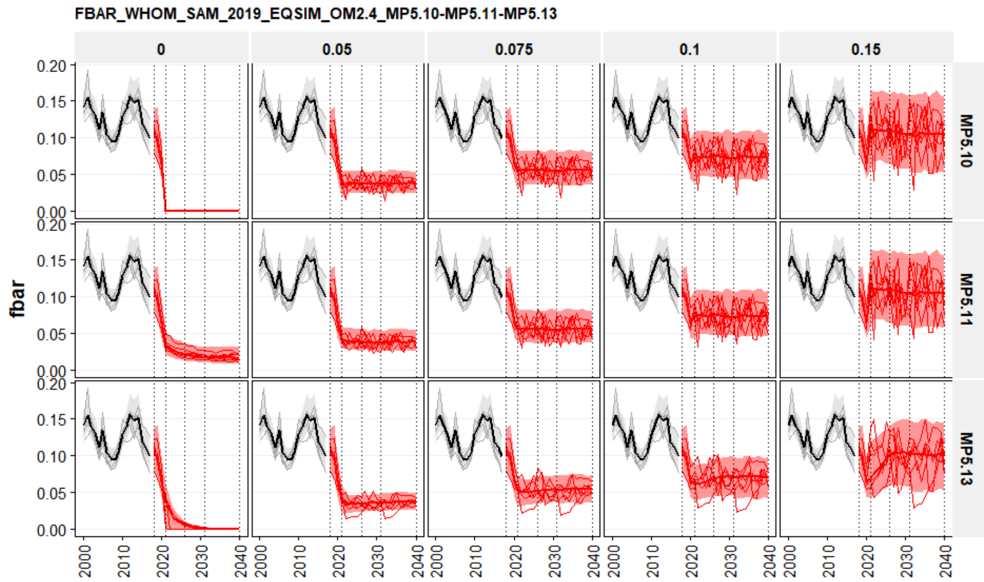
8.2 SAM assessments in EqSim simulator

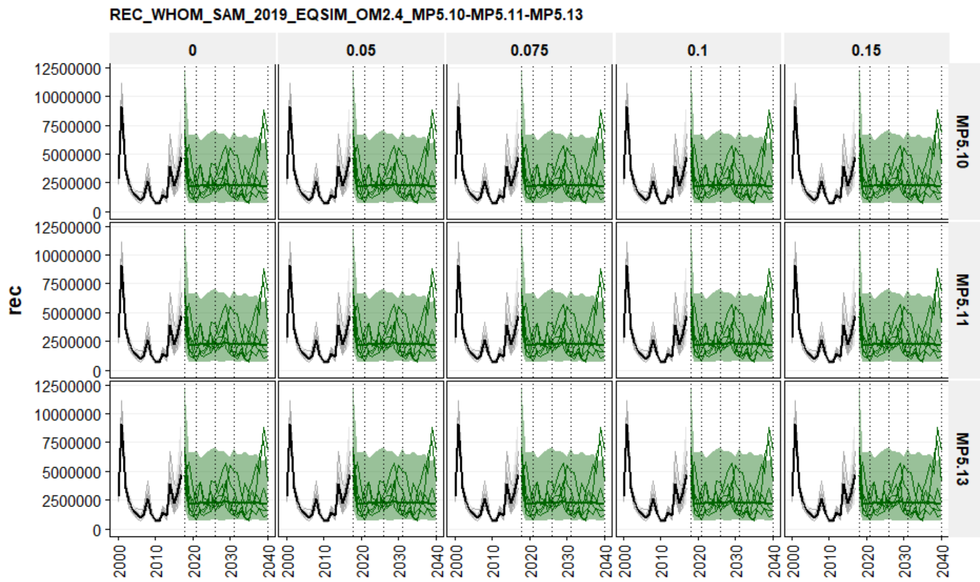
8.2.1 ICES advice rule

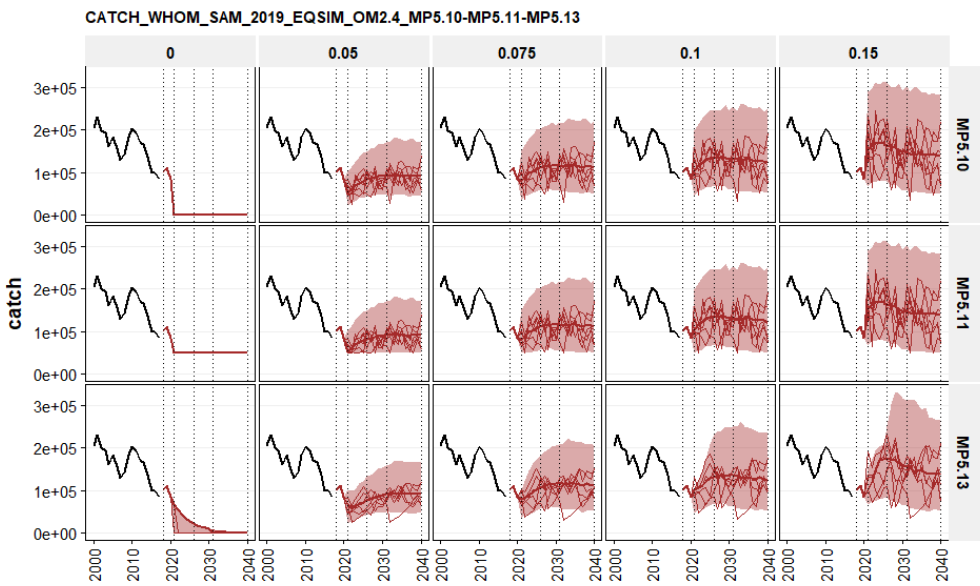
Simulation summary 2019 assessment, SAM, EqSim.

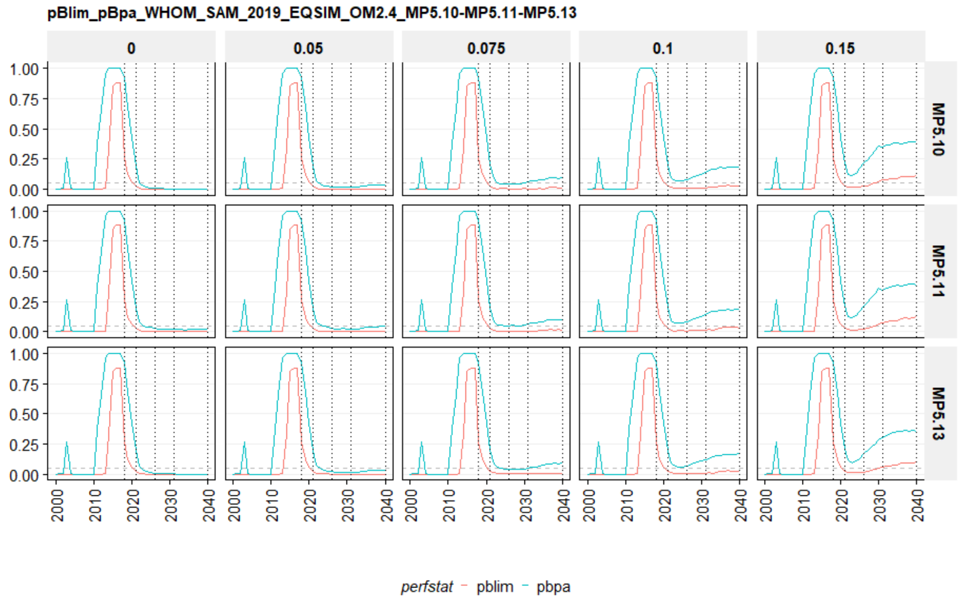
MP5.10 ICES AR; MP5.11 ICES AR with minimum TAC of 50kT; MP5.13 ICES AR with 20% IAV constraint above Btrigger.






REC_WHOM_SAM_2019_EQSIM_OM2.4_MP5.10-MP5.11-MP5.13
0
0.05
0.075
0.1
0.15
MP5.10
MP5.11
MP5.13
rec
12500000
10000000
7500000
5000000
2500000
0
2000
2010
2020
2030
2040


CATCH_WHOM_SAM_2019_EQSIM_OM2.4_MP5.10-MP5.11-MP5.13
0
0.05
0.075
0.1
0.15
MP5.10
MP5.11
MP5.13
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040



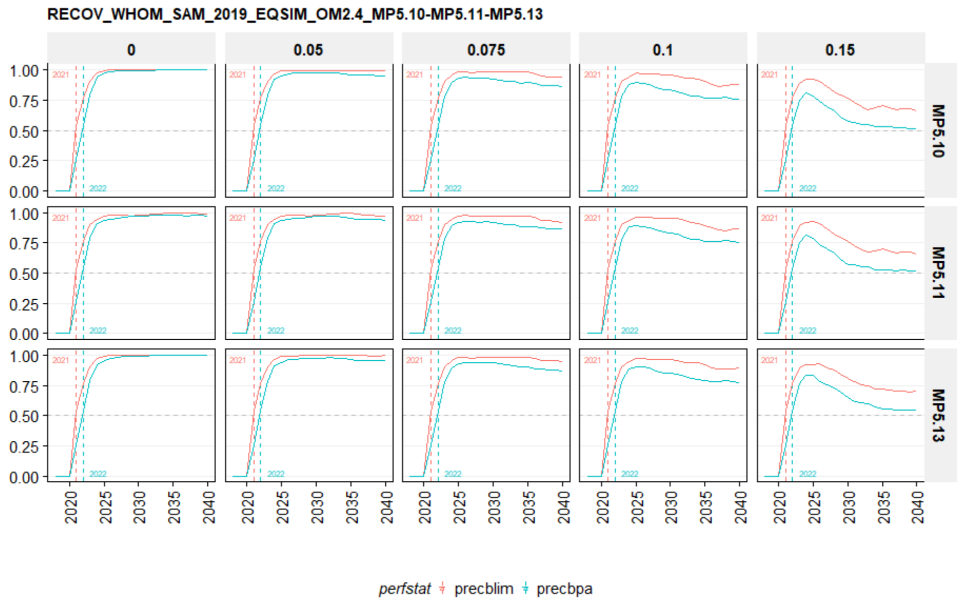
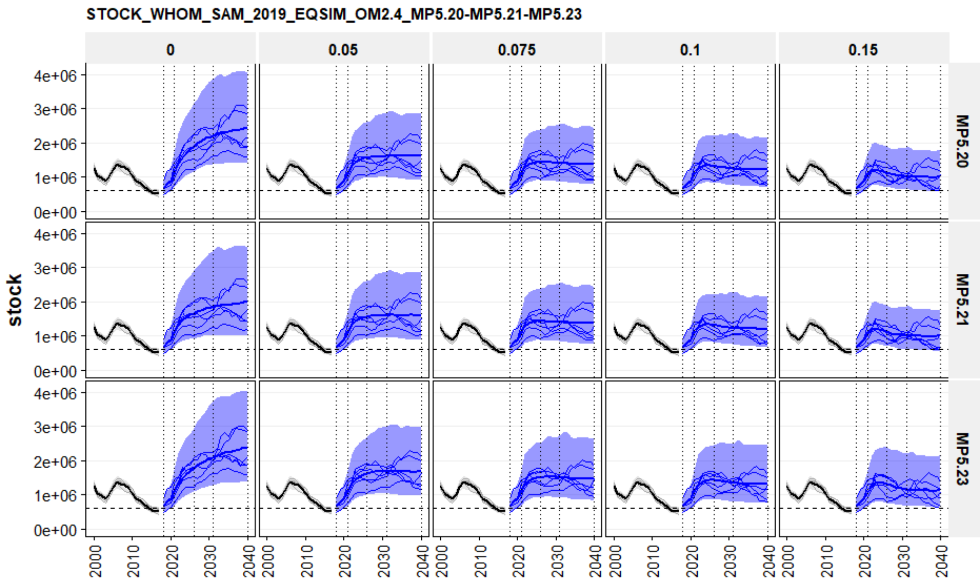


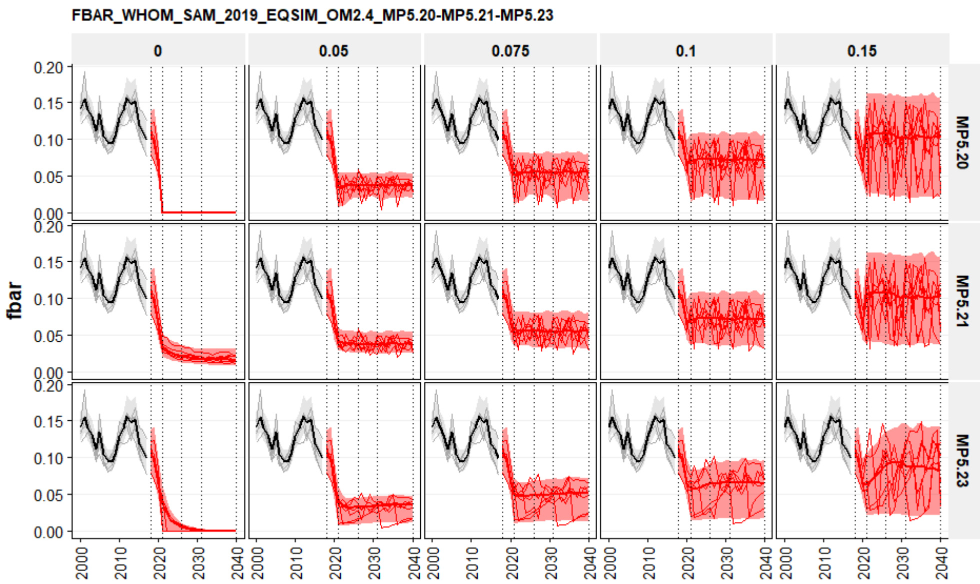
Figure 8.4 Simulation summary 2019 assessment, SAM, EqSim. MP5.10 ICES AR; MP5.11 ICES AR with minimum TAC of 50kT; MP5.13 ICES AR with 20% IAV constraint above Btrigger.

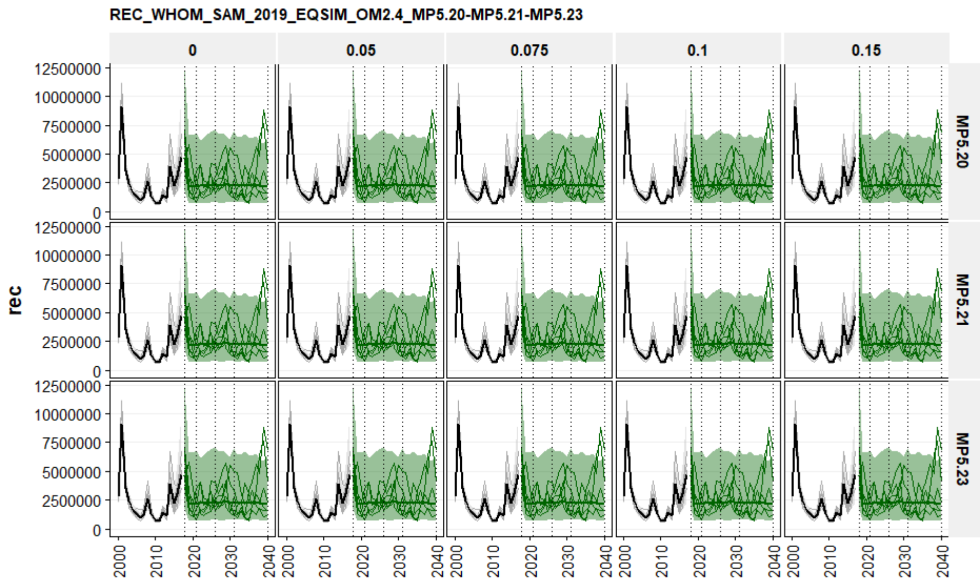
8.2.2 Double breakpoint rule

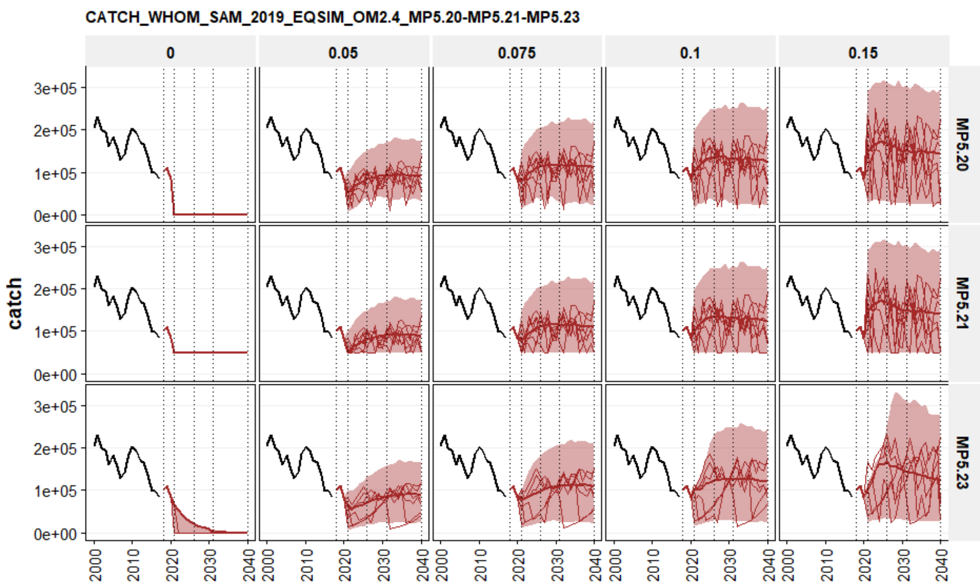
Simulation summary 2019 assessment, SAM, EqSim.

MP5.20 Double BP; MP5.21 Double BP with minimum TAC of 50kT; MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.






