ICES CIEM

WORKSHOP ON MANAGEMENT STRATEGY EVALUATION FOR NORTH SEA HERRING (WKMSEHERRING)

VOLUME 7 | ISSUE 53

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 © 2025 International Council for the Exploration of the Sea

ICES Scientific Reports

Volume 7 | Issue 53

WORKSHOP ON MANAGEMENT STRATEGY EVALUATION FOR NORTH SEA HERRING (WKMSEHERRING)

Recommended format for purpose of citation:

ICES. 2025. Workshop on Management Strategy Evaluation for North Sea Herring (WKMSEHerring). ICES Scientific Reports. 7:53. 179 pp. http://doi.org/10.17895/ices.pub.28868798

Editor

José A. A. De Oliveira

Authors

Benoit Berges, Niels Hintzen, John Tyler Trochta, Valerio Bartolino, Bernhard Kuehn, Marta Ballesteros, Jürgen Batsleer, Florian Berg, Helen Bradbrook, Thomas Carruthers, Imogen Cessford, José De Oliveira, Carryn de Moor, Daniel Howell, Cecilie Kvamme, Lisa Laustsen, Paul Marchal, Tanja Miethe, Henrik Mosegaard, Richard Nash, Snorri Palmason, Campbell Pert, Nis Sand Jacobsen, Claus Reedtz Sparrevohn, Michael Spence, Norbert Rohlf, Hannah Sterling, Esben Sverdrup, Tamara Talevska, Vanessa Trijoulet, Ashley Wilson

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

Contents

1 Executive Summary

The European Union, Norway, and the United Kingdom jointly requested ICES to identify appropriate combinations of Ftarget and Btrigger in a harvest control rule that, together with possible TAC constraints, could form part of a long-term management strategy for North Sea autumn spawning herring (*Clupea harengus*) in Subarea 4 and divisions 3.a and 7.d, (North Sea, Skagerrak and Kattegat, eastern English Channel). A Management Strategy Evaluation (MSE) process was set up to evaluate such combinations, check that the management strategy was robust to a 10% banking and borrowing mechanism, and to investigate sensitivity to a number of exploitation pattern scenarios. Key sources of uncertainty (related to recruitment and natural mortality) were captured by a reference set of operating models, and results were integrated over these.

Key findings are that several management strategies are possible for a similar level of precaution (less than 5% risk), with similar levels of catch (less than 5% difference in average yield between the possible rules). These different rules involve either a high Ftarget and high Btrigger, or a lower Ftarget combined with a lower Btrigger, and provide a trade-off between maximising catch and minimising interannual variability in catch. However, a high Ftarget-high Btrigger combination (e.g. Ftarget=0.34, Btrigger=1.7 million tonnes) results in a marginally higher yield (average annual catch of 0.37 million tonnes) but with a realised fishing mortality well below Ftarget (0.23 against Ftarget=0.34) because such high theoretical Ftarget rules result in the stock being frequently below the Btrigger value. This combination of control points are also associated with more unstable catches (IAV=18.5%), lower SSB (1.3 million tonnes), and the more frequent suspension of TAC constraints (as a result of SSB being below Btrigger). In contrast, a low Ftarget-low Btrigger combination (e.g. Ftarget=0.21, Btrigger=0.8 million tonnes) results in marginally lower yield (average annual catch of 0.36 million tonnes) but with realised fishing mortality close to Ftarget (0.2 against Ftarget=0.21), substantially more stable catches (IAV=9.9%), higher SSB (1.5 million tonnes), and less frequent suspension of any stability mechanisms in place.

Assuming current fishing conditions (2022-3), as opposed to longer-term recent (2013-2021) or historical (1998-2003) conditions, leads to higher risk and more variable fishing mortality on ages 0-1, providing assurance that assuming current conditions (as was adopted for the reference set of operating models) is both reasonable and errs on the side of precaution with respect to risk. Increasing fishing mortality on ages 0-1 has a clear negative impact on risk, SSB and catch in the long-term. Furthermore, shifting the selection pattern towards younger ages is generally negative across all performance metrics. Management

strategies appear to be robust to implementing a 10% banking and borrowing scheme, even under an extreme version that deliberately forces unrealistic annual fluctuations in catch (the opposite of its intended purpose). It should be stressed that Ftarget and Btrigger selected from the MSE should not be confused with reference points set by ICES and should not be used to indicate stock status.