REC_WHOM_SAM_2019_EQSIM_OM2.4_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
rec
12500000
10000000
7500000
5000000
2500000
0
2000
2010
2020
2030
2040


CATCH_WHOM_SAM_2019_EQSIM_OM2.4_MP5.20-MP5.21-MP5.23
0
0.05
0.075
0.1
0.15
MP5.20
MP5.21
MP5.23
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040

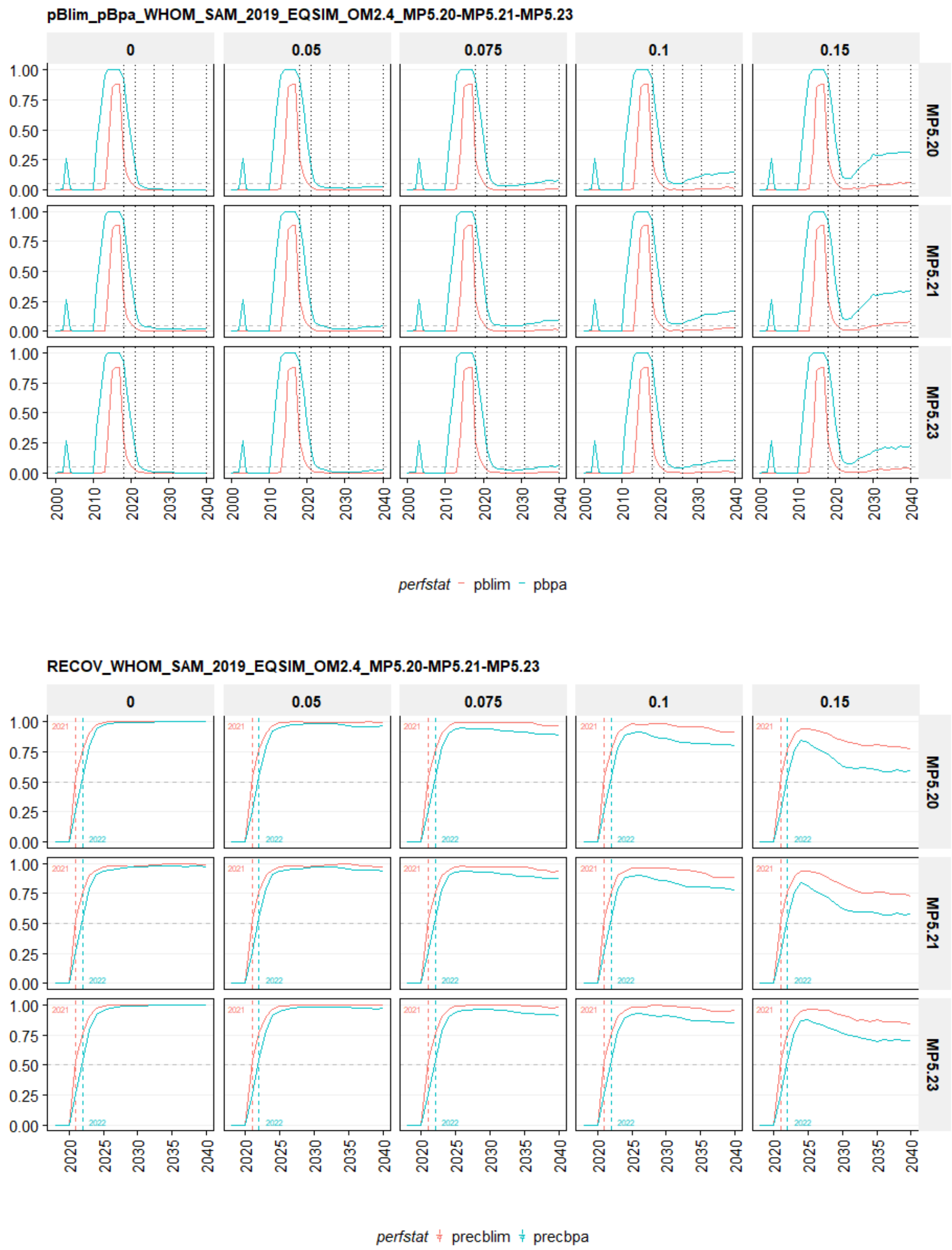
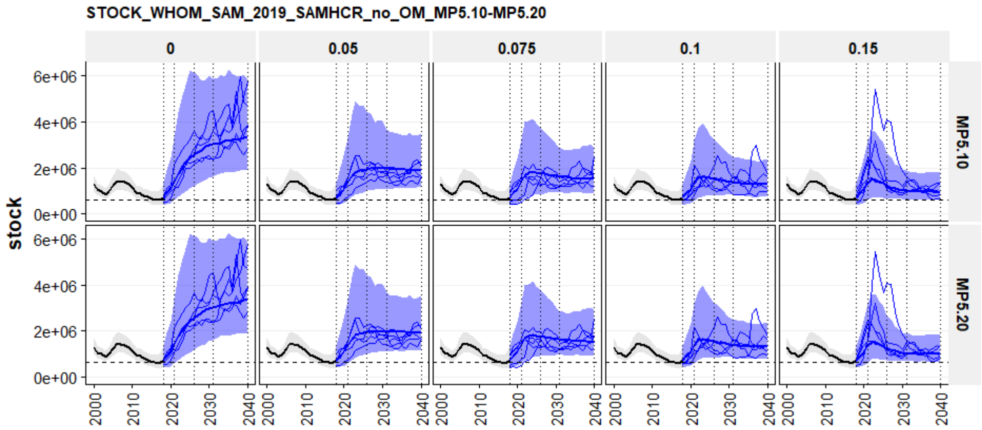


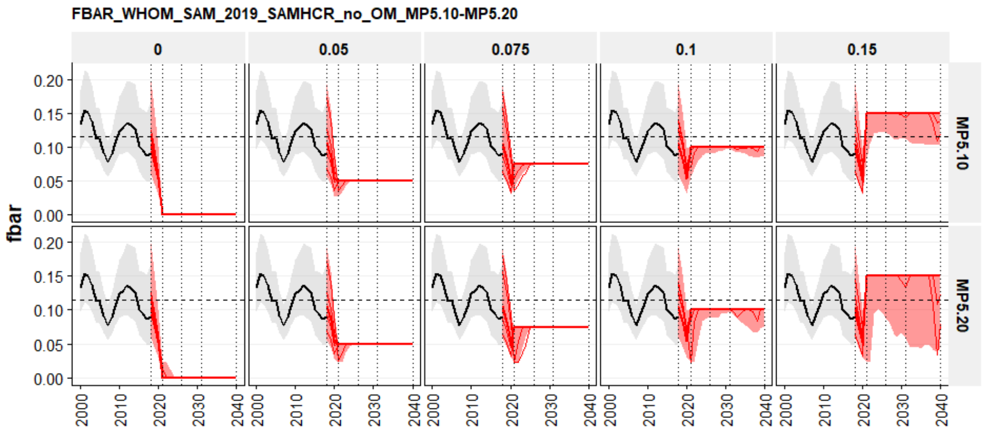
Figure 8.5 Simulation summary 2019 assessment, SAM, EqSim. MP5.20 Double BP; MP5.21 Double BP with minimum TAC of 50kT; MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

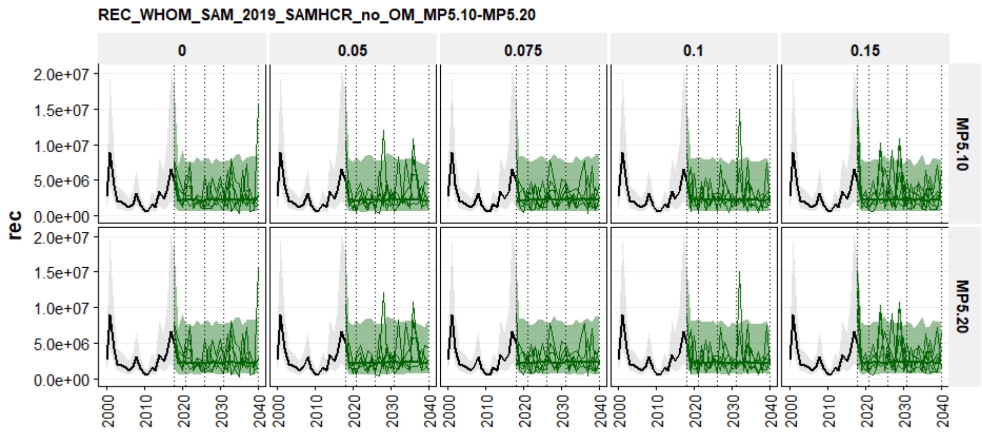
8.3 SAM and SAM HCR forecast

SAM Assessment and SAM HCR forecast.

MP5.10 ICES Advice Rule; MP5.20 Double BP; No IAV of fixed catch options. Minimum F in Double BP is 20% of Fmsy.







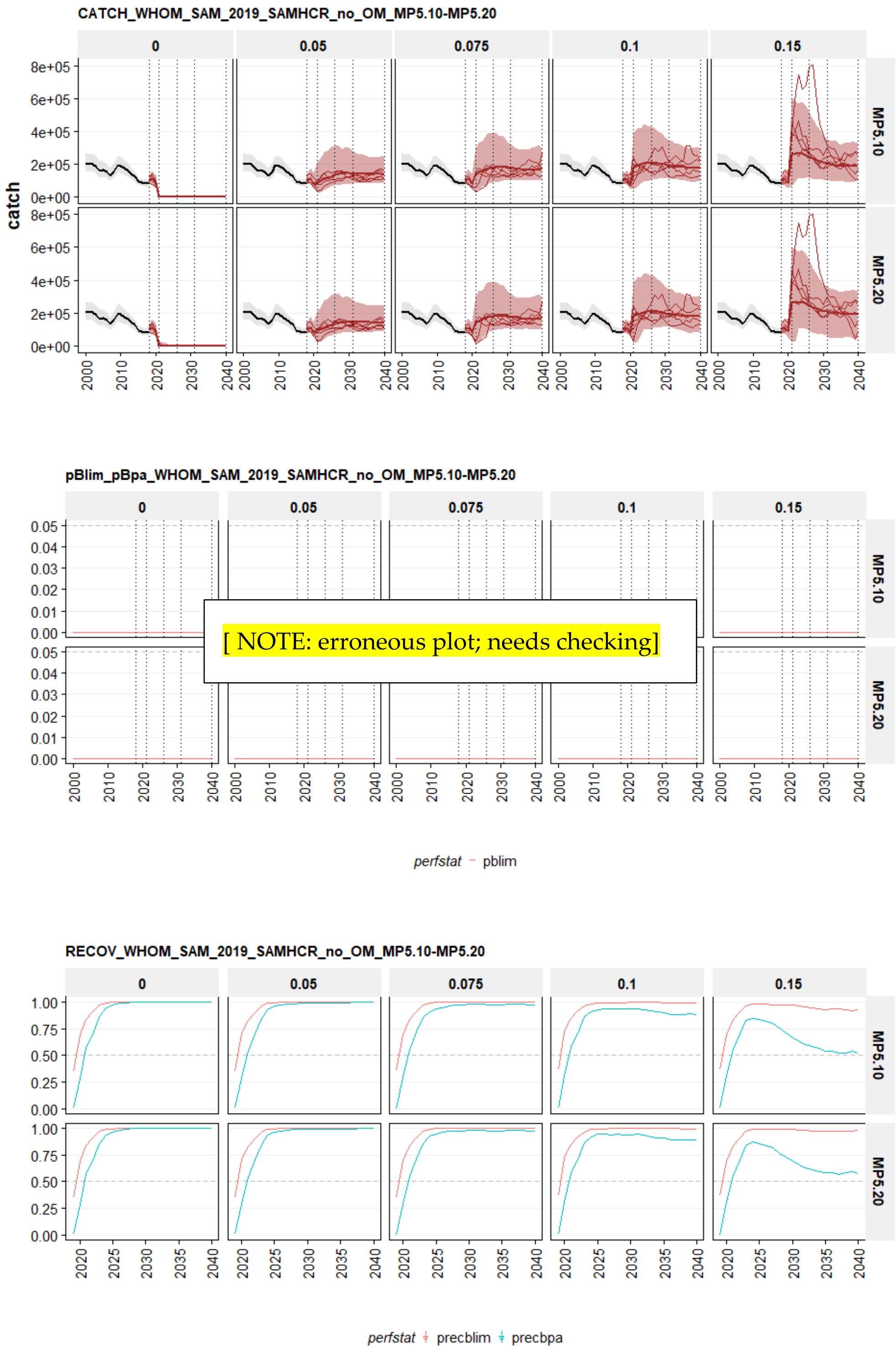


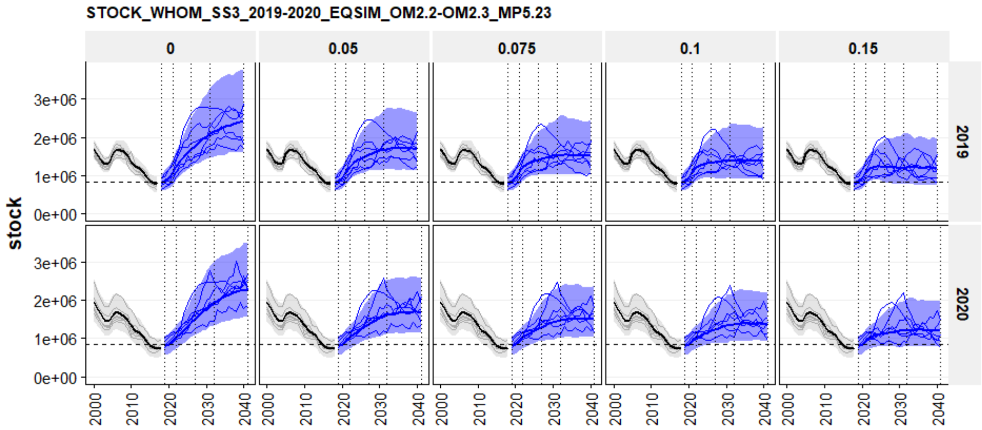
Figure 8.6 Simulation summary of SAM Assessment and SAM HCR forecast. MP5.10 ICES Advice Rule; MP5.20 Double BP; No IAV of fixed catch options. Minimum F in Double BP is 20% of Fmsy.

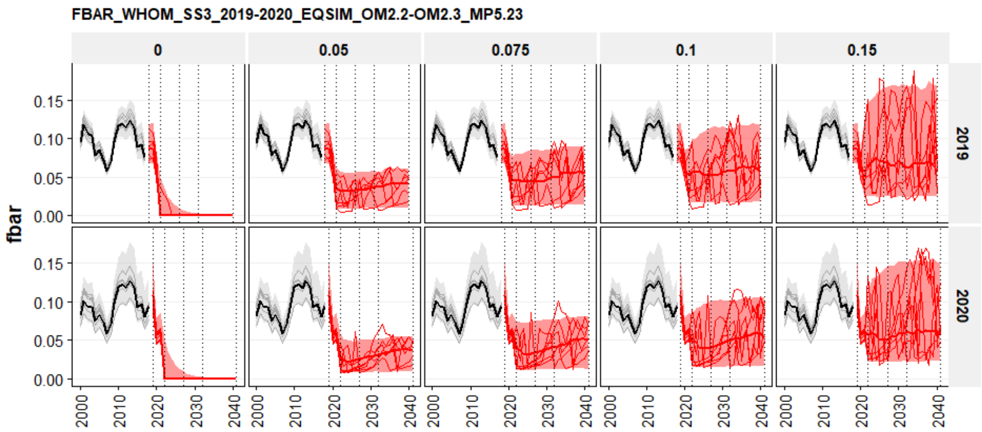
8.4 Sensitivity analysis to assessment year as basis for the simulation

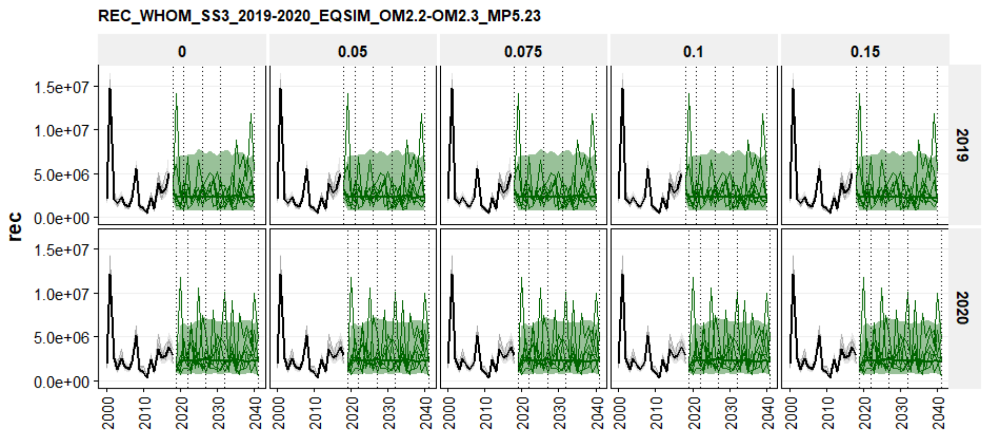
8.4.1 Comparison of 2019 and 2020 SS3 assessment and EqSim simulator

MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

SS3 assessment of 2019 and 2020 using Eqsim simulator







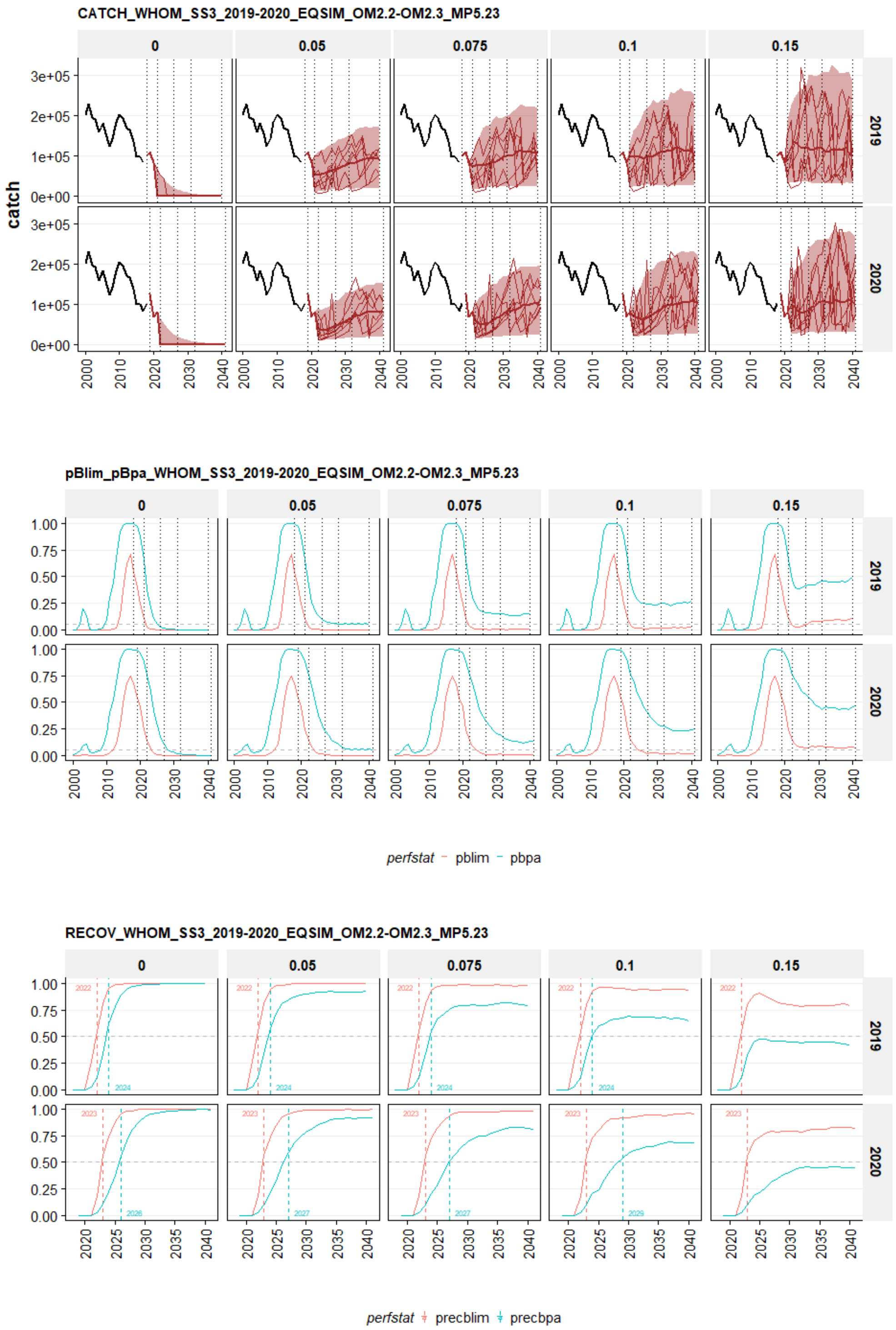
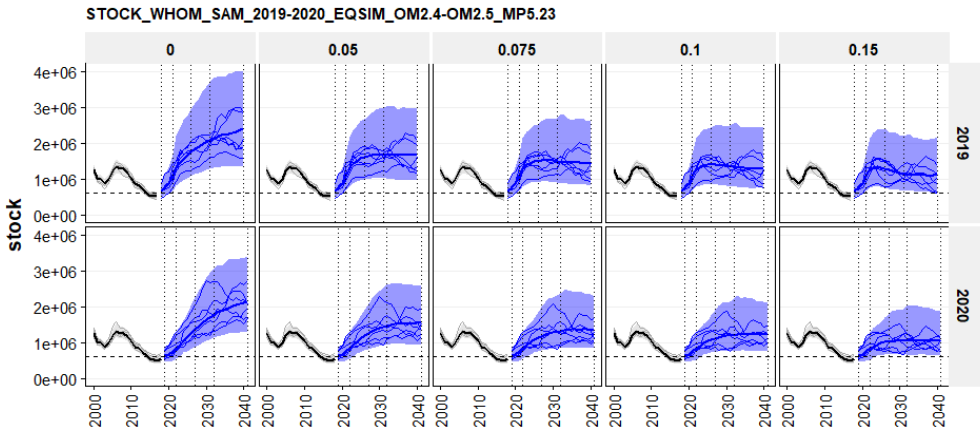


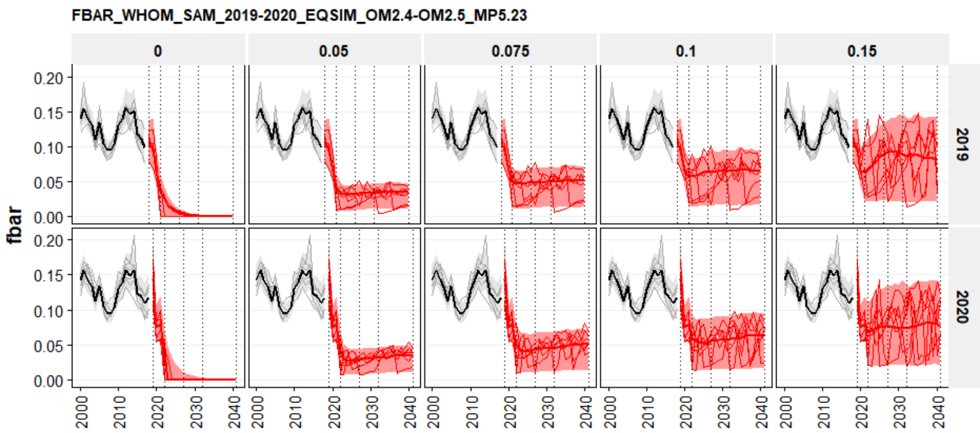
Figure 8.7 Simulation summary of SS3 Assessment 2019 and 2020 and EqSim simulator. MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

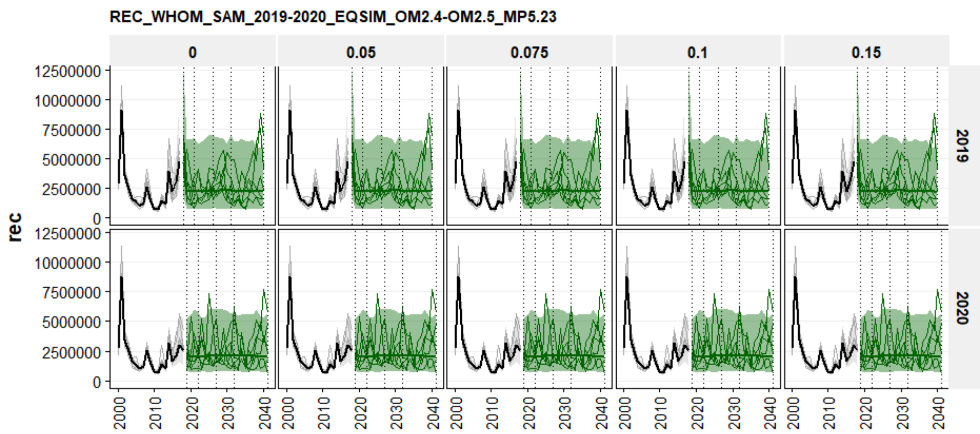
8.4.2 Comparison of 2019 and 2020 SAM assessment and EqSim simulator

MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

SAM assessment of 2019 and 2020 using Eqsim simulator







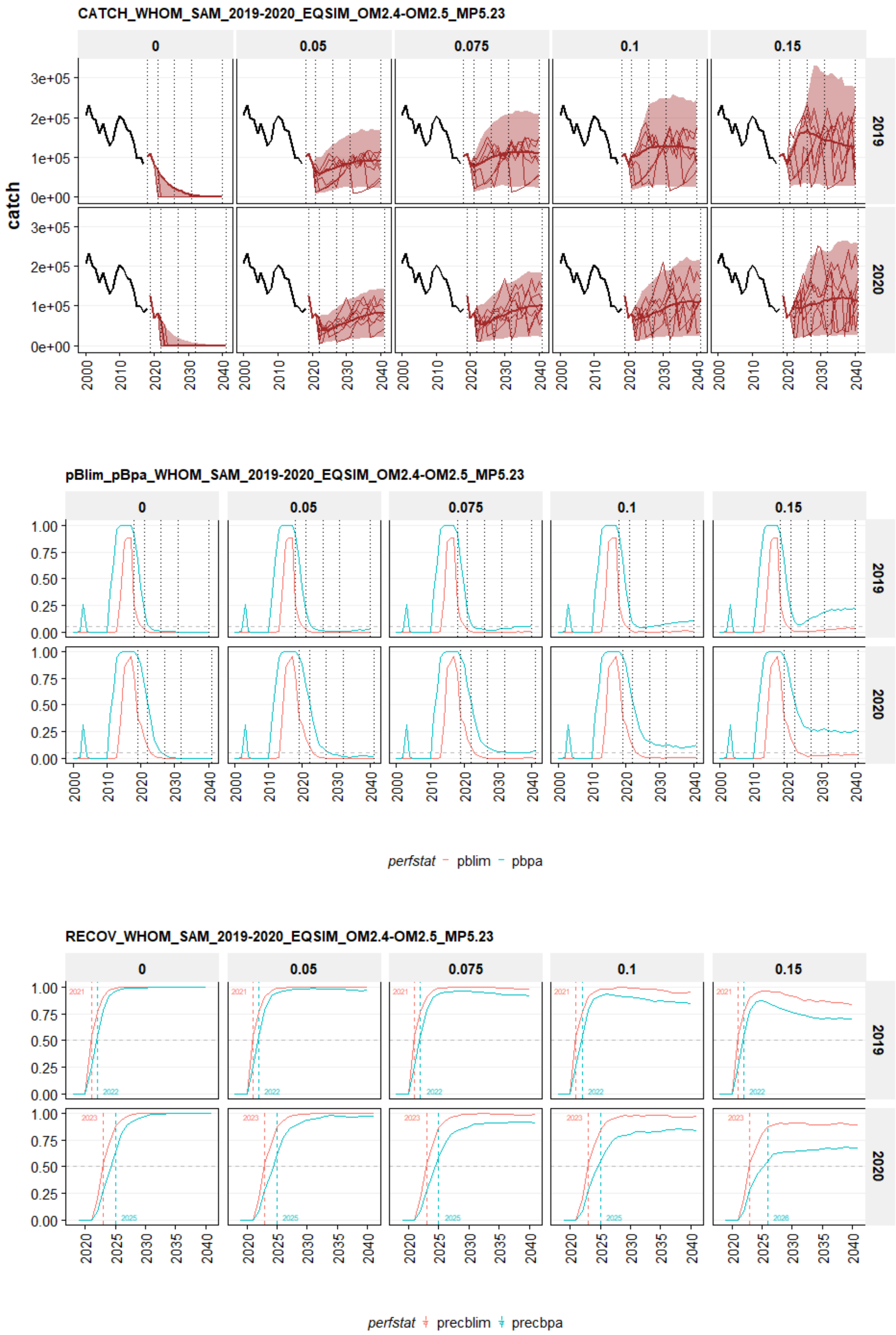


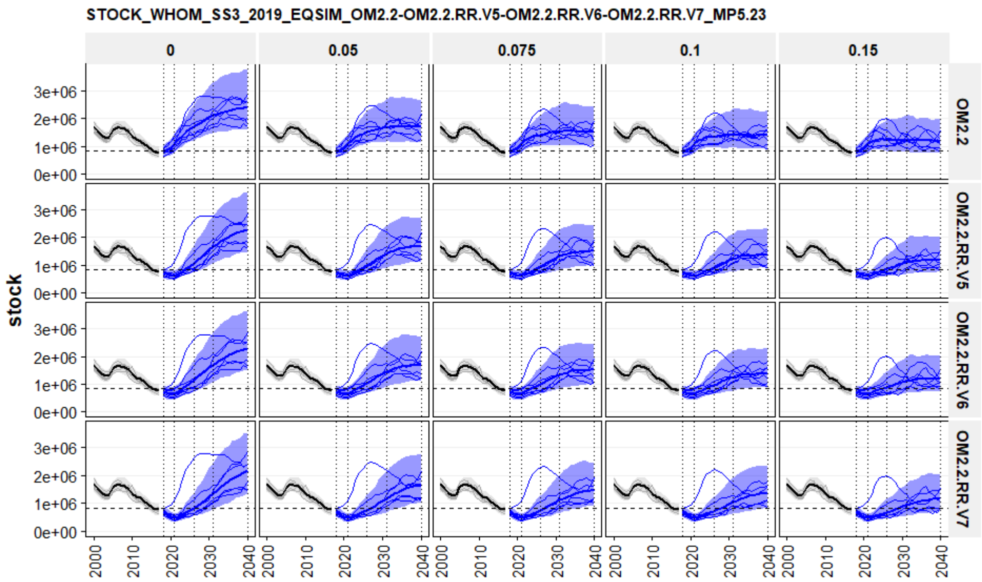
Figure 8.8 Simulation summary of SAM Assessment 2019 and 2020 and EqSim simulator. MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

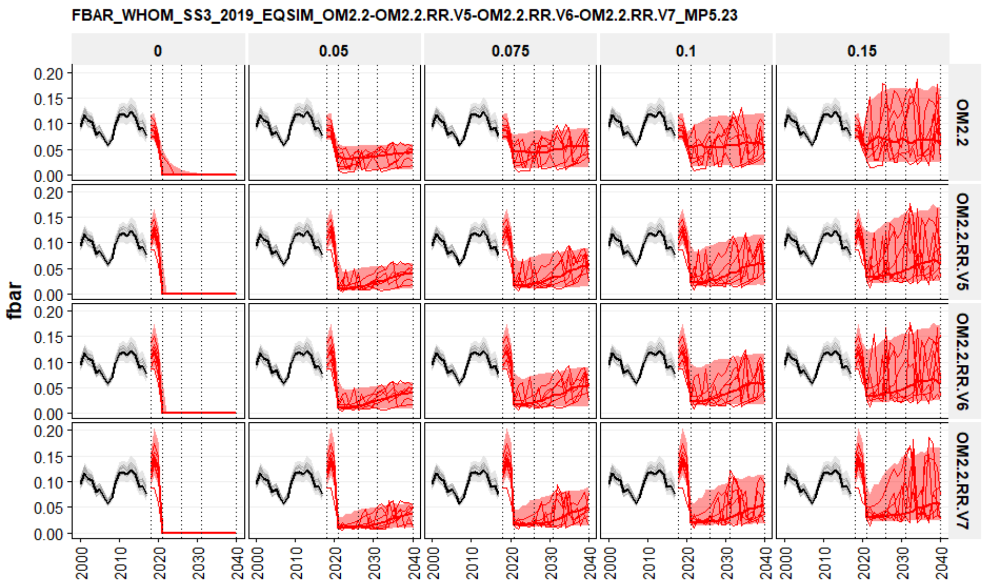
8.4.3 Sensitivity of 2019 SS3 with different recruitment assumptions

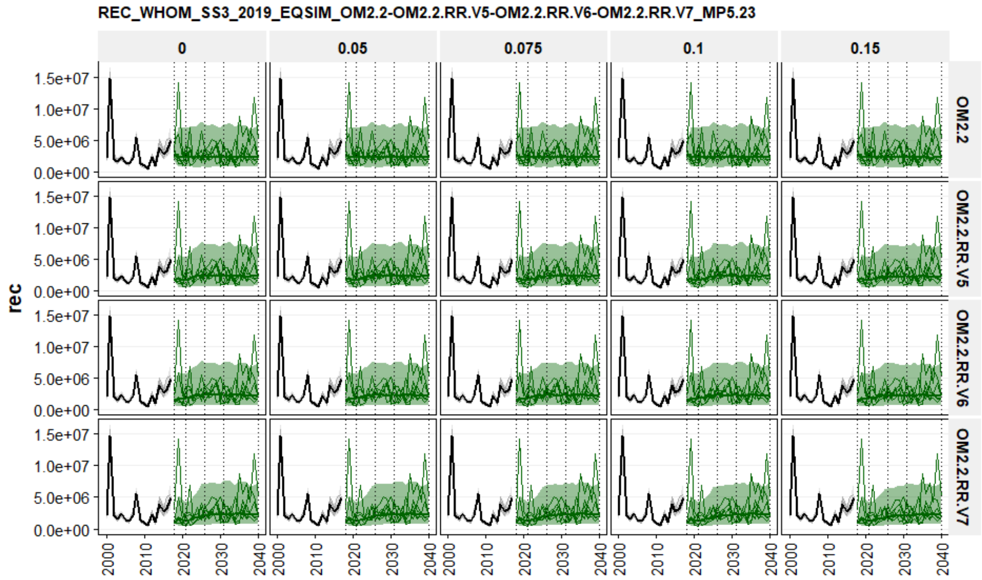
2019 SS3 assessment using Eqsim simulator and different recruitment assumptions. MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

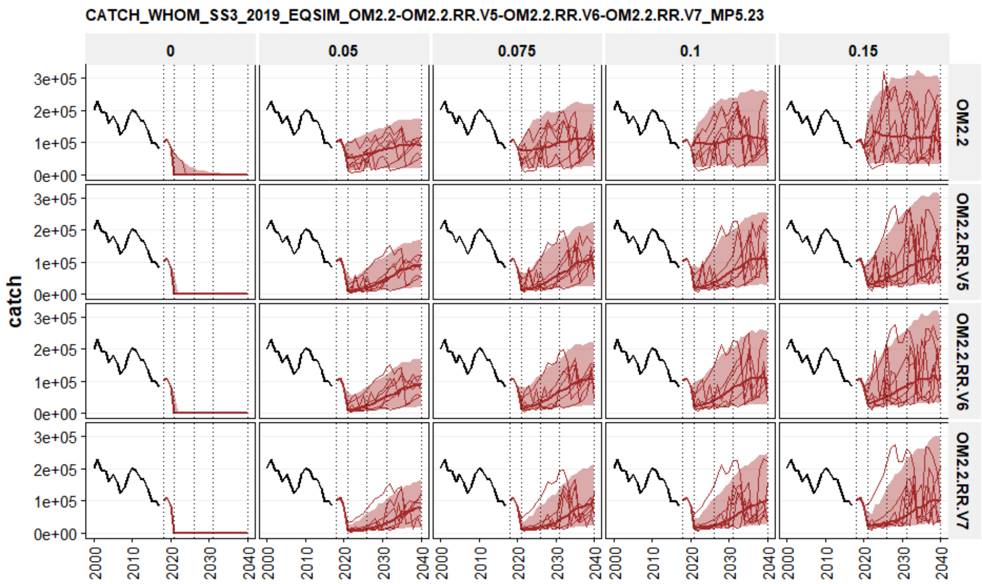
Recruitment assumptions:

- base case: recruitment from 2019 SS3 assessments. The geo mean of the full time series (ex 1982) is 2,572,649. The mean of the last 5 years from the 2019 assessment is 3,612,642
- RR.V5: 2014-2018 recruitments reduced to the GM of the time series from 2002-2013. Apart from 2008, this represents an extended period of low recruitment. The geo mean for 2002-2013 is 1,620,516
- RR.V6: 2014-2018 recruitments reduced to ½ of the baseline estimate. The mean of the last 5 years is then 1 806 321, similar to scenario RR.V5.
- RR.V7: 2014-2018 recruitments reduced to the mean of the 5 lowest from the full time series. The mean of the 5 lowest is 1,001,051. This is the most pessimistic scenario.