2 Introduction

A WKMSEHerring Scoping meeting was held during 17-18 January 2024 in Copenhagen to plan the work needed to answer the joint request to ICES to advise on a Long-Term Management Strategy for North Sea autumn-spawning (NSAS) herring in North Sea, Skagerrak and Kattegat and Eastern English Channel, given in Annex 1 of this report (which includes clarifications on the request). The scoping meeting was followed by a series of online meetings to finalise the set of operating models, sensitivity tests and robustness tests (21 February, 15 April, 22 July, 28 October 2024); a summary of these meetings can be found in the Scoping Report associated with this meeting (Annex 9). A participants list can be found in Annex 2. Descriptions and main results of the MSE are given in the main body of the report, with Annexes 3-6 providing further details. A stakeholder engagement session was held as part of this process and is reported in Annex 7. The external reviewers' report is given in Annex 8.

Formal TORs for the meeting were developed once the request was signed off and agreed between ICES and the advice requesters. These are as follows:

2024/WK/FRSG45 – The **Workshop on Management Strategy Evaluation for North Sea Herring (WKMSEHerring)** chaired by José De Oliveira, UK will be established and will meet at ICES HQ, Copenhagen 10-12 December 2024, and online 23 January 2025 and 13 February 2025 to:

a) Develop a work plan as outlined below, following on from the January 2024 meeting with advice requesters to agree on the approach and tools for the MSE, in response to the joint request from the advice requesters on a long-term management strategy for North Sea autumn spawning herring.

b) The December 2024 meeting to include results for the conditioning of the suite of pre-agreed operating models, results for the management strategy options and sensitivity tests stipulated in the request, and any additional results of interest forthcoming from the MSE analysis. This meeting will include stakeholder engagement.

WKMSEHerring will report by 01 April 2025 for the attention of ACOM.

Description	Date	Meeting format and composition	Additional information
Start: 1st meeting	Jan-24	In-person + managers	17-18 Jan, ICES HQ, Copenhagen
	Mar-24		HAWG
Explorations meeting	Dec-24	In-person	10-12 Dec, ICES HQ, Copenhagen
Final explorations meeting	Jan-25	On-line + managers	23 Jan
Tuning final model	Feb-25	On-line + managers	13 Feb
	Mar-25		HAWG
	Apr-25		Report to ICES

Priority	High priority. This workshop will facilitate a special request to ICES from EU-NO-UK.

Scientific justification	
Resource requirements	
Participants	The WK will be attended by experts contributing to the joint request, HAWG experts including the North Sea herring stock assessor, and two external reviewers.
Secretariat facilities	ICES HQ for the in-person meeting and online meeting setup.
Financial	Budget set in response to the joint request
Linkages to advisory committees	Direct link to ACOM.
Linkages to other committees or groups	HAWG
Linkages to other organizations	

There is a github repository that contains the software code used (https://github.com/ices-taf/wk_WKMSEHerring.git); it is not public, but access can be provided on request.

3 Background

3.1 NSAS herring and assessment model

The assessment for North Sea Autumn-Spawning (NSAS) herring is using commercial and survey data and spans the 1947-2023 period. The assessment is conducted annually at the Herring Assessment Working Group (HAWG; ICES (2024a)). The model used is the SAM stock assessment model (Nielsen and Berg (2014)) in a single fleet configuration. In parallel to the single fleet assessment, a SAM multi-fleet assessment model (Nielsen et al. (2021)) is also conducted yearly to inform the short-term forecast on fleet-wise fishing selectivity. The NSAS stock assessment was benchmarked in 2018 (ICES (2018)) and 2021 (ICES (2021)) and underwent a management strategy evaluation (MSE) in 2019 (ICES (2019)). Despite the latter, there is no agreed management strategy to date for this stock and under the ICES framework, the ICES MSY approach advice rule[1] has taken precedence for the advice since 2018.