STOCK_WHOM_SS3_2019_EQSIM_OM2.2-OM2.2.RR.V5-OM2.2.RR.V6-OM2.2.RR.V7_MP5.23
0
0.05
0.075
0.1
0.15
OM2.2
OM2.2.RR.V5
OM2.2.RR.V6
OM2.2.RR.V7
stock
3e+06
2e+06
1e+06
0e+00
2000
2010
2020
2030
2040


FBAR_WHOM_SS3_2019_EQSIM_OM2.2-OM2.2.RR.V5-OM2.2.RR.V6-OM2.2.RR.V7_MP5.23
0
0.05
0.075
0.1
0.15
OM2.2
OM2.2.RR.V5
OM2.2.RR.V6
OM2.2.RR.V7
fbar
0.20
0.15
0.10
0.05
0.00
2000
2010
2020
2030
2040


REC_WHOM_SS3_2019_EQSIM_OM2.2-OM2.2.RR.V5-OM2.2.RR.V6-OM2.2.RR.V7_MP5.23
0
0.05
0.075
0.1
0.15
OM2.2
OM2.2.RR.V5
OM2.2.RR.V6
OM2.2.RR.V7
rec
1.5e+07
1.0e+07
5.0e+06
0.0e+00
2000
2010
2020
2030
2040


CATCH_WHOM_SS3_2019_EQSIM_OM2.2-OM2.2.RR.V5-OM2.2.RR.V6-OM2.2.RR.V7_MP5.23
0
0.05
0.075
0.1
0.15
OM2.2
OM2.2.RR.V5
OM2.2.RR.V6
OM2.2.RR.V7
catch
3e+05
2e+05
1e+05
0e+00
2000
2010
2020
2030
2040

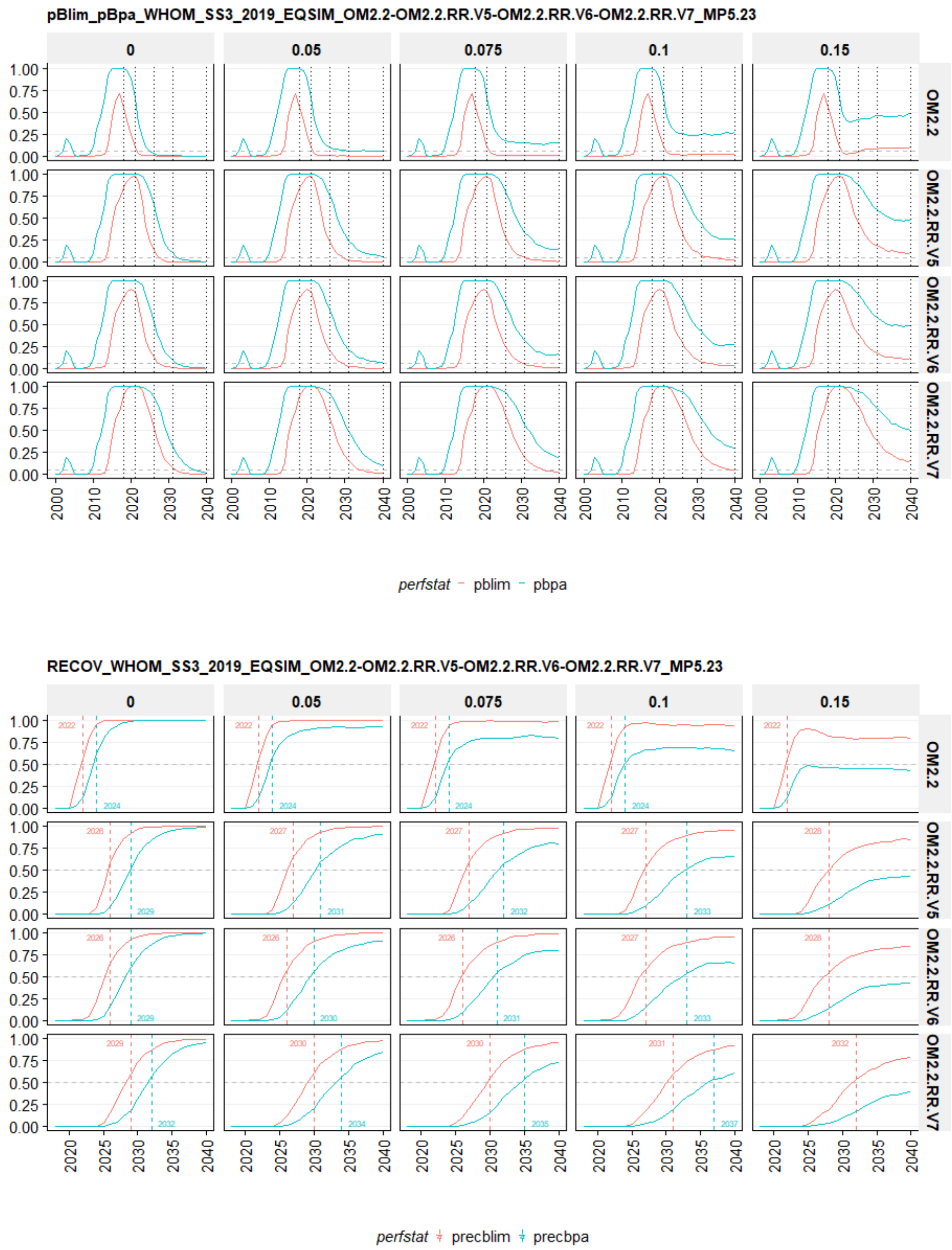


Figure 8.9 Simulation summary of SS3 Assessment 2019, EqSim simulator and different recruitment as-sumptions. MP5.23 Double BP with 20% IAV constraint above Btrigger. Minimum F in Double BP is 20% of Fmsy.

8.4.4 First year rebuild to a biomass reference point under different recruitment assumptions

Continued from section 8.4.3.

perfstat	scenario	recscenario	0	0.025	0.05	0.075	0.1	0.125	0.15
precblim	SAM 2019	def	2021	2021	2021	2021	2021	2021	2021
precblim	SAM 2020	def	2023	2023	2023	2023	2023	2023	2023
precblim	SS3 2019	def	2022	2022	2022	2022	2022	2022	2022
precblim	SS3 2019	RR.V5	2026	2026	2027	2027	2027	2028	2028
precblim	SS3 2019	RR.V6	2026	2026	2026	2026	2027	2027	2028
precblim	SS3 2019	RR.V7	2029	2029	2030	2030	2031	2031	2032
precblim	SS3 2020	def	2023	2023	2023	2023	2023	2023	2023
precbpa	SAM 2019	def	2022	2022	2022	2022	2022	2022	2022
precbpa	SAM 2020	def	2025	2025	2025	2025	2025	2025	2026
precbpa	SS3 2019	def	2024	2024	2024	2024	2024	2025	Inf
precbpa	SS3 2019	RR.V5	2029	2030	2031	2032	2033	2036	Inf
precbpa	SS3 2019	RR.V6	2029	2029	2030	2031	2033	2037	Inf
precbpa	SS3 2019	RR.V7	2032	2033	2034	2035	2037	Inf	Inf
precbpa	SS3 2020	def	2026	2026	2027	2027	2029	2031	Inf

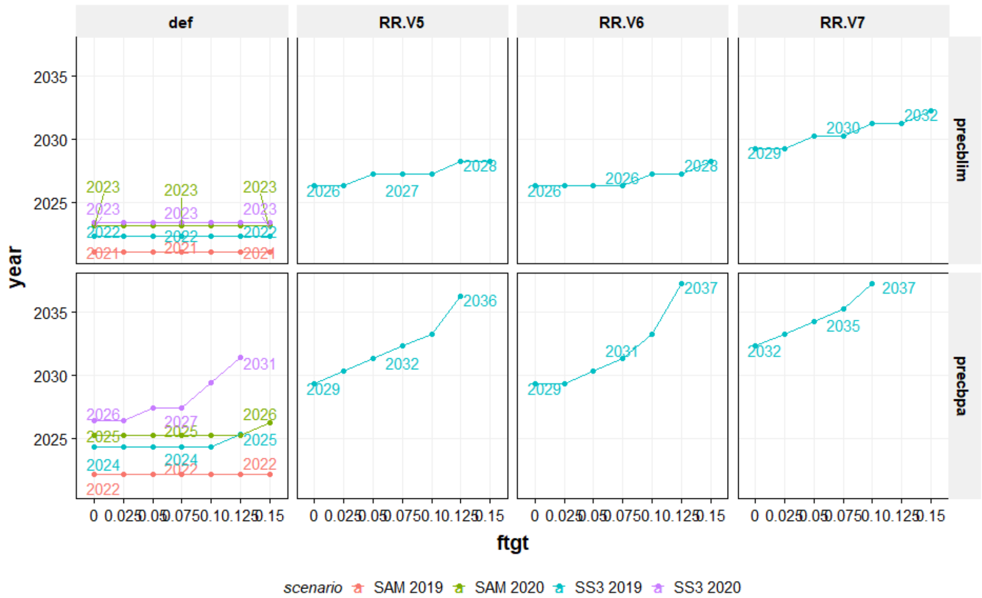


Figure 8.10 First year to rebuild to a biomass reference point under different recruitment assumptions

8.4.5 First year rebuild to a biomass reference point when 'wrong' reference points are used

What happens if SAM reference points are used for SS3 assessment or when SS3 reference points are used for the SAM assessment.

perfstat	scenario	rpscenario	0	0.025	0.05	0.075	0.1	0.125	0.15
precblim	SAM 2019	def	2021	2021	2021	2021	2021	2021	2021
precblim	SAM 2019	WR	2022	2022	2022	2022	2022	2022	2022
precblim	SAM 2020	def	2023	2023	2023	2023	2023	2023	2023
precblim	SAM 2020	WR	2025	2025	2025	2025	2025	2025	2025
precblim	SS3 2019	def	2022	2022	2022	2022	2022	2022	2022
precblim	SS3 2019	WR	2022	2022	2022	2022	2022	2022	2022
precblim	SS3 2020	def	2023	2023	2023	2023	2023	2023	2023
precblim	SS3 2020	WR	2023	2023	2023	2023	2023	2023	2023
precbpa	SAM 2019	def	2022	2022	2022	2022	2022	2022	2022
precbpa	SAM 2019	WR	2024	2024	2024	2024	2024	2024	2024
precbpa	SAM 2020	def	2025	2025	2025	2025	2025	2025	2026
precbpa	SAM 2020	WR	2027	2028	2028	2029	2030	2032	Inf
precbpa	SS3 2019	def	2024	2024	2024	2024	2024	2025	Inf
precbpa	SS3 2019	WR	2022	2022	2022	2022	2022	2022	2022
precbpa	SS3 2020	def	2026	2026	2027	2027	2029	2031	Inf
precbpa	SS3 2020	WR	2023	2023	2023	2023	2023	2023	2023

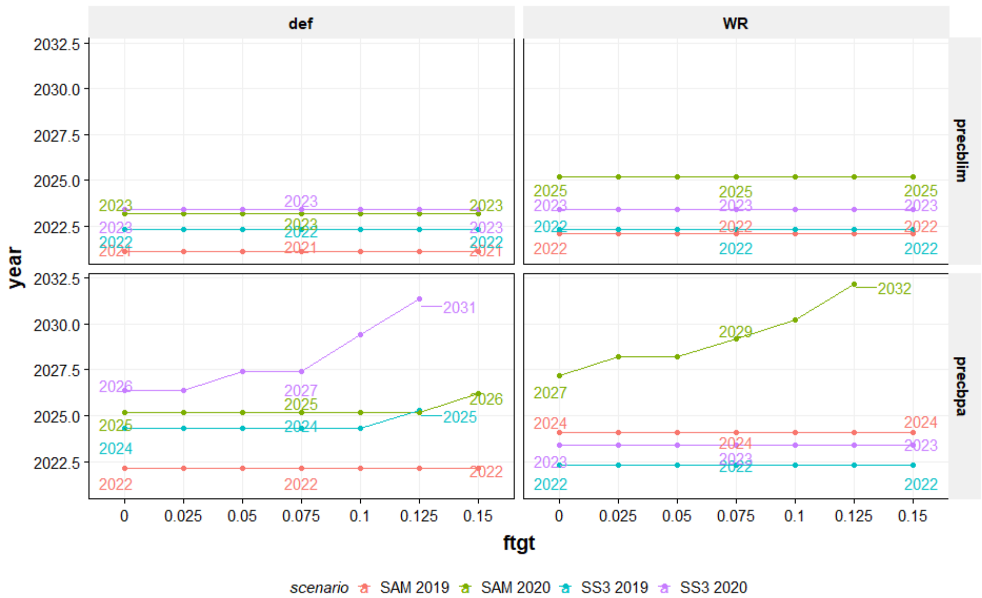


Figure 8.11 First year to rebuild to a biomass reference point when 'wrong' reference points are used

Annex 4: Summary template for HCR

A summary of the MSE is provided below in the format requested by ICES MSE guidelines (developed by the series of WKGMSE workshops)

BACKGROUND		
Motive/initiative/background.	There is no agreed management plan currently in place for western horse mackerel. The Pelagic Advisory Council (PELAC) initially proposed a plan in 2007 and this was used for the purposes of catch advice until 2010. Since this time, advice has been based on the MSY approach and the PELAC has continued efforts to develop a new management plan. The evaluation that is presented, has been conducted by a group of experts, on request from the PELAC. On the basis of the evaluation, the PELAC has selected a preferred management procedure.	
Main objectives	Rebuilding of the stock to levels well above B_{lim}. In recent years, the median SSB estimate from the stock assessment has been just above B_{lim} . During periods of lower than average recruitment, management should lead to catches that maintain the stock above B_{lim} with a high probability whilst avoiding large interannual changes in TAC.	
Formal framework	EC request to ICES to evaluate the MSE conducted by the PELAC technical focus group	
Who did the evaluation work	PELAC technical focus group on horse mackerel (Pastoors *et al*., 2020)	
Method		
Software Name, brief outline include ref. or documentation	1. A further adaptation of SimpSIM which was, initially developed from the *EqSim* and FLR R packages for the 2016 blue whiting MSE (ICES, 2016). Code of the adapted version is on github (https://github.com/ices-eg/wk_WKREBUILD; folder EqSimWHM) . The *EqSim* based simulator was conditioned on the SS3 assessment model as carried out during WGWIDE 2019. Additional runs were also conditioned with the WGWIDE 2020 assessment and with the exploratory SAM assessments (2019, 2020) as reported in the relevant expert groups (ICES, 2019; ICES, 2020b). The *EqSim* simulator is based on an age-structured operating model, with catches, numbers and F at age derived from simulated assessments. Assessment/advice error on SSB and F is incorporated following guidelines from WKMSYREF3 (ICES, 2014) 2. SAM HCR forecast based on the exploratory SAM assessment. The WHOM assessment model scripts and results used in this study are available at https://www.stockassessment.org/set-Stock.php?stock=WHOM2019. The HCR forecast in SAM that is used in this study was first developed and presented in Trijoulet et al. (Trijoulet et al., 2021). The code is available at https://github.com/vtrijoulet/SAM/tree/master2 and the main function is forecast2(). The forecast is a stochastic forecast that allows HCRs in the management procedure. It was set up to replicate as closely as possible the settings in the *EqSim* simulator.	
Type of stock	Medium life span (>20 years), pelagic	
Knowledge base *	Analytic assessment (SS3) and exploratory assessment (SAM)	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Segmented regression with a breakpoint at B_{lim} . Four scenarios considered (base case and three alternative (more pessimistic) scenarios whereby recent increases in recruitment (since 2014) is reduced).	Log-normal, CV from residuals
Weight at age	Random selection from stochastic time series of catch/stock weight at age.	Weight at age estimates from each of the 1000 model replicates were used to generate a time series stochastic weights via application of an error term derived from variability in the historic weight at age datasets to the replicate mean. The error term is considered age specific and autocorrelated.
Maturity	Maturity during the simulations is time invariant (as in the assessment)	No stochasiticity implemented

Natural mortality	0.15 – all ages and years	No stochasiticity implemented
Selectivity	From assessment	The MSE draws at random from the 1000 individual selection profiles available from the replicates for each year of the projection period.
Initial stock numbers	From assessment	Using 1000 model iterations to derive starting values.
Reference points	OM assessment-specific reference points are estimated (because of different partitioning between SSB and F for the exploratory SAM and WGWIDE SS3 assessments)	
	WGWIDE/SS3 reference points B_{lim} 834,480t B_{pa} 1,168,272t F_{MSY} 0.074	SAM reference points B_{lim} 611,814t B_{pa} 856,540t F_{MSY} 0.115
Decision basis **	SSB in the TAC year	
Number of iterations	1000	
Projection time	23 years	
Observation and implementation models (no assessment in the loop)		
Type of noise	*EqSim*: historic advice error on F and SSB SAMHCR: uncertainty directly taken from SAM assessment (function: simulate)	
Comparison with ordinary assessment?	Advice uncertainty is larger than model parametric uncertainty (SS3 estimates relative low uncertainties). The historic advice error for the last ten years was derived according to the methodology detailed in WKMSYREF3 (ICES, 2014) SAMHCR is based on uncertainty estimated within the SAM model.	
Projection: If yes - how?	*EqSim*: no forecast included. TAC derived directly from SSB estimate taking into account the advice error that was estimated for the forecast. SAMHCR: is essentially a multi-annual forecast using different HCR types.	
Projection: Deviations from WG practice?	Yes, WG runs a deterministic projection based on the SS3 model	
Implementation	Catches in numbers at age from projection according to the rule. No implementation error assumed.	
Harvest rule		
Harvest rule design	Proposed HCR: Double breakpoint on SSB: If SSB < Blim, Flow = 0.2 * Ftarget If Blim<SSB<Bpa: linear slope between Flow and Ftarget If SSB > Bpa: F = Ftarget Fixed F and ICES advice rule (with breakpoint at Bpa) HCRs were also evaluated.	
Stabilizers	Optionally, 20% TAC change limits (only applied above $B_{trigger}$ (B_{pa}) Scenarios with fixed minimum catch (50 kt) were run.	
Duration of decisions	Annual	
Revision clause	LTMP specifies a revision in no more than 5 years from first year of implementation.	
Presentation of results		
Interest parameters	The probability of stock size exceeding a biomass threshold *e.g.* B_{lim} or B_{pa}. For this MSE, the stock is considered to have achieved rebuilding when the probability that biomass exceeds B_{pa} is at least 50% for three consecutive years. Other parameters of interest are yield and inter-annual variation in yield.	
Risk type and time interval	Type 3	
Precautionary risk level	5%	
Experiences and comments		
Review, acceptance:	Under review (WKWHMRP 2021)	
Experiences and comments	Running two different operating models (SS3 and SAM) was useful to explore the sensitivity of the rebuilding probability to the different model formulations. Running two different evaluation methods (EqSim and SAMHCR) was useful to explore how the results of the management procedure are dependent on the evaluation method. Model results were comparable but not identical. Main difference is due to the different perception of the stock relative to reference points.	

Annex 5: Review

Review of:
Western Horse Mackerel Technical Focus Group on Harvest Control Rule Evaluations 2020
M.A. Pastoors, A. Campbell, V. Trijoulet, D. Skagen, M. Gras, G.I. Lambert, C.R. Sparrevohn, S. Mackinson

Completed by:
Jonathan J. Deroba
NOAA NMFS
Woods Hole, MA, USA
jonathan.deroba@noaa.gov

Introduction

I would like to thank the authors for their efforts and for the opportunity to provide a review. I have a great deal of appreciation for the time and work required to conduct this type of research. For lack of a better alternative, I structured my review similar to the way a peer review might be written for a journal article. More specifically, I used page numbers and referenced specific sections of text that inspired a comment, question, or recommendation. I first provide a review of "Annex 1: PELAC proposal for a rebuilding plan for Western horse mackerel" and then a review of the Evaluations document. Lastly, I provide commentary as to whether each Term of Reference was met.

Review of Annex 1

Page 1, numbered issue 6: I strongly support the preemptive development of a rebuilding plan because developing something after it is "too late" adds far more stress and urgency to the situation that only hinders analyses. To resolve the issues of instability in the stock assessment, research could be conducted external to the assessment process that provides estimates of parameters that inform scale, such as catchability or natural mortality. These parameters could then be fixed in the assessment or strong priors used during stock assessment fitting. This suggestion is well beyond the scope of this working group and should not be considered a shortcoming of what was done. I only mention it so that perhaps it is considered more broadly within ICES.

Page 3, article 5: Article 5, combined with language in the TORs, specifies the exact harvest control rule to use for rebuilding. More typically, an MSE is conducted that compares the performance of a broad range of control rules so that an option robust to major uncertainties can be identified and applied. The current process and TORs seems limiting in the sense that the question at hand is "can the given control rule achieve rebuilding", which is very different than what I would consider typical, "find the control rule that will likely meet the rebuilding objectives". I suggest future work and processes focus on the latter because it is very plausible that other control rules may be preferable to the one specified *a priori* in this case.

PELAC Report

Page 6, Horse Mackerel Stock ID: At this point, I wondered how this topic was considered in the MSE, but it was not explicitly addressed. If the intent of this work was to argue that this needs to be addressed in the future or that stock structure should be considered an uncertainty in the MSE, then I suggest adding some text putting this work within the context of the MSE.

Page 7, recommendation pertaining to length composition: I want to support the suggestion of the authors to align the length composition data with ICES standards. Improving data standardization will improve repeatability, transparency, and ultimately stakeholder acceptance.

Page 7, Contribution of recruitment to SSB: I was unclear as to how this work fit in with the broader work related to the MSE and rebuilding. I do not understand how some of these bits of work fit together for a common purpose.

Page 8, statement that “In probability theory, there is a very simple formula for variance of a weighted sum of independent components”: Although a minor issue with no consequence in the context of the MSE, I doubt that a time series of recruitments are independent. I don’t know how a violation of this assumption would affect this work.

Page 8, Figure 3: I think this Figure 3 is missing from the document.

Page 10, the text explaining the process of using fixed selectivity in SAM: I can understand the issue that when the only age structured data available is the catches, that SAM might be inclined to fit it tightly given no possibility of contrary signals. But, preventing that tight fit by imposing fixed selectivity implies that the analyst does not “believe” the catches. I suggest providing some supporting graphics demonstrating that the fits to catches were unreasonably precise. As a contrast to fixing selectivity, what happens to the fit if you fix the observation variance of the catches at a relatively high value but still allow time varying selectivity? I do not think this necessarily needs to be added to the report, but offer it up as a consideration in the future.

Page 10, in regards to the fact that time-invariant selectivity improved the retrospective patterns: I do not understand this result because typically added flexibility in time varying dynamics is what reduces retrospective patterns, not less flexibility. Perhaps this is related to the overfitting of the catch data. I have no suggestions here, nor am I requesting follow-up, but this is curious and perhaps the analysts also consider it worthy of more exploration.

Page 13: Not enough detail was provided here for me to sufficiently follow the methods. Perhaps add a bit more. This section seems to have been written for someone with extensive prior knowledge.

Page 16, Conclusions: In regards to the fact that scale instability in the stock assessment complicates, if not invalidates, the calculation of Fcv and Fphi, perhaps each time series involved could be divided by its mean to remove scale. Fcv and Fphi could be calculated from these “scale-less” time series, which should remove the effect of instability. Of course this does not solve all the problems that instability causes, but it might help with Fcv and Fphi. Not suggesting this be added to this report because I think this issue requires dedicated exploration.

Page 16, Conclusions: These conclusions are well supported and I encourage ICES to follow-up on this topic given its broad importance.

Page 16, reference point table: The time series of SSB and F were broadly similar between SS3 and SAM, and so the reason why the reference points differ so much between these platforms is unclear. I suggest exploring why, and adding results comparing the recruitment time series and selectivity patterns because these seem like the most likely explanations for the discrepancy.

Page 17, HCR evaluations: The reference points (e.g., Blim, Bpa, Fmsy) were all seemingly specified in Annex 1 and assumed known without error in the MSE. This assumption is unlikely to be true and I suggest adding text somewhere in the document mentioning that the performance of the HCRs is likely to degrade when reference points are imprecise or biased. I would say the extent to which performance is robust to this issue is unknown, but a thorough evaluation is beyond the scope of the current evaluation.

Page 18, SAM HCR: I generally suggest a bit more details about many of the methods. In the case of SAM HCR projections, I suggest specifying whether process errors were carried forward into the projections, which process errors, and providing their variance (or SD) estimates.

Page 26, SAM HCR forecast tool: Given the ubiquity of SAM and the need for MSE within ICES, I encourage the expansion of the SAM HCR forecast tool to include assessment and implementation errors. Early modifications could use the short-cut approach, but efforts should strive toward full feedback and consistency with the WKGMSE guidelines.

Page 31, “This may be due to the fact that the SAMHCR is operating as a forecast only…”: You could test this hypothesis by turning the assessment and implementation errors off (or really low) in

EQSIM so that it is also forecast only. If this hypothesis holds, then EQSIM_SAM and SAMHCR projections should get more similar. Some alternative explanations might be differences in how recruitment is handled in projections, or the fact that SAMHCR also may have process error in survival.

Page 31, "Selection of preferred HCRs…": The way this text was written made it seem as though PELAC was a management body that chose the HCR to be applied, but I do not think that is strictly true. Providing some context in the document as to the role of PELAC in the management context would help outsiders understand the situation better. The PELAC also recommended the double break point rule, but I do not think that choice was clearly favorable over the ICES AR, which performed about as well in terms of rebuilding metrics. Some stronger justification should be added. I also wondered whether the double break point rule was counter to WKGMSE advice that control rules with sharp declines in F should be avoided.

Page 31, "Selection of preferred HCRs…": It was not clear to me how the entirety of the work included in the report was used to reach this preferred HCR.

Page 35, "Stock Synthesis results are in general a bit smoother compared to SAM.": I found this outcome counterintuitive. SAM likely has several random walk processes that typically create smoother time series than models like SS. Looking into the future, I suspect SAM will be considered for this stock at the next Benchmark and I would encourage an exploration into why SS was smoother in this case. I suspect SS is being constrained somehow, perhaps in estimation of recruitment deviations. Data weighting might also play a role, especially when using conditional age at length in SS.

Page 38, 100,000 tonnes: Is this value close to MSY? I ask only because the TORs state this as desired.

Terms of Reference

a)

i) This portion of the TOR was met.

ii) The PELAC report contained the bare minimum of requirements described in WKGMSE. Aligning the work more closely with WKGMSE guidelines might include testing the short-cut approach against a full feedback for at least a subset of simulations, running some simulations with no implementation error to serve as a point of comparison, and expanding the uncertainties covered by operating models.

iii) I felt the criteria used were prespecified in Annex 1 to some extent, and so in so far as the PELAC report included most of those criteria this portion of the TOR was met. Criteria used to compare HCRs is likely also best done by managers, and usually requires some iteration between analysts and management bodies.

b) This TOR was met. Given that this TOR and Article 5 were very prescriptive in what the evaluations were to do, I consider the work sufficient to evaluate whether the HCR largely specified in Article 5 meets the rebuilding criteria also specified in Article 5. The work completed, however, should not be considered a thorough MSE that followed best practice. The work did not: evaluate a broad range of operating models that encompass major uncertainties, include a broad range of HCRs, use a full feedback or condition the short-cut method to ensure a reasonable approximation. This critique does not reflect any failure of the analysts, as such a thorough approach was not their remit. Rather, this critique is to ensure that the results are interpreted in the relatively restrictive context in which it was requested.

c) I believe this TOR was met, but suggest adding explicit language tying the relevant work in the report to this specific TOR.

d) I believe this TOR was met, but suggest adding explicit language tying the relevant work in the report to this specific TOR. I also suggest reporting the MSY catch level so that this TOR can be more thoroughly evaluated.

e) The work was not described in relationship to the TORs, so this TOR was not met, but could easily be achieved.

Review of:
Updated Diagnostics and Simulations of the Western Horse Mackerel Rebuilding Plan Evaluations
M.A. Pastoors, A. Campbell, V. Trijoulet, M. Gras, G.I. Lambert, C.R. Sparrevohn, S. Mackinson, R. Ourens

Completed by:
Jonathan J. Deroba
NOAA NMFS
Woods Hole, MA, USA
jonathan.deroba@noaa.gov

Introduction

The authors should be commended for a thorough, clear, and well written update to the previous analysis. I have no major concerns and think the Terms of Reference have been sufficiently met. Most of my new comments are editorial in nature. I structured my review similar to the way a peer review might be written for a journal article. More specifically, I used page numbers and referenced specific sections of text that inspired a comment, question, or recommendation. At the end, I provide commentary as to the extent to which each Term of Reference was met.

Review Comments on Updated Analysis

Page 30: For the figure here and throughout, define what the numbers and labels mean in the banners and legend. There was also occasionally inconsistency in the language or abbreviations used in the text and captions and what appeared in the figures.

Page 34, the stock and catch weights at age: I suggest some added detail here about exactly how variation in weights at age were incorporated into the simulation period. As I understand it, the assessment assumes time invariant von B and a time invariant length-weight relationship, which I think equates to time invariant weights at age. Rather than make this assumption in projections, variability was added somehow using the historic weights at age. This sentence in particular, “Variability is incorporated within the time series by applying an error term derived from an analysis of the variability in the historic weight at age datasets,” needs more detail. How was the error term derived and applied?

Page 36: As with the weights at age described on page 34, I do not understand the details as to how the degree of noise in the selectivity pattern was determined, especially given time invariant selectivity used in the stock assessment.

Page 46, the last sentence: I think this sentence, “Using the appropriate or wrong reference points does not affect…” is incorrect and not entirely clear. Comparing the top row of figure 2.28, the SS3 assessments had the same rebuild year to Blim (comparing the blue and purple lines between the top left and top right panels), but the SAM assessments differed by one year in the case of SAM2019 and two years in the case of SAM2020. I suggest making this clarification. I think all the lack of clarity here could be fixed with a couple quick sentences, “For a given assessment, assessment year, and set of reference points, ftgt did not affect rebuild year to Blim. Rebuild year to Blim was also unaffected by using the appropriate or wrong reference points for SS3, but differed by 1 or 2 years for SAM depending on the assessment year.”

Page 47: Interesting question above Figure 2.28 as to why SAM2019 was unaffected while results for SAM2020 showed relatively large impact. I will be interested to see if you uncover anything here.

Page 51, the first paragraph of 3.1.1.6: The juxtaposition of pointing out that the assessment assumes time invariant selectivity, but that the simulations want to include “appropriate variability” in selectivity could be viewed as a criticism of the assumption used in the assessment. I do not think that is the intent, so perhaps make the distinction as to why time invariant assumptions are acceptable in one place but not the other. Or, even changing the text from saying, “However, in order to ensure…” to something like, “The assumption of time invariant selectivity was relaxed in projections to acknowledge that process errors in selectivity are likely to occur in the future,” might be sufficient. This issue also applies to weights at age, where the assessment assumes time invariant dynamics, but the projections do not.

Page 53, implementation model: I do not think the absence of implementation error is of any consequence because any reasonable assumption as to the degree of implementation error would get swamped by all the other noise in the projections. I wonder if there is even any anecdotal information about how reasonable this assumption is, however (does slippage occur? Are discards well quantified?). Future simulations could consider a bare minimum short-cut approach to implementation error, similar to the short-cut approach used for assessment error. That is, realized catch equals the TAC times some lognormal noise. As noted, however, I do not suspect this to have much effect on results and conclusions.

Page 57, “because not all simulated data resulted in a converged assessment…”: This rate of non-convergence is quite high, especially considering that the default for the simulation function only includes observation error (i.e., no process errors). While beyond the scope of this analysis, if SAM is considered for this stock in the future, then I suggest the reason for such relatively high non-convergence be explored.

Terms of Reference

a)

i) This portion of the TOR was met.

ii) This portion of the TOR was met. Future work could move this MSE closer to the ideal situation described in WKGMSE guidelines, such as testing the short-cut approach against a full feedback for at least a subset of simulations and expanding the uncertainties covered by operating models.

iii) This portion of the TOR was met. Any additional considerations as to the appropriateness of the criteria should likely be done with managers.

b) This TOR was met.

c) This TOR was met.

d) This TOR was met.

e) This TOR was met.

Review of

Pastoors, M.A., Campbell, A., Trijoulet, V., Skagen, D., Gras, M., Lambert, G.I., Sparrevohn, C.R., and Mackinson, S. 2020. Report on Western Horse Mackerel Technical Focus Group On Harvest Control Rule Evaluations 2020. Report for Pelagic Advisory Council. 46 p.

and

ANNEX I: PELAC proposal for a rebuilding plan for Western horse mackerel. Dated 28 July 2020

Prepared 24-Jan-21 by:

A.R. Kronlund (Rob)

Interface Fisheries Consulting, Ltd.

E-mail: interfacefisheries@gmail.com

1. Recommendations

Recommendations in the review are listed here with page numbers for reference.

2. Introduction

Purpose and Terms of Reference

The two documents for review are (a) an evaluation of harvest options to support a management plan for the Western Horse Mackerel (WHM) fishery (Pastoors et al. 2020), and (b) a proposal for a rebuilding plan for WHM ("Annex 1"). The stated purposes of the documents are:

1. to develop operational evaluation tools for management plan evaluation,
2. to evaluate potential Harvest Control Rules (HCRs) so that the Pelagic Advisory Council (PelAC) can decide whether to adopt a (management) procedure, and
3. to ensure stock recovery to safe biomass levels and a long-term stock exploitation that is consistent with the precautionary approach and with achieving the objective of maximum sustainable yield (MSY).

My understanding is that evaluations leading to the selection of a management procedure must meet an acceptable scientific standard as determined by peer review. In addition, the performance of a successful management procedure must meet the constraints of the ICES precautionary and MSY approaches. Terms of reference (TOR) that guided this review are included below (***emphasis mine***):

a) Review the approach (i.e., analytical methods, application and interpretation) described in the Pelagic Advisory Council (PelAC) report (listed below) for the evaluation of the proposed Harvest Control Rule (HCR) in the rebuilding plan. The review should consider:

i) Whether the tools used (methods), and the model conditioning done (data), are appropriate for the stock;

ii) Whether the ***minimum requirements for simulation testing HCRs***, as developed by WKGMSE process, are met;

iii) The appropriateness of the ***criteria used to draw conclusions*** on the performance of the HCRs.

b) On the basis of (a), determine whether the HCR evaluations presented in the PelAC are sufficient to evaluate the proposed rebuilding plan against ***precautionary criteria*** and the ***rebuilding plan targets and measures*** detailed in Article 5 of the proposed rebuilding plan.

c) Depending on the outcome of (b), either:

  i) analyse the results of the HCR evaluation, and develop conclusions on whether the proposed rebuilding plan ***can be considered precautionary and be used as the basis for ICES fishing opportunity advice*** for the stock; **or**

  ii) Propose ***additional analyses or diagnostics*** that would allow for sufficient evaluation of the proposed plan

d) Should the proposed plan include elements that are in contradiction with ensuring that the stock is fished and maintained, also in the future, at levels which can produce MSY, comment specifically on such elements, and ***their consequences for ensuring MSY***.

e) Deliver a report containing the key decisions and conclusions in relation to the TOR.

Definitions used in this Review

In this review I use the term operating model (OM) to mean a hypothesis about stock and fishery dynamics as represented by a mathematical model that applies assumptions and uses relevant data. A management procedure (MP) is a set of:

1. stock and fishery monitoring data,
2. an assessment method,
3. a harvest control rule (HCR) used to translate assessment outputs into a catch limit, and
4. any meta-rules that modify the catch limit to meet a specific objective or constraint.

There may be multiple OMs and MPs under consideration. The HCR component of a MP typically has operational control points (OCPs) that can be used to “tune” the behavior of the MP. Sometimes the OCPs coincide with reference points (at least by name) but this is generally not a requirement (although may be under the ICES system?). Performance relative to objectives is the critical outcome of interest in the search for feasible MPs (rather than the use of a particular set of OCPs such as biological reference points). For example, an OM may be based on a statistical catch age model, but the assessment method in the MP applies a surplus production model (Cox and Kronlund 2008). Reference points based on maximum sustainable yield (MSY) calculated by the age-structured model will have different values than those estimated than the surplus production model due to the differences in data and production functions. But that is irrelevant, because it is the management outcomes as measured by performance statistics (related to stated objectives) that eliminate MPs that do not perform adequately from consideration.

I use the term (closed loop) feedback simulation to mean an algorithm that mimics a real fishery management system. Here, the OMs simulate future stock and fishery monitoring data that are used by MPs to update a sequence of catches. Those catches are iteratively applied to the OM accounting of the stock states at each time-step, i.e., each year of simulated data is input to the MP assessment method for output to the HCR and any meta-rules. This process is meant to mimic the (usually) annual cycle of data gathering and analysis that would be conducted for the actual stock and fishery (e.g., Cox and Kronlund 2008). The term Management Strategy Evaluation (MSE) is reserved for a participatory process that identifies objectives guided by policy and the needs of resource users. An MSE processe uses the results of feedback simulation to provide for a strategic choice of an acceptable management procedure by eliminating those MPs that do not perform. The accepted MP is applied to the real stock and fishery. The term forecast means a stochastic projection of future states that does not require an iterative calculation and updating of OM states (e.g., similar to the ICES "short-cut" methods).