The NSAS stock is harvested by 4 fleets (Figure 3.1):

- A fleet: human consumption in the North Sea and Eastern Channel
- B fleet: bycatch of herring (in the sprat and Norway Pout fishery) in the North Sea
- C fleet: human consumption in 3.a (Skagerrak-Kattegat area)
- D fleet: bycatch of herring (in the sprat and Norway Pout fishery) in 3.a

[1] https://www.ices.dk/advice/Pages/technical_guidelines.aspx

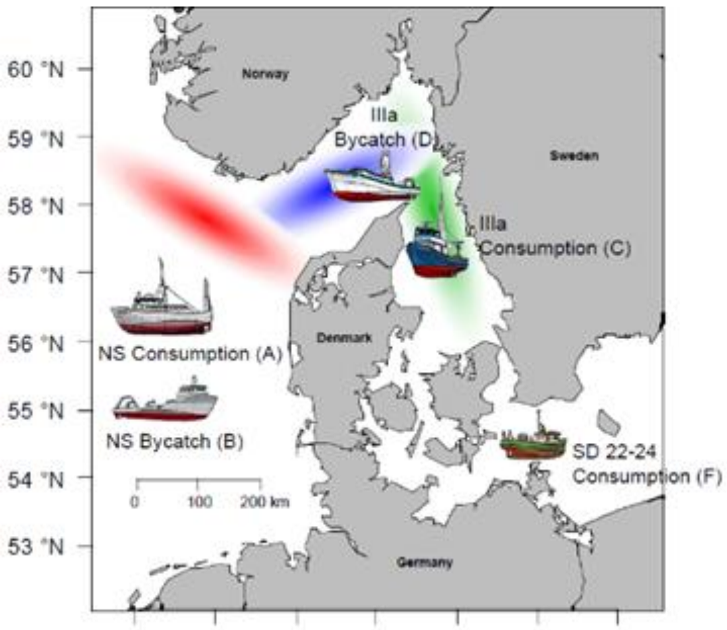


Figure 3.1: Conceptual drawing of the spatial coverage of the A-D and F fleets. The A and B fleets operate in the North Sea (red shaded area). The C and D fleets operate in the Western Baltic area, green and blue shaded areas respectively.

The corresponding data for catches at age are available from 1947 but are only disaggregated by fleet from 1997. Most of the catches are realised by the A-fleet (Figure 3.2). However, other fleets are of importance because of the different fleets selectivity and mixing with the Western Baltic Spring Spawning (WBSS) herring stock. In term of selectivity, the A fleet targets ages 2+ winter rings (wr), the C fleet targets ages 1-3 wr and the fleets B and D bycatch juveniles (ages 0-1 wr) (Figure 3.3).

In term of surveys, the assessment model is informed by 5 surveys:

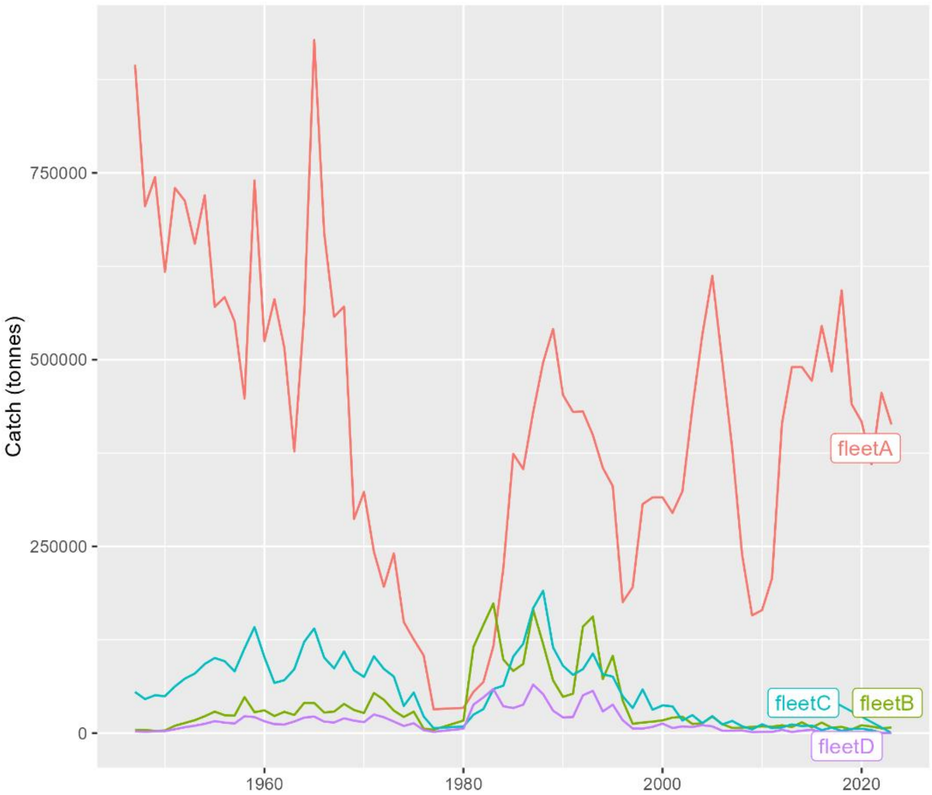
- Larvae abundance index, LAI: survey focuses on the early larvae life stage of NSAS and covers the four different stock components: Orkney/Shetland, Buchan, Central North Sea (CNS), Southern North Sea (SNS). The influence of this survey is limited but remains important as it provides information on stock components. This survey is also known as the IHLS (International Herring Larvae Survey).
- IBTS0 (age 0): late larvae survey (MIK net) taking place Q1 of each year on all stock components except the southern North Sea components (so called Downs). This is usually a good indicator of recruitment.
- IBTS-Q1 (age 1): bottom trawl survey taking place Q1 of each year which provides clear information on the recruitment survivors to the fishery.
- IBTS-Q3 (age 0-5): bottom trawl survey taking place Q3 of each year.
- HERAS (age 1-8+): acoustic survey covering the full extent of the NSAS and WBSS stocks and is conducted yearly in June/July. The derived indices cover age 1+ and are very influential to the stock assessment