Compliance with ICES PA and MSY approaches

I understand that the ICES reference points Blim and Bpa are intended to represent biomass levels related to impaired recruitment capacity, while MSY Btrigger is an OCP where a reduction in fishing mortality is invoked as spawning biomass declines. MSYBtrigger should be positioned at Bpa or higher. At least for short-lived species, MSYBtrigger may be set to Bpa, or perhaps as for WHM when there are insufficient data to reliably estimate MSYBtrigger. ICES also uses fishing mortality rate reference points with Fpa generally less than Fmsy; Fpa represents a threshold to fishing mortality rates associated stock decline and eventually collapse (Flim). Actions initiated as Bpa (or Fpa) is breached are intended to promote the necessary conditions for MSY implied by the MSY framework. MSYBrigger is considered a lower bound to fluctuations of SSB around Bmsy (at equilibrium, or with sufficient historical data to support estimation of the likely range of SSB variation around MSY?). As informative data accrue, updated values of MSY Btrigger are likely to be higher than Bpa which represents a lower boundary under the ICES PA approach.

Thus, it seems that compliance of the WHM analysis with the ICES PA and MSY approaches would at minimum:

1. require estimation/setting of values for Blim, Bpa, Fpa, and MSY Btrigger, and
2. rely on simulation evaluation of the various MPs to provide the evidence to reject poor choices of management options and identify those that might reasonably be expected to achieve the desired risk tolerance for breaches of limits and achieving target levels.

As a general comment, the authors of Pastoors et al. (2020) assumed that the reader has a working knowledge of ICES harvest policies and the implications of ICES software (SS3 adopted from the US, SAM, EqSim, etc.) with respect to model assumptions and structure. This choice may well be consistent with ICES practice and the terms of reference but can be daunting for the uninitiated. Reviewers external to ICES might struggle (as I did) to understand the policies and practices routinely applied by convention. It might be worthwhile to provide some contextual information on key features to orient readers external to ICES. Hence, I note in this review where more information may benefit a wider audience, including decision-makers and stakeholders who receive ICES advice. I note my lack of extensive familiarity with ICES standards and modelling methods as apology where I have made incorrect assumptions or interpretations.

***Recommendation 1**. Include a definition and/or references to key considerations for the ICES PA and MSY approaches in order to identify which elements are essential for compliance.*

Minimum requirements for simulation testing HCRs as defined by WKGMSE

The WKGMSE document (ICES 2013) outlines 9 requirements in section 6.1 and paraphrased here for the purposes of this review (***emphasis mine***):

SIM1. The OMs should be according to established standards for population dynamics models, and be ***sufficiently detailed to provide the information needed in the decision proces***s.

SIM2. All natural variation should be modelled as stochastic processes and ***where assumed stationary may warrant sensitivity analyses***.

SIM3. Autocorrelations and time trends should be considered.

SIM4. When deciding on parameters for distributions, ***the guideline should be to obtain a plausible range of realities*** e.g., plausible range of initial year class strengths.

SIM5. Observation model: If the short-cut approach is used, the variances should lead to a range of stock 'estimates' ***comparable with the statistical properties of the routine assessments***, including retrospective errors and autocorrelations over time and age.

SIM6. Does the rule allow sufficient action if the biology or management falls outside the assumed range?

SIM7. It is not always obvious how reality checks can be done, but ***to the extent possible conformance with historical experience should be demonstrated***.

SIM8. Leaving out sources of variability, or deviating from the routine practice in the decision process may be permissible, but if such simplifications may be questionable, ***sensitivity tests should be done***.

SIM9. Ensure that measures that are compared with ***reference points are derived the same way as the original reference points***.

The labels SIM1-SIM9 will be referenced below in text and the numbered recommendations.

Motivation

The analyses are in part motivated by concerns over stock condition that point to the need for a rebuilding strategy and possible adoption of a rebuilding plan (the strategy informs the plan, which may have steps and measures beyond those of the management procedure applied for rebuilding purposes). At the least, there appears to be a desire to manage the stock at a level higher than recently estimated biomass. Regardless of whether the need for a rebuilding strategy has been formally determined, it is certainly in keeping with a precautionary approach to fisheries management to prepare for that event before it is needed. A procedural paradigm is proposed for developing the (rebuilding) management strategy with emphasis on identifying the form of HCR that meets the constraints outlined above.

Previous Work

The authors usefully provide some background on previous work on the WHM stock. They point to the feedback simulation-evaluation completed by Cox et al. (2018) and note that analysis demonstrated a clear recovery potential under different fishing scenarios that depended primarily on recruitment assumptions and target fishing mortality. The authors cautioned that the starting conditions of the populations simulated by Cox et al. (2018) did not include uncertainty, and therefore the simulated performance of management procedures evaluated may be optimistic. However, they do not explicitly explain how OMs assumed by Cox et al. (2018) differed from the Stock Synthesis (SS) and SAM implementations they adopted, or how the latter two implementations dealt with the concern about starting conditions. Nor do they consider how adopting the ICES "short-cut" approach to the projections might also have produced results that are too optimistic by ignoring assessment errors and lags that will be inherent in the real management system.

Questions:

1. Would addressing the starting conditions in the analysis by Cox et al. (2018) have provided acceptable OMs (and would it have been a viable choice to address this shortcoming to preserve the advantages of full feedback simulation)?
2. Did the SS and SAM analysis address the concern of starting conditions?
3. The authors might consider commenting on the following question: Which is worse, a lack of uncertainty in the starting conditions or ignoring assessment errors and lags? The latter are likely to be important in actual application given the low current abundance and noted retrospective behavior of the assessments.

***Recommendation 2**. Describe how the SS3 and SAM implementations parameterize initial conditions in a manner that overcomes the lack of uncertainty in initial conditions in the analysis by Cox et al. (2018).*

***Recommendation 3**. Discuss the relative importance of stochasticity in the starting conditions versus not incorporating stock assessment errors and lags in the ICES short-cut approach used for evaluating rebuilding measures (related to SIM5, SIM7).*

3. Rebuilding Goals

I will begin by summarizing the goals of the rebuilding strategy as I understand them from Pastoors et al. (2020) and the plan described by Annex 1

1. Comply with the ICES precautionary and MSY approaches (described above).
2. Avoid biological limits. Although not explicitly stated in the papers I assumed this to mean that there should be less than a 5% chance of SSB falling below Blim in every year of the evaluation period as per Risk 3 of ICES (2013).
3. Achieve the rebuilding target defined by SSB greater than Bpa for 3 consecutive years within 10 years of initiating the rebuilding plan.
4. Comply with the MSY approach (*achieve MSY or proxy*). I assumed this to mean that the SSB should exceed MSY or proxy 50% of the time over some evaluation period.
5. Limit year to year TAC variability to 20% (achieved via a meta-rule defined in the MP).

Although the order of priority of objectives was not stated, I have listed them in the order I would consider them to apply.

These goals would benefit from two additional steps. First, as far as possible they should be made measurable by formally defining the outcome of interest, the desired certainty of achieving that outcome and the time frame for evaluation. Second, each measurable objective should have a matching performance statistic explicitly defined in mathematical notation. For example, goal (2) would need criteria to decide when SSB > Bpa, e.g., with at least 50% probability in each of the 3 years? Or simply that the estimated ratio SSB/Bpa > 1 in three consecutive years? I do, however, *commend* the stipulation that SSB > Bpa in 3 consecutive years because it may reduce the likelihood of a false declaration of the rebuilding target being achieved and hence too early an end to rebuilding efforts.

Questions:

1. Goal (2) requires further explanation by the authors because, as far as I can tell, performance statistics were not calculated that are related to this goal. If that is true, there is no basis provided for determining which HCRs fail to meet that constraint.
2. The 10-year time frame for rebuilding is not one of the four stanzas (current 2018-2020, short 2021-2025, medium 2026-2030, and long 2031-2040) defined in Pastoors et al. (2020). Were performance statistics for a 10-year time frame reported elsewhere? Or are the combined results for the "current", "short" and "medium" stanzas considered to apply?
3. The plan would seem to be considered successful if the stock achieved the rebuilding target in (goal 3) by year 10. If that happens, however, goal 4 can only be met if Bpa = MSYBtrigger. Is it possible that MSYBrigger might be updated within the 10 year time horizon of the plan to be a value greater than Bpa?

There are no equations that unambiguously describe the calculations for performance statistics (my experience is that different people may well have different understanding of the performance statistics without equations).

***Recommendation 4**. Clarify why the rebuilding proposal in Annex 1 identifies a 10-year time frame for rebuilding while the evaluations do not include performance statistics evaluated over the same time frame.*

Performance statistics described by Pastoors et al. (2020) include the first year of achieving rebuilding to Blim and Bpa. These statistics were calculated as the first year where the probability of being above Blim or Bpa was larger than 50%. However, there is no defined rebuilding objective related to this statistic although it is cited in the selection of the preferred MP as important. In Annex 1, requiring 3 consecutive years of SSB > Bpa is a different performance statistic but is not explicitly evaluated. Much of the short-term performance of the MPs relies on the expected strength of the 2014-2018 year classes. However, these effects maybe transient if that recruitment does not materialize at the expected strength or returns to pre-2014 lower levels in the very near future. Thus, MP selection on the basis of minimum time to Bpa alone (estimated to occur in 2021 under the preferred MP) may not be sufficient. It should also be recognized that MP performance is relative; one can only say that one MP performs better than another relative to an objective under a given OM scenario (whereas the true state of the stock is unknown).

***Recommendation 5**. Provide a formal listing of (measurable) rebuilding/management objectives relative to the ICES PA and MSY approaches and catch goals. Each objective should have a matching performance statistic defined with equations. Propose an order of priority for the objectives so that it is clear how MPs are eliminated from consideration because (for example) they do not meet imperative objectives or provide unacceptable trade-offs of management outcomes.*

***Recommendation 6**. Restate Goal 4 as a measurable objective related to achieving MSY or proxy as a measurable objective. Be clear on how Goal 4 applies if the rebuilding target of Bpa is achieved late in the 10 year rebuilding time frame. It could be that Goal 4 provides a means of considering how management transitions from a rebuilding mindset to harvest at target levels.*

4. Supporting Analyses

Stock ID

The main conclusions of the genetic work relevant to stock assessment is that the northern-southern population boundary is about the middle of the coast of Portugal. The authors conclude that this distribution implies that the current division between western and southern horse mackerel is not adequate as the northern part of 9a is currently included in the southern population.

However, there is no further discussion of the issue in the document or how it might affect the performance of the tested MPs. First, a map might be helpful showing the distribution of the stock and the distribution of catches and effort. Second, at least some scenario planning around how scale mismatch of control created by stock and fishery monitoring data not matching the stock distribution might be worthwhile, at least for qualitatively assessing the utility of MPs and suggesting next steps.

Questions:

1. How important is the scale mismatch of control problem created by the stock and assessment area not overlapping?
2. Is it something that could compromise model assumptions and the attainment of the necessary conditions for a successful MP as described above (and therefore need to be investigated in the as the rebuilding/management plan is updated?).

***Recommendation 7**. Discuss the expected effects on assessment/management due to possible scale mismatch of control indicated by the genetic analysis. At this point the discussion may be qualitative rather than quantitative, but the issue of importance is whether proposed management measures are likely to be compromised.*

Length Composition of the Catches

The authors note that "In the current SS assessment framework, length compositions are used as the key metric for catches in combination with age-length keys to generate age compositions dynamically." They are unfortunately forced to report that these data are not readily available from some sources due to variation in data archiving practices over time.

Simulation of compositional data is very challenging. The usual outcome is that simulated age- or length-composition data are too informative, and do not have realistic bias or errors adequately specified. The ICES short-cut methods would not use simulated length data in the assessment method of the MP but full feedback methods used in future could. This may seem to be an advantage of the short-cut method, but I wonder if it would be equally difficult to adequately mimic the behavior of SSB and recruitment over time that would be expected from an age- and length-informed OM? I would expect that at least some recruitment information is being contributed by the length composition data obtained from catches that are available (see figure below).

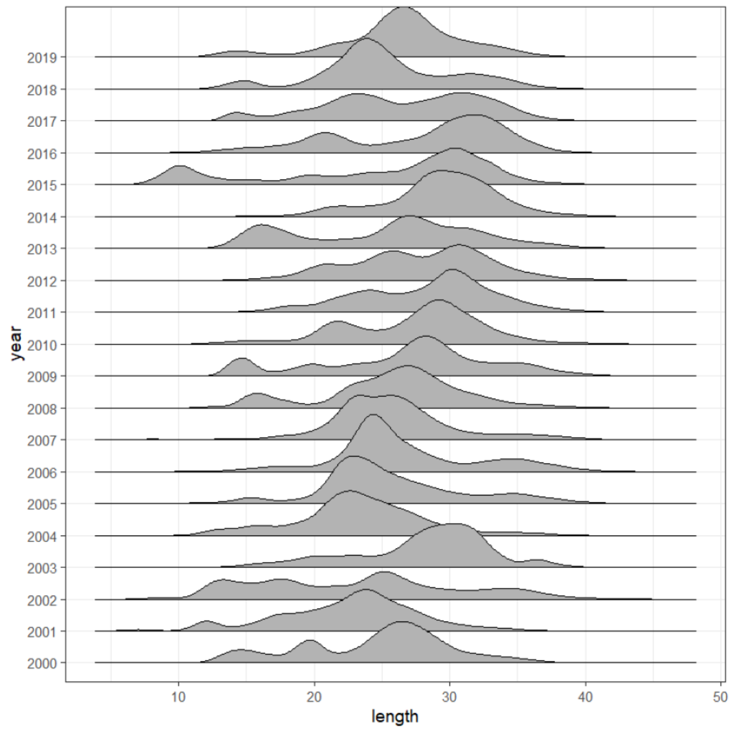


Figure 7.2.10.2: Western horse mackerel. Stacked length frequency distribution of the catch data as used in the assessment model.

I note underfitting of recent index values in the SS modeling from WGWIDE (2019, figure below) which may be due to a conflict with the length composition data of the catch, which themselves do not appear to be fit well (Figure 7.2.11.1 of WGWIDE 2019). I note this because of the use of a "short-cut" method in the projections that are unlikely to capture the (restrospective) assessment errors that have been found previously, nor the effects of analysts trying to intervene each year to address lack of fit due to data conflicts. It could be that a rather simple MP that avoids the need for such intervention might be worth considering (but not blindly, as robustness to dangers like hyperstability of abundance indices needs to be explored via OM scenarios as per SIM4).

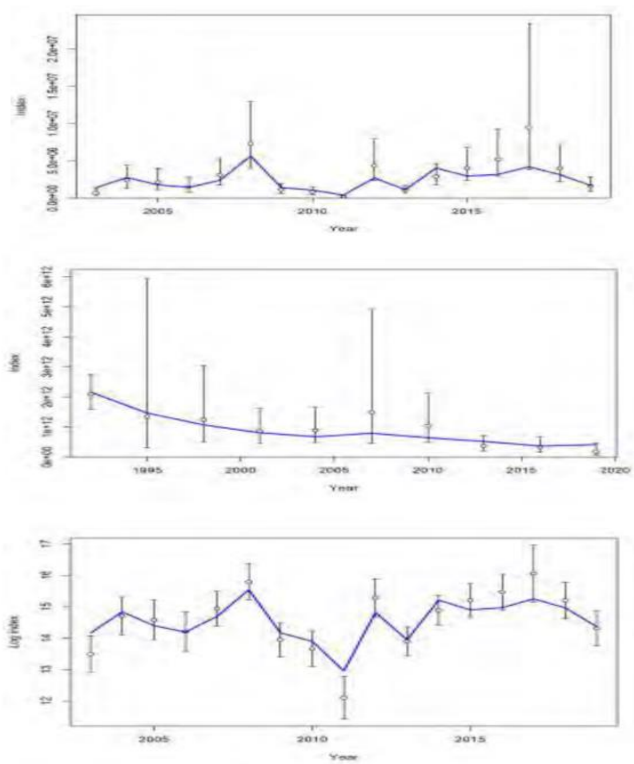
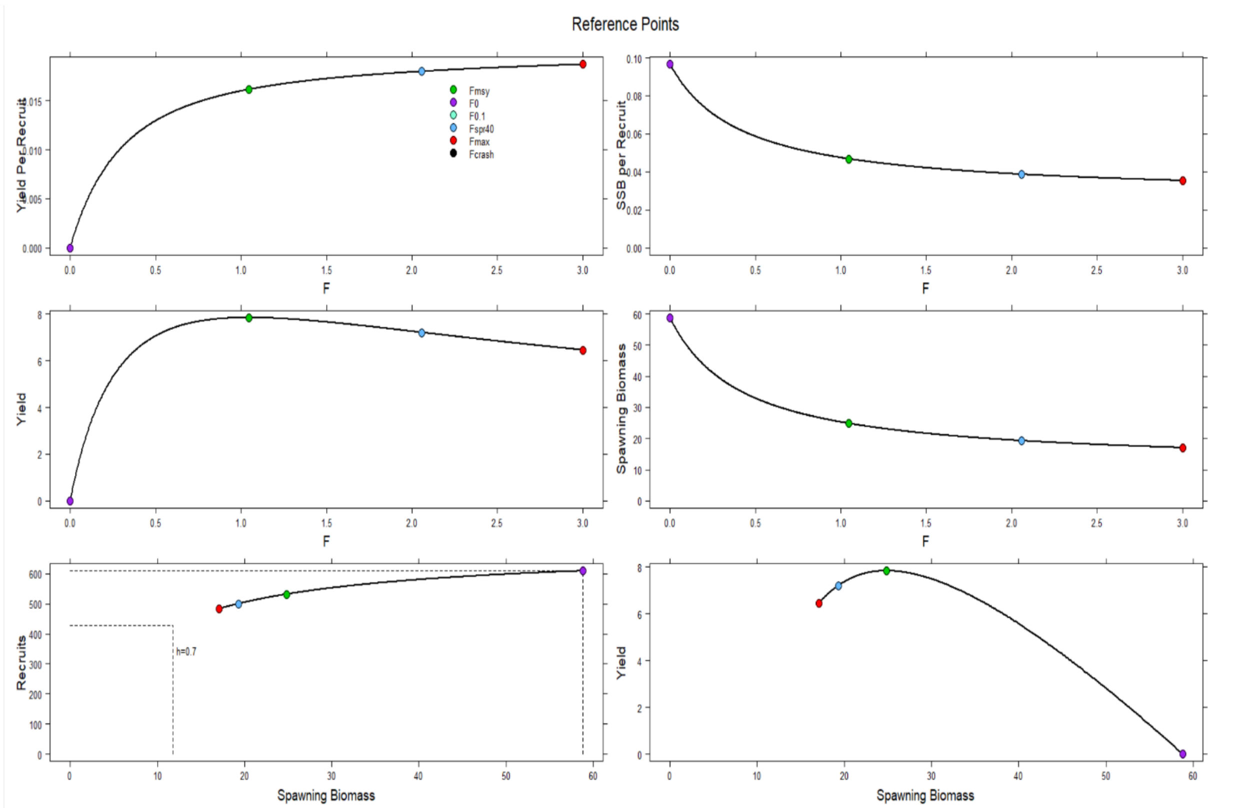


Figure 7.2.11.1: Western horse mackerel. Model fitting. Fitting of the model to the fisheries-independent indices. From top to bottom: IBTS, egg survey, PELACUS.

SSB per Recruit Analysis

What is being added by this analysis that is not calculated (or could potentially be calculated) by using the OM, with the advantage that if alternative OMs are considered then the per recruit analysis could reflect those hypotheses? Figures such as the example below for a given OM configuration can provide a useful check of whether results “make sense”. Could similar conclusions about the lag in recruitment and increased SSB be inferred from the age at 50% maturity or generation time calculated within the OM using Seber’s (1982) generation time equation?



5. Approach to the HCR Evaluation

The evaluation for WHM essentially has two operating models (based on WGWIDE19 SS3 using Stock Synthesis 3 and WGWIDE19 SAM implemented in SAM) and two different forecasting algorithms (EqSim and SAM HCR). The latter are of the ICES "short-cut" methodology. Thus, the assessment method component of the tested MPs does not necessarily mimic the behavior that could be expected from application of an annual assessment model to the real stock. This issue may be important for WHM given the history of retrospective behavior and the difficulty of fitting available length composition data (and possibly future signal the challenges of incorporating length composition data that may be available in future). Only EqSim was applied to the results from SS3 and not SAM HCR. The HCRs varied in some, but not all, of the OCPs (Btrigger is not varied). Where applicable the operating model reference points were adopted as the OCPs and note that MSY Btrigger=Bpa. The latter choice might typically be applied to short-lived species but as WHM was characterized as being relatively long-lived by the authors this choice may warrant a short explanation. The projection period for evaluating MP performance was 23 years (to the year 2040) with consideration of four stanzas (e.g., current, short, medium and long-term as described above and in Pastoors et al. 2020). Note again however that Annex 1 identifies 10 years as the rebuilding time frame.

Stock Assessment

Stock Synthesis (SS) and SAM software are applied to the stock and fishery monitoring data. The primary difference in terms of data choice appears to the be inclusion of length composition data in the SS analysis.

***Recommendation 8**. Provide a table that summarizes all data types (with years indicated for temporal data) for each of the SS and SAM OM implementations as per the example (related to SIM1).*

Regarding the recommendation above, included below is an example from the 2020 Gulf of Alaska Sablefish assessment:

Source	Data	Years
Fixed gear fisheries	Catch	1960 - **2020**
Trawl fisheries	Catch	1960 - **2020**
Japanese longline fishery	Catch-per-unit-effort (CPUE)	1964 - 1981
U.S. fixed gear fishery	CPUE, length	1990 - **2019**
	Age	1999 - **2019**
U.S. trawl fisheries	Length	1990,1991,1999, 2005 - **2019**
Japan-U.S. cooperative longline survey	CPUE, length	1979 - 1994
	Age	1981, 1983, 1985, 1987, 1989, 1991, 1993
Domestic longline survey	CPUE, length	1990 - **2020**
	Age	1996 - **2019**
NMFS GOA trawl survey	Abundance index	1984, 1987, 1990, 1993, 1996, 1999, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019
	Lengths	1984, 1987, 1990, 1993, 1996, 1999, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019

In each case, the software was configured to use the same data (except for the length composition data) and efforts to resolve fitting issues such as retrospective patterns are briefly described by the authors.

Question:

1. What is the question being addressed by the distinction in models? If the question is to examine the influence of length composition data, a more self-consistent analysis could be conducted by using only SS with the length composition data “turned on” or “off”. The concern here is that differences in perceptions about the stock reconstruction could depend not only on the length composition data, but also on differences in model assumptions between the SS and SAM implementations. The results suggest some large differences in the estimated values of reference points due to model choice as noted by the authors.

Here is where a reviewer faces a challenge in evaluating the implications of software choice. The SS3 and SAM implementations differ in assumptions and parameterization. However, there is no easily accessible summary of the differences. It would be helpful to provide a listing of key structural elements of each parameterization and any fixed parameter values (see examples of tables of notation and structural equations in Cox et al. (2018). Admittedly, a reviewer fully versed in ICES methods might have a clear understanding of the implications, but an external reviewer is unlikely to have such “insider” knowledge and the general descriptions of the software provided throughout the paper do not help a great deal (e.g., Section 8.2 points to code standardization and improvements, but does not provide the key equations that might indicate how recruitment deviations are simulated, how the OM is updated, or equations for the HCRs and meta-rules).

Recommendation 9*. Include a comparative table of notation and key structural differences between the SS and SAM model configurations (not the software control files but the actual stock and fishery dynamics, e.g., see Cox and Kronlund (2008) and Cox et al. (2018). This material would help satisfy SIM1.*

Pastoors et al. (2020) relies on reference to previous analyses, and appropriately so. However, it would be useful to include at least some figures that show key data or model estimates related to conclusions in the document, and particularly those that describe the differences in the SS and SAM fits. It might be useful to augment Figure 3 of the document by constructing a multi-panel plot that, in addition to the time series of estimated SSB, also shows the following:

- Catch (landings?) time series and estimated biomass overlaid,
- Stock index or indices,
- Estimated recruitment time series (or the log recruitment deviations)
- Estimated fishing mortality time series.

Such multi-panel plots can provide a “story board” of the model interpretation of observed and estimated time-series that can help to communicate how each model is integrating the removals, compositional data, and stock indices. An example from a Pacific Herring analysis (Kronlund et al. 2018) is show here to illustrate what might be helpful:

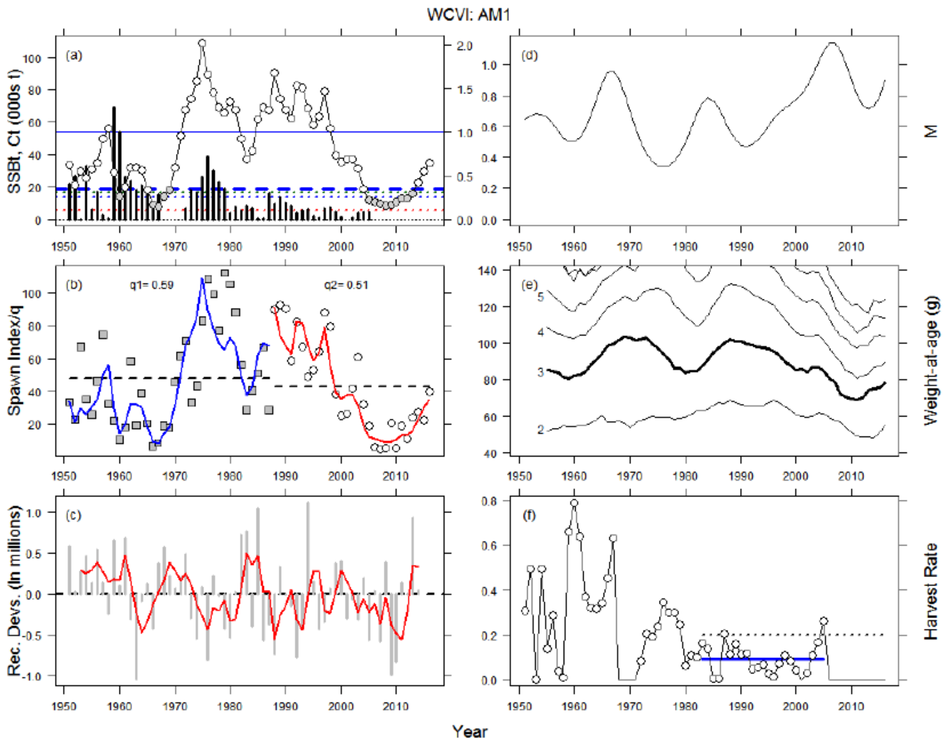


Figure 3. Assessment model AM1 stock reconstruction for WCVI Pacific Herring. Panel (a) shows the 1951-2016 time series of estimated spawning biomass (circles) and catch (bars). The lower 20% of spawning biomass estimates are shaded grey. Reference lines are shown at estimates of $0.1B_0$, (red dashed line) $0.25B_0$ (blue dashed line), $0.3B_0$ (green dashed line), the 1996 fixed cutoff value (thick blue long dashed line), and unfished spawning biomass (solid blue line). Panel (b) shows observed surface (grey squares) and dive (open circles) survey indices scaled to spawning stock biomass. A trailing 3-year moving average smoother indicates trend for the surface (solid blue line) and dive (solid red line) survey indices. The horizontal dashed line is positioned at the mean observed value for each survey series. Estimated q_1 (surface survey) and q_2 (dive survey) are shown in the panel. Recruitment deviations (grey bars) from a Beverton-Holt stock-recruitment function are shown in panel (c). A 3-year trailing moving average smoother (red line) shows the trend in deviations. Estimated natural mortality is shown in panel (d). Panel (e) shows a 3-year trailing moving average of observed weight-at-age for age classes 2-6. Estimated harvest rates are shown in panel (f). Reference lines are shown at the intended harvest rate of 0.2 (horizontal dotted line) and at the average harvest rate from 1983 to the first year of commercial fishery closure or 2016. All model estimates are the maximum likelihood estimates (DFO 2016a).

Similar plots can also help to assess the importance of possible non-stationarity as per SIM2 (e.g., there seems to be some non-stationarity in weight at age for WHM as indicated in Fig 7.2.5.1 of WGWIDE 2019).

Recommendation 10*. To support the description of the differences in OM conditioning, consider including figures in addition to SSB vs. year to illustrate how each of the SS and SAM implementations are interpreting the stock and fishery monitoring data.*

Estimation of Reference Points

The authors note that the "…reference points for the SAM assessment are based on the 2018 assessment. Bpa and Blim are lower than the values for the SS assessment, while the Fmsy is higher." This is unsurprising given reference points can change due to updating of data (2018 vs. 2019), when new data are introduced (e.g., use of the length composition data), and because of changes in model assumptions (SS vs. SAM). Here there would be benefit in summarizing the key differences in model assumptions recommended earlier. Plotting the production function from each model might be helpful for the MSY reference points as well as comparative figures (F as a function of SSB) of the limit, trigger and "pa" reference points. The effect of year is minor as noted by the authors; the performance of MPs will be more affected by the differences in OMs (SS vs. SAM). Absolute values of reference points are reported.

Questions:

1. What is the ICES process when a change to operating model rescales the stock reconstruction and reference points?
2. How often would reference points be revised over the 23-year evaluation period of Pastoors et al. (2020) and the 10-year period of Annex 1?
3. Related to (2), are reference points for the 23-year evaluation period fixed at the absolute values or recalculated at each time step?

A danger of focusing on absolute reference points is that decision-makers and stakeholders can get fixated on the specific values which are strictly a consequence of the data and model assumptions used to calculate their values at each assessment year. For example, changes in model assumptions changes can alter the status determination of the stock (e.g., from needing rebuilding to above Blim) without any change in data.

***Recommendation 11**. Consider qualifying estimates of reference points with the definition of the reference point (or a reference to the method) and year the estimate was calculated. Take care to condition decision-makers and stakeholders that the scale of estimates changes with updated data, the addition of new data not previously used (e.g., forthcoming length-composition data), and changes to model assumptions as illustrated by differences in the SS and SAM reconstructions).*

Simulated States

Given recruitment for this stock, like most others, is key determinant of management performance I might have expected some evaluation of whether the biomass projections show behavior consistent with the historical reconstruction of the operating models (see SIM5), as well as some alternative scenarios for autocorrelation to mimic various phenomena (e.g., pulse behavior of large year classes, consideration of any depensatory effects at low abundance if plausible).

In addition to the *F*=0 scenario included in the WHM evaluations, another potentially useful simulation to check model behavior and provide a benchmark for assessing relative performance of MPs is a "perfect information" scenario (perhaps related to SIM7 as a reality check/bound on other scenarios).

Such analyses would be consistent with ICES (2013) guidance on validation that note expectations to confirm that the statistical properties of simulated data are consistent with the behavior of the conditioned operating model (note that not all the guidance is applicable to the short-cut approaches). In particular the guidance suggests:

- Run the evaluation with zero F in future to check the behaviour of the population model.
- Run the management decision model with perfect knowledge, and compare this with the management decision model with assessment error included to check the impact of this assessment error. It may be that the management plan is not precautionary even under perfect knowledge. This is also useful as a code check.

- Justify the approach used to characterise noise in parameters used when approximating the assessment model for short-cut MSE by making use of “reality” checks to ensure future noise is consistent with historically observed noise.
- Run the evaluation by forcing *F*s to be in the range of *F*s experienced historical in order to check how the properties of the assessment model in the loop compares with its historical behaviour in practice.

Recommendation 12*. Add material (figures, tables) that illustrates that the properties of projected biomass are consistent with historical estimates as per ICES (2013, SIM2, SIM5) (also see Weidenmann et al. 2015 for suggestions).*

HCR Evaluations

The description of HCRs in Section 8.1 of the document is clearly presented. As noted above, Btrigger was either not defined (Constant *F* strategy) or was not varied. A range of target *F*’s were tested that included *F*=0 as a reference level. Meta-rules were applied to impose a TAC floor (50 kt) or a maximum 20% variation in TAC above Btrigger. Presumably, the latter was implemented by trimming the output from the HCR at +/- 20% variation from the previous year, although this is not stated. It is not clear however, if one of the proposed HCRs mimics the status quo procedure, at least approximately. This would be useful for determining if proposed MPs can be expected to offer an improvement to status quo, or if status quo itself is capable of acceptably meeting objectives.

Recommendation 13*. Either identify which HCR is intended to approximate status quo management or if possible, add an HCR to represent status quo management (can the effective HCR be inferred by examining historical performance even if no formal HCR exists?).*

A priori, it can be expected that the Constant F rule would have the lowest AAV (Absolute Annual Variability) though it does not include a “precautionary ramp” to promote stock growth as biomass approaches low biomass. The “Double Break-Point” rule might have the highest AAV across all stock levels due to the inflection points at Btrigger and Blim.

I suggest that this section of the paper could benefit from some restructuring. First, it is not clear what exact measurable objectives and matching performance statistics are being considered in MP rejection/selection. The reader has only been told that an acceptable MP needs to meet the ICES precautionary approach requirement (assumed here to be less than a 5% chance of breaching Blim in any given year) and the MSY approach requirement. There are four stanzas for calculating performance (e.g., current, short, medium, long) but no indication of what objectives might be of interest relative to each stanza. Trade-offs in performance statistics are not plotted for easy comparison and should be presented graphically as well as in tabular format as discriminating the performance among MPs is the essence of decision-making in this analysis.

Selection of preferred HCRs for Western Horse mackerel

The authors report that PelAC selected the following preferred option for the Western horse mackerel rebuilding plan:

- Evaluation method: EqSim

• Assessment: Stock Synthesis (WGWIDE 2019), because this is the basis for the assessment and advice.

• Target fishing mortality at Fmsy = 0.074 (approximated by 0.075 in the simulations)

• Blim at ICES Blim (834 480 t)

• Btrigger at ICES MSY Btrigger (1 168 272 t)

• Double breakpoint rule with 20% constraint on IAV above Btrigger

• Minimum F when stock is below Blim at 20% of Fmsy = 0.015.

Pragmatism of process appeared to guide the choice of assessment rather than a choice based on plausibility of the SS OM ("Given that the EqSim with SS3 evaluation is closed to the ICES advisory practice, this was used as the basis of the suggested rebuilding plan by the PELAC"). This choice, while pragmatic in terms of improving conditions for adoption of the proposed plan, does forfeit a possible benefit of two hypotheses about stock and fishery dynamics, namely assessing robustness of successful MPs to the true underlying state of the stock (related to SIM4).

The HCR design appears to have been selected primarily based on minimum time to Blim and Bpa, respectively. Hinging MP choice on that single criterion (after meeting the requirement to avoid Blim) is a convenience allowed by the expected strength of the 2014-2018 recruitment but has less general applicability over a range of stock conditions that might be encountered over a 10-year (or 23-year) time horizon. A rebuilding strategy should be likely to provide acceptable performance (or at least prevent disaster) even without the happy event of recent above average recruitment. Average catch and catch volatility received less attention in the rationale but are important trade-off outcomes for stakeholders. Thus, the analysis is essentially a best assessment approach (SS3) with an approximation to the intended MP based on a short-cut method.

Author Preferred HCR:

The TAC setting mechanism during the rebuilding plan shall be as follows:

a. When the stock (SSB) is estimated to be below Blim in the assessment year, the TAC will be fixed with a fishing mortality equivalent to 20% of Fmsy = 0.015.
b. When the stock (SSB) is estimated to be between Blim and Bpa in the assessment year, the TAC will be fixed with a fishing mortality equivalent to:
c. 0.015 + (SSB-Blim)/(Bpa-Blim) * (Fmsy-0.015).
d. When the stock (SSB) is estimated to be above Bpa in the assessment year, the TAC shall be fixed with a fishing mortality equal to Fmsy (0.074), subject to the constraint that the change in TAC compared to the current (assessment) year does not exceed 20%.

Given the selection of preferred MP (HCR) is hinged on SS as an OM, there is no consideration of robustness to misspecification of the true stock and fishery dynamics. It would have been helpful to have seen the entire analysis done using SS, but with inclusion/exclusion of the length-composition data and full feedback simulation of 1-2 MPs after 'pre-screening' using the short-cut methods (if necessary). This would at least have provided a starting point for subsequent updates of the rebuilding/management plan that could expand the consideration of robustness. Again, the effects of the interaction of SS as the assessment method in the MP and the preferred HCR is unknown since that combination was not simulation tested.

Nevertheless, there is a need in contexts where rebuilding is (very nearly) mandated not to defer action. In this sense it might be acceptable to adopt the preferred MP as an interim procedure *if it can be clearly demonstrated to be superior to status quo management in simulation*. A requirement of an interim procedure is that it have a reasonable expectation of promoting stock condition that are better and not worse than currently exist and ideally allow data needed to improve a rebuilding strategy to accrue.