model. This survey also provides the weight at age and the maturity at age for the assessment.

A summary of data sources by age is shown in Figure 3.4.

In addition to these input data, natural mortality for herring is estimated by the North Sea SMS multi-species model (ICES (2024b)).

(a)



(b)

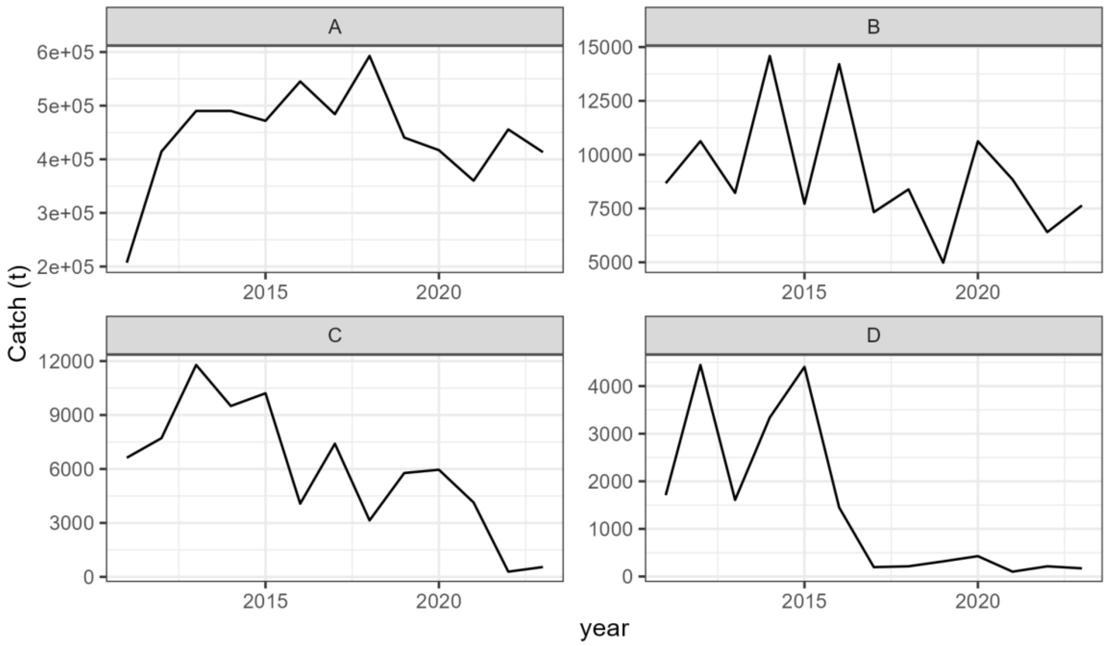


Figure 3.2: Catch per fleet (A-D). (a): Catches over the 1947-2023 period. (b): Catches over the period 2010-2023 (note the different scaling in y axis).

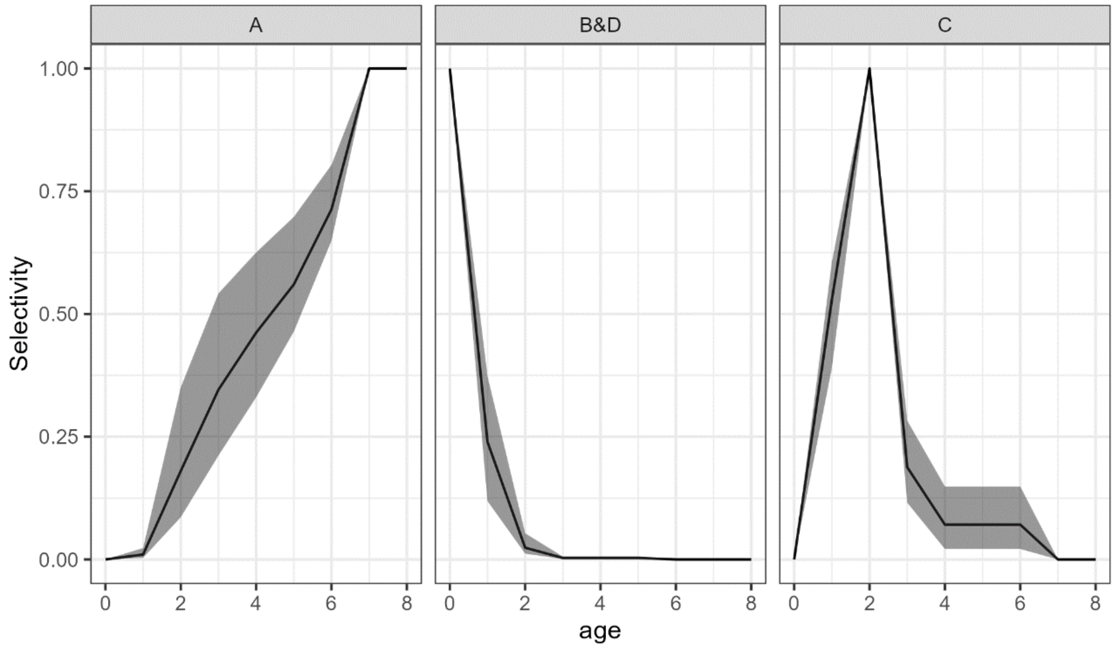


Figure 3.3: fishing selectivity at age by fleet. Fleets B and D are bycatch fleets and exemplify the same fishing selectivity.