6. Rebuilding Plan (Annex 1)

Annex 1 Plan Terms (numbering and ***emphasis mine***)

1. "Rebuilding plan timeframe": the timeframe for achieving the rebuilding plan target is a ***maximum of ten years***, although ***all attempts will be made to realise that target within five years.***

2. "Rebuilding plan target": when the spawning stock biomass is greater than Bpa for a ***minimum of three consecutive years***.
3. The applicable biomass reference points for the Western horse mackerel stock shall be as follows: Blim = 834 480 tonnes and MSY Btrigger = Bpa = 1 168 272 tonnes. It should be noted in case of this rebuilding plan the value of MSY Btrigger is identical to Bpa and should be read as one wherever mentioned in the text. Should this relationship change in the future ***the plan is no longer valid***. The maximum fishing mortality associated with Maximum Sustainable Yield (Fmsy) for the Western horse mackerel stock shall be as follows: Fmsy= 0.074. These values are based on the 2019 inter-benchmark report (ICES 2019).
4. ***The rebuilding plan may be superseded by a long-term strategy for the stock*** when, according to ICES, the spawning stock biomass is assessed to have been above Bpa for three consecutive years.
5. Should any other underlying assumption, or the definitions of the stocks in Article 2, of the rebuilding plan change based on new scientific knowledge ***this rebuilding plan will be deemed no longer to be applicable***.

Plan item (1) above may be problematic. First, as noted above performance for a 10-year time frame was not evaluated (what measurable objectives needed to be satisfied over that period?). Second, adding a clause that "all attempts will be made to realize that target within five years" implies some management change to make that happen; if the preferred MP does not accomplish that outcome what is the process for adapting the rebuilding strategy within the plan?

I commented on the need to define plan item (2) previously.

I suggest that rebuilding strategies should encourage adaptation of rebuilding strategies (an approach to review the objectives, and an adjustment of these if the objectives are not being achieved) as new understanding, new data, and new analyses arise. To simply state the plan is no longer valid if reference points change (items 3 and 5 above) is a almost a guarantee the plan will be declared invalid. We know that reference point estimates will change simply by updating data. Instead, the plan might anticipate the need to evolve by including advice on determining how frequently the rebuilding strategy should be evaluated for progress which may vary according to:

1. Time-prescribed interim objectives agreed to in the development of the rebuilding plan.
2. Life history (short-lived fish may require more frequent progress evaluation than long-lived species).
3. Schedule of anticipated data collection or availability of new data or analytical resources for updating assessments or simulations.
4. Exceptional circumstances such as unexpected data, or new understanding of the stock and fishery.

***Recommendation 14**. Rebuilding plans are not "set and forget" and stocks seldom rebuild according to the intended plan. The plan should build-in opportunities for adaptation as new data, new understanding, and new analyses are conducted. Interim milestone should be specified to evaluate progress and check for exceptional circumstances that would warrantee revision of the rebuilding strategy that informs the plan (related to SIM6).*

Finally, for item (4) rebuilding strategies intended to recover stocks are often presented as separate from management strategies. However, invoking mechanisms for rebuilding a stock in isolation of the overall management strategy can lead to:

1. Deferral of action until thresholds (e.g., Blim) to serious or irreversible harm are breached; or
2. Conflict between rebuilding measures and those measures intended to provide harvest opportunities, such that recovery efforts are thwarted or delays in rebuilding to target levels are incurred.

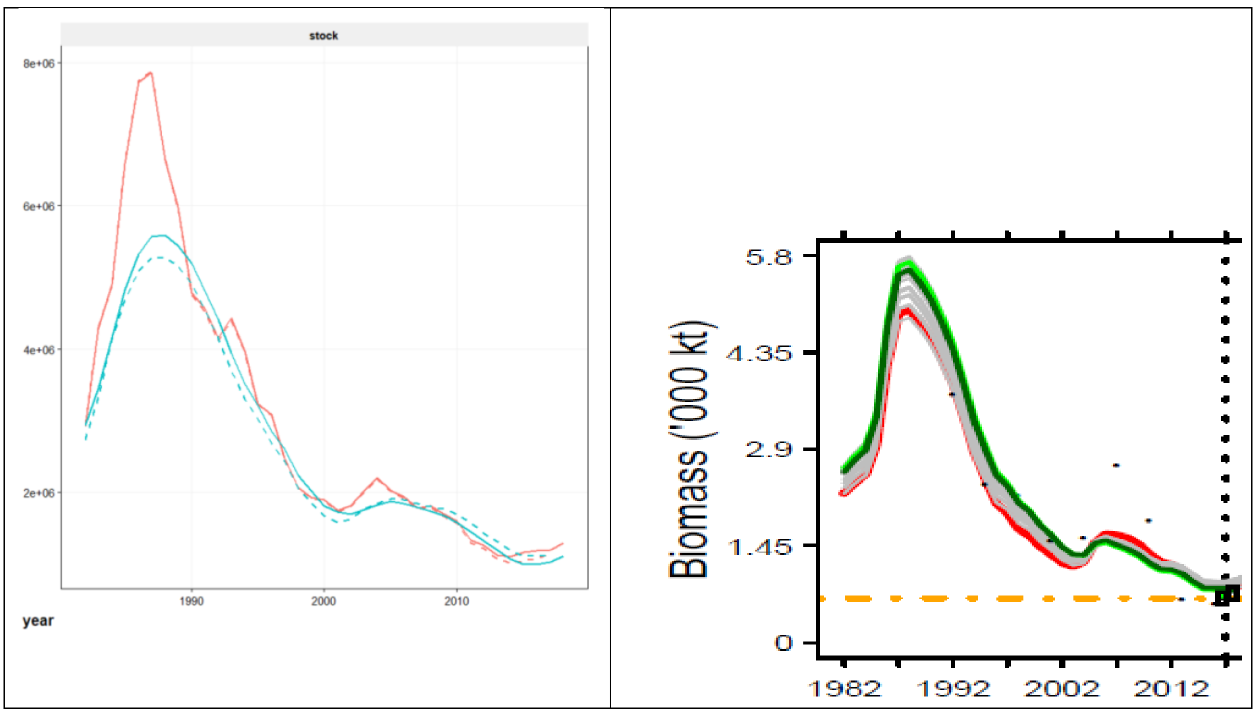
In fact, separation of rebuilding scenarios from the overall management strategy design is not helpful to achieving sustainable outcomes for a stock and the dependent fisheries. First, the management strategy should aim to avoid approaching a situation where rebuilding is needed. However, a stock can decline despite this intent; such a situation should be anticipated and revisited as a stock approaches limits so that current, rather than average, conditions are considered. Under a precautionary approach to managing fisheries, a rebuilding strategy should be considered integral to the design of a management strategy. Therefore, consideration of management at target levels should not be deferred in the development of a rebuilding strategy.

7. Summary

The authors have added to the previous literature that indicates rebuilding potential for WHM. Their analysis relies on HCR evaluation via ICES "short-cut" methods; they did not conduct a full feedback simulation of MP performance such as that completed by Cox et al. (2018). Such short-cut approaches do not produce the full range of output of a stock assessment model and may not capture complex feedbacks between the state of the system and the variance, bias, and correlation of errors (Weidenman et al. 2015).

This is my most serious concern with applying the preferred MP. There is no discussion in the document of how well the ICES "short-cut" approach might reproduce critical behaviors that can kill stocks; these relate to the relationship between assessment errors, the HCR and management system lags. These behaviors can be exacerbated by the interaction of MP components or, better still, if diagnosed in simulation they can often be "tuned out" by deliberate MP design. Evaluations of the HCR alone do not necessarily capture the interaction of the HCR and assessment method.

There are reasons noted by the authors to be concerned about the effects of the assessment method, namely the retrospective issues and consequently the amount of "manual tuning" that might be required at each stock assessment step to achieve statistical fit. For that reason, would there be value in considering the analysis of Cox et al. (2018) further by updating the approach to resolve the concern over initial conditions? The initial conditions issue seems no more serious than having to fix certain selectivity parameters (Executive Summary, page 2) or forfeiting the introduction of lags and assessment errors by adopting a "short cut" approach. And in fact, the stock reconstruction of Cox et al. (2018) and that produced by SAM do not seem markedly different (SAM in blue on left, Cox et al. (2018) on right):



Having said that I appreciate that ICES standards may require specific model implementations to be used as input to the advisory process. Time constraints to deliver an analysis can conflict with computing requirements of a full-feedback approach. This may preclude a full suite of feedback simulations being completed over all OM and MP combinations of interest. I also understand that analyses such as the "short-cut" approach can offer a useful sketch of possible behavior of a management option. If there is no intention to proceed to a full feedback simulation analysis as the next step, then I recommend a full feedback analysis of the preferred HCR and actual assessment model to confirm that performance remains acceptable. The closer the MP mimics the actual procedure applied to the real stock and fishery the better.

***Recommendation 15**. If applied, test the preferred MP in a full feedback simulation using the stock assessment method planned for the MP as soon as feasible to check whether the interaction of the assessment and the HCR causes the MP to fail in simulation.*

The issue for a rebuilding strategy is determining what to do with surplus production when it becomes available to meet rebuilding and management objectives. Selecting a MP to accomplish this task is a actually a process of rejecting those MPs that are likely to fail in practice; there is no guarantee a given "preferred" MP will work. However, simulation-evaluation of the MP intended for use over a range of OM scenarios hopefully closes the gap between precaution in theory and practice.

8. Minor Comments

1. Please provide descriptive captions for all tables and figures.
2. Phrases such as "ensure MSY" are quite bold, given one cannot ensure that MSY will be achieved. It would be fair to restate such goals as "identify management procedures that can be reasonably expected to acceptable meet stock and fishery objectives." Such objectives may include achieving a target of Bmsy (50% of the time over some specified evaluation period).

9. References

Cox, S. P., Benson, A. J., Doherty, B., and Johnson, S. 2018. Evaluation of potential rebuilding strategies for the Western Horse Mackerel (*Trachurus trachurus*) fishery. Landmark Fisheries Research, Ltd. 46 p.

Kronlund, A.R., Forrest, R.E., Cleary, J.S., and Grinnell, M.H. 2018. The Selection and Role of Limit Reference Points for Pacific Herring (*Clupea pallasii*) in British Columbia, Canada. DFO Can. Sci. Advis. Sec. Res. Doc. 2018/009. ix +125 p.

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

ICES. 2020. Working Group on Widely Distributed Stocks (WGWIDE). ICES Scientific Reports. 2:82. 1019 pp. http://doi.org/10.17895/ices.pub.7475.

Seber, G.A.F. 1982. The estimation of animal abundance and related parameters, 2nd edition. Edward Arnold, London.

Wiedenmann, J., Wilberg, M.J., Sylvia, A. Miller, T.J. 2015. Autocorrelated error in stock assessment estimates: implications for management strategy evaluation. Fish. Res. 172: 325-334.

Review of

Pastoors, M.A., Campbell, A., Trijoulet, V., Skagen, D., Gras, M., Lambert, G.I., Sparrevohn, C.R., Mackinson, S., and Ourens, R. 2021. Updated Diagnostics and Simulations of the Western Horse Mackerel. Dated March 22, 2021. *Supplement to*:

Pastoors, M.A., Campbell, A., Trijoulet, V., Skagen, D., Gras, M., Lambert, G.I., Sparrevohn, C.R., and Mackinson, S. 2020. Report on Western Horse Mackerel Technical Focus Group On Harvest Control Rule Evaluations 2020. Report for Pelagic Advisory Council. 46 p.

and

ANNEX I: PELAC proposal for a rebuilding plan for Western horse mackerel. Dated 28 July 2020

Prepared 26-Mar-2021

A.R. Kronlund (Rob)

Interface Fisheries Consulting, Ltd.

E-mail: interfacefisheries@gmail.com

Summary

The "update document" was produced in response to the review of Pastoors et al. (2020) at WKWHMRP in February 2021. The document contains supplementary outputs, new simulations, and enhanced descriptions of various topics identified during the February workshop. The authors have completed considerable work that addresses the review comments; as a reviewer it is encouraging to participate in a process where clear efforts are made to improve harvest advice following review. Most of my comments are editorial in nature and can be addressed in revision as documents are finalized. I will pass a marked version of the document to the authors for their consideration.

Terms of reference (TOR) that guided this review are included below (***emphasis mine***):

a) Review the approach (i.e., analytical methods, application and interpretation) described in the Pelagic Advisory Council (PelAC) report (listed below) for the evaluation of the proposed Harvest Control Rule (HCR) in the rebuilding plan. The review should consider:

- i) Whether the tools used (methods), and the model conditioning done (data), are appropriate for the stock;
- ii) Whether the ***minimum requirements for simulation testing HCRs***, as developed by WKGMSE process, are met;
- iii) The appropriateness of the ***criteria used to draw conclusions*** on the performance of the HCRs.

b) On the basis of (a), determine whether the HCR evaluations presented in the PelAC are sufficient to evaluate the proposed rebuilding plan against ***precautionary criteria*** and the ***rebuilding plan targets and measures*** detailed in Article 5 of the proposed rebuilding plan.

c) Depending on the outcome of (b), either:

- i) analyse the results of the HCR evaluation, and develop conclusions on whether the proposed rebuilding plan ***can be considered precautionary and be used as the basis for ICES fishing opportunity advice*** for the stock; **or** ii) Propose ***additional analyses or diagnostics*** that would allow for sufficient evaluation of the proposed plan

d) Should the proposed plan include elements that are in contradiction with ensuring that the stock is fished and maintained, also in the future, at levels which can produce MSY, comment specifically on such elements, and ***their consequences for ensuring MSY***.

e) Deliver a report containing the key decisions and conclusions in relation to the TOR.

As a result of the additional work, the authors have met ToR elements (a), (b) and (c) as I understand the ICES requirements. Recalling my initial review, I would consider the author's preferred HCR to be an acceptable interim choice subject to the ongoing evaluation of performance cited in Table 6.1 of the update document, specifically the comments on checking progress of the rebuilding plan. I say "interim" because rebuilding plans are not "set and forget". They need a planned schedule for evaluating progress and allowing adaptation as new understanding, data, and analyses accrue. This will be important for Western Horse Mackerel as more is learned about the strength of the 2014-2018 year classes. And I restate that regardless of whether the ICES requirement for a rebuilding strategy has been met, it is precautionary management to prepare for that event before it is needed. The authors should be commended for striving to achieve that goal.

Editorial Comments

Generally, throughout the document provide complete explanations in figure and table captions and define first occurrence of acronyms. I expect the authors were pressed for time given the amount of work completed and this issue can be addressed in a final revision.

p. 10. Suggest that each of the bulleted scenarios be given a label (e.g., "BaseCase", "HalfBaseCase") for easy reference and recall of their meanings.

p. 13, Item 11. Need to be clear on what is "acceptable". I might assume this to be a less than a 5% chance of a Blim breach in each and every simulation year once the stock is rebuilt, where rebuilt may be achieving a 50% chance of being above Bpa in three consecutive years. I would suggest adding a very clear statement of how the (proposed) "rebuilt" criteria are defined. I appreciate that ICES may not have policy established in this regard, so that the authors need only be clear on what they mean and point to precedent or suggested guidance (e.g., WKREBUILD?).

Figure 2.1. Very helpful for the naïve reviewer.

Figure 2.2. Are observed or modelled age proportions shown? Age class coherence looks almost too well-behaved to be observed but clarification would be helpful. I might add a vertical line where the method changed from presumably random samples to ages derived from an age-length key (if I understand correctly what occurred).

p. 20. Thanks for defining that the model year begins in January. The description would be completed by noting that, for example, biomass is beginning of year biomass if that is the case.

Figure 2.6 is a useful argument *against* absolute reference points. Clarify that retrospective peels are being shown in the figure as opposed to plotting the results obtained in each historical assessment year.

p. 22. Since the 1982 recruitment is a 1-in-40 year event, what is the occurrence of "similar" recruitments in the 1000 simulated SSRs?

Section 2.4, Table "main objectives" row. What does stable and sustainable catches mean here? Also check the maturity schedule and natural mortality rows: they are indicated as "no" but I thought I read that both are fixed parameters? And for "interest parameters" I find the "(Blim or Bpa)" reference confusing. When is the stock rebuilt: when it is above Bpa in 3 consecutive years with at least 50% probability? Blim would seem not to be relevant for a rebuilt definition, rather a Blim breach would indicate that rebuilding measures should have been initiated some time ago. Again, ICES may not have a formal definition, so it is worthwhile that the authors make it clear what they mean.

Section 2.7, could the authors confirm:

a) I suspect that the stock is not considered rebuilt just by exceeding Blim (although that is a necessary condition). It is likely that the stock also must exceed Bpa to be considered rebuilt.

b) I think what is being calculated within each replicate is whether a given year is (at least) the third year where where SSB>Blim or SSB>Bpa. Such years are coded with a “3”. Then the cumulative total of the "3s" is calculated by year within each replicate. Finally, the cumulative sums in each year are divided by 1000 to obtained the probabilities.

c) The year "rebuilding" to Blim or Bpa is declared is the year where the cumulative probability exceeds 50%.

Figure 2.16. Not sure how much of the trending in (mean?) weight at age might be sampling versus real signal. However, if real signal then the projection does not seem to capture the trending behavior (hard to tell based on 3 iterations, should probably also add the mean simulated trend). Something to look more closely at in *future work*.

Section 2.13. A useful reference for including in this discussion might be: Kerr, L. A., Hintzen, N. T., Cadrin, S. X., Clausen, L., Worsøe Dickey-Collas, M., Goethel, D. R., Hatfield, E. M.C., Kritzer, J. P., and Nash, R.D.M. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. – ICES Journal of Marine Science, 74: 1708–1722.

p. 57. Were the 1,000 converged simulations randomly selected from the set of converged simulation within the 1,700 attempted simulations? If it took 1,700 attempts to get 1,000 successful simulations, then this issue needs to be investigated in future work as that is a high failure rate.

Table 6.1. “*The default probability for rebuilding above the target is 95% but for certain stocks a lower probability may be more relevant in the short- to medium-term…*” In future, one should probably evaluate whether this is feasible. I would think a lot of whether this is possible is influenced by how Bpa is selected – establishing a 95% certainty for SSB above the rebuilding target should not be done in absence of evaluating whether, say a biomass target like Bmsy can be obtained with at least 50% probability. In other words, achieving a SSB greater than Bpa with 95% probability may result in having to exceed Bmsy with (for example) 70% probability (much larger than one might want for a target reference point unless ecosystem considerations suggest SSB should be well above Bmsy). The system performance should be evaluated from the limit reference point to the rebuilding target, and then to the target reference point to determine what is feasible. Again, this is something to be considered for future work.