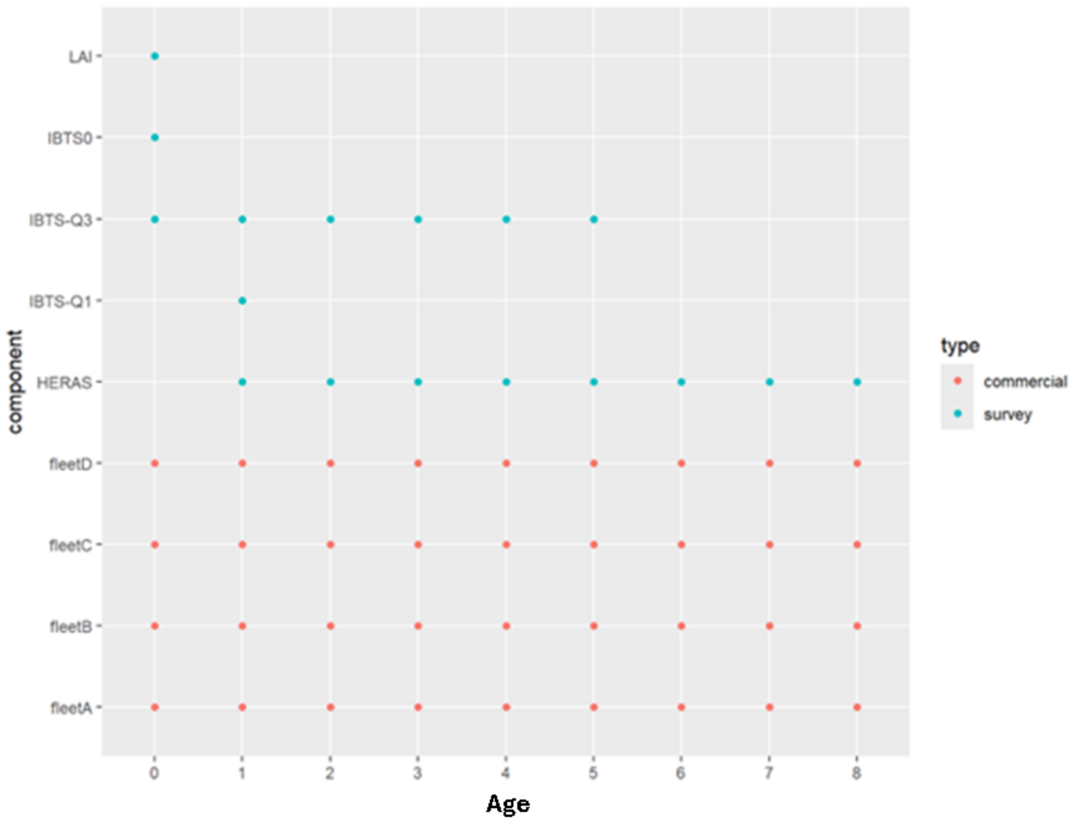


Figure 3.4: Summary of fishery dependent (red markers) and fishery independent (blue markers) input data to the assessment.

3.2 Intermixing with WBSS

Intermixing between the NSAS and WBSS herring stocks is taking place in the 3.a area and along the Norwegian coast (Figure 3.5). Further background information can be found in Bekkevold et al. (2023). Most of the mixing is taking place in the 3.a area where the C and D fleets operate, but also in the eastern part of 4a and 4b in the so called “transfer area” (Figure 3.6). The MSE undertaken in WKMSEHerring only concerns the NSAS stock, but the intermixing has implications for the management of the WBSS stock, as it is currently below its biomass reference point Blim (Figure 3.7; within the ICES framework, Blim is estimated through the fitting of segmented regression stock recruitment functions, following dedicated technical guidelines; ICES (2023)). As a result, the stock is subject to zero-catch advice from ICES (2019-2024). In practice, the TACs for the C and D fleets follow management rules with 1) the C fleet TAC being scaled based on the A fleet TAC and 2) the D fleet TAC being kept constant at 6659 t. The consequence is that TACs for the C and D fleets are not set as zero and to avoid putting fishing pressure on the WBSS stock, they are transferred almost entirely since 2022 from the 3.a area to the North Sea (Figure 3.6). When transfer of the C fleet TAC takes place, the TAC is taken in the North Sea and the fleet selection pattern follows the one from the A fleet. The way this transfer has been implemented following negotiations has led to a yearly overshoot of the overall TAC for NSAS herring.

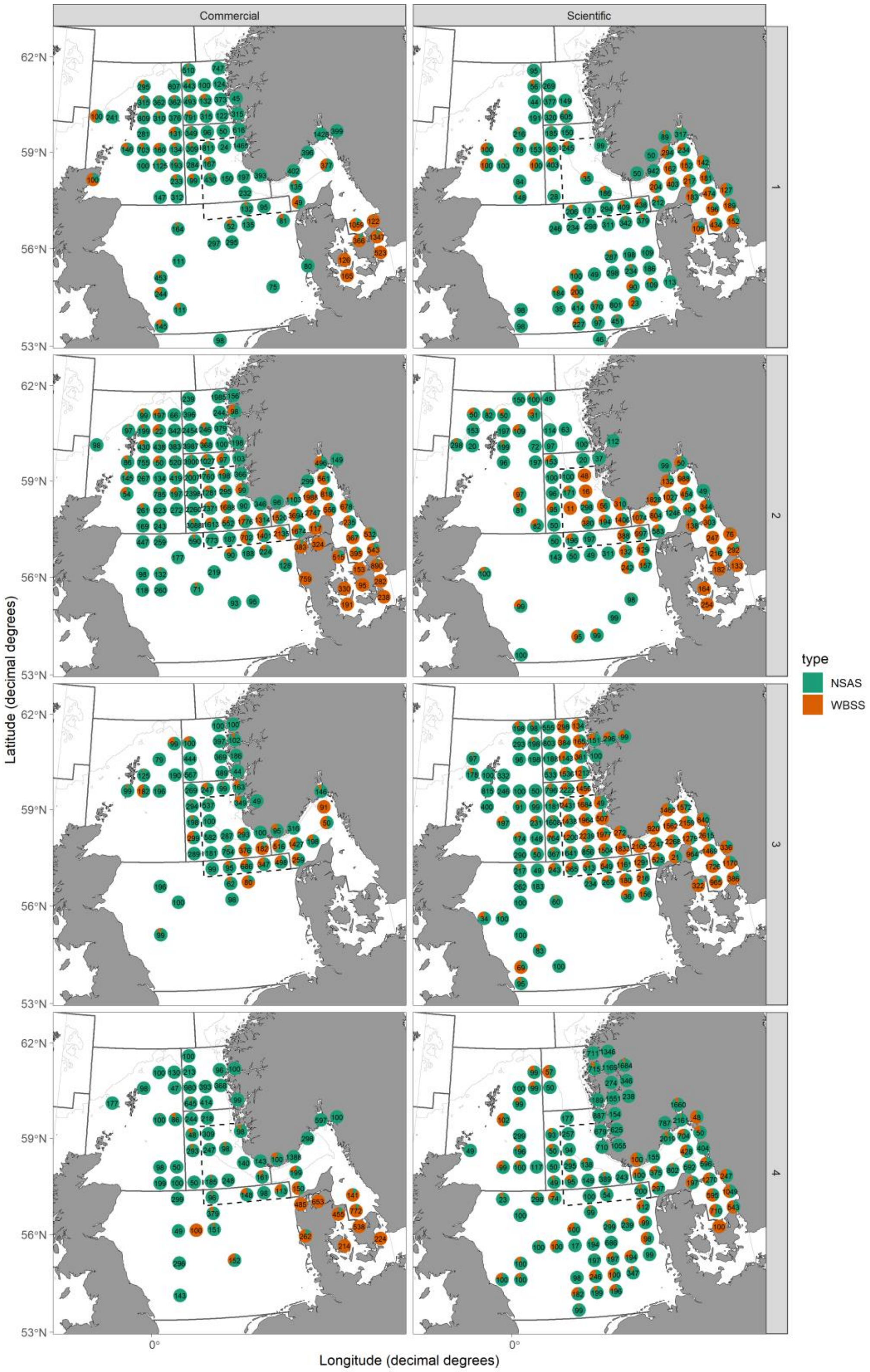


Figure 3.5: Proportions of NSAS herring (green) and WBSS herring (orange) based on vertebral counts (Norwegian data) and otolith microstructure (Danish and Swedish data) from 1980-2020. Mixing levels between NSAS and WBSS from commercial and scientific data (columns) during Q1-4 (rows).

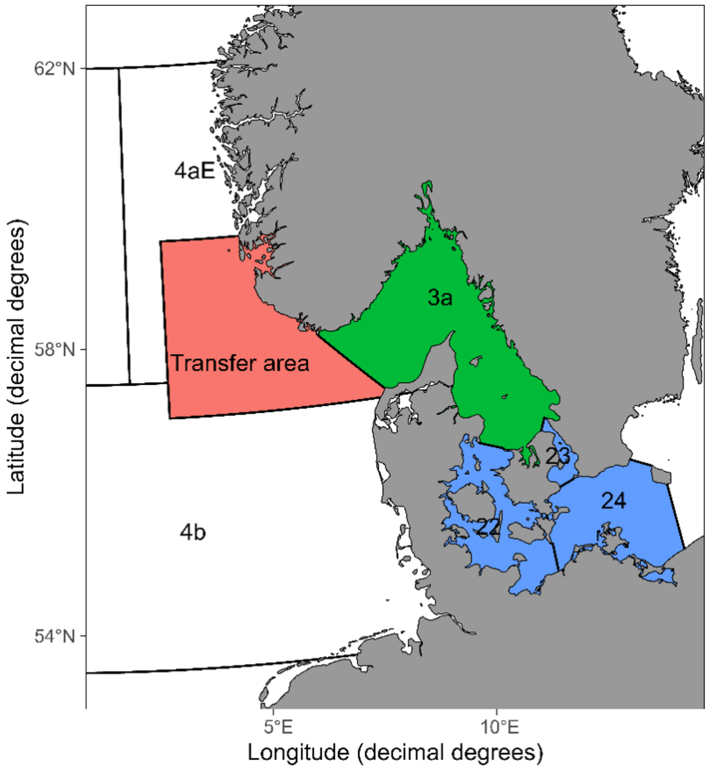


Figure 3.6: Depiction of the transfer area.

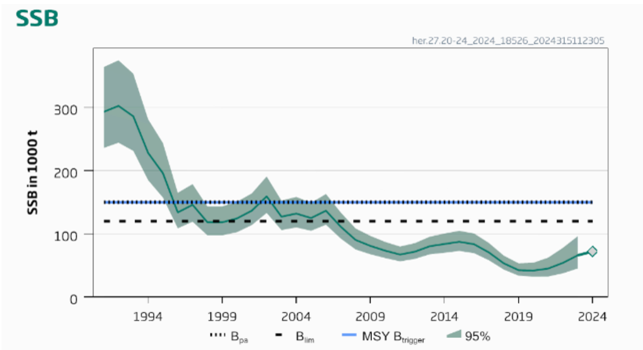


Figure 3.7: SSB trajectory of the WBSS stock.

3.3 Stock assessment and advice

In order to provide advice, the NSAS stock assessment model is run yearly in mid-late March at HAWG. The data lag for all data sources is 1 year except for the IBTS0 (age 0) and IBTS-Q1 (age 1) indices. It means that in 2024, the IBTS0 and IBTS-Q1 data run up to 2024 and all other data sources run up to 2023 (Figure 3.4). The management lag for the stock is of 1 year. It means that the catch advice given in 2024 is for the 2025 calendar year. Here are the corresponding definitions:

- Data year: the final year of the data (excluding IBTS0 and IBTS-Q1).
- Intermediate year: data year+1, also corresponds to the year the assessment is conducted.
- Forecast year: year where catch advice is given (data year+2).
- Continuation year: forecast year+1 (or data year+3)

The reference assessment model is single fleet, aggregating catch at age from fleets A-D. The multi-fleet version of the model is also run to estimate fleet selectivity for the forecast. The stock trajectory is shown in Figure 3.8. The assessment model used during the HAWG working group is stable overall with a limited retrospective pattern (ICES (2024a)). The assessment for the current MSE undergoes slight adjustments (Section 3.4) and exemplifies similar properties. The estimated observation variance per survey is shown in Figure 3.9, demonstrating the importance of the different data sources for the model. Overall, the most informative survey is the HERAS survey across the core ages.

Importantly, the natural mortality used in the assessment is the output of the SMS multi-species model from WGSAM (ICES (2024b)). In the SMS model, two natural mortalities are considered: M1 (background mortality) and M2 (predation mortality). The total mortality is the addition of these two components for each quarter of the year M=M1+M2. Whilst the SMS model effectively estimates the predation mortality M2, the background mortality M1 is taken as a fixed value. The background mortality or residual mortality is the natural mortality that is not accounted for in M2, either by predators not included in the model or by other natural mortality causes. The total natural mortality applied in the assessment is a rescaling of the M from SMS; such rescaling is obtained by profiling the stock assessment model to estimate a scalar additive to M. This assessment model profiling method was benchmarked in 2021 at IBPNSHerring 2021 (ICES (2021)). The method was last applied at the 2024 HAWG meeting (ICES (2024a)) where a new vector of natural mortality was made available. The profiling is shown in Figure 3.10 and resulted in an additive rescaling of addM=0.02. The reference points for the stock were also updated using the EqSim package (ICES (2024c)) (Table 3.1).

In the absence of an agreed management plan, the management advice is based on the ICES MSY approach advice rule, which is a biomass/fishing pressure hockey-stick rule based on the reference points shown in Table 3.1. This Harvest Control Rule (HCR) is shown in Figure 3.11. More specifically, the advice process is as follows (Figure 3.12):

1. The assessment is conducted in the intermediate year, with input data up to the data year.
2. The stock is projected in the intermediate year using the already-known TAC for that year.

3. The stock is projected in the forecast year with a fishing pressure constraint. This target fishing pressure is calculated from SSB at spawning time (in autumn) in the forecast year using the hockey-stick HCR shown in Figure 3.11.
4. TAC is derived from catches corresponding to fishing pressure in the forecast year.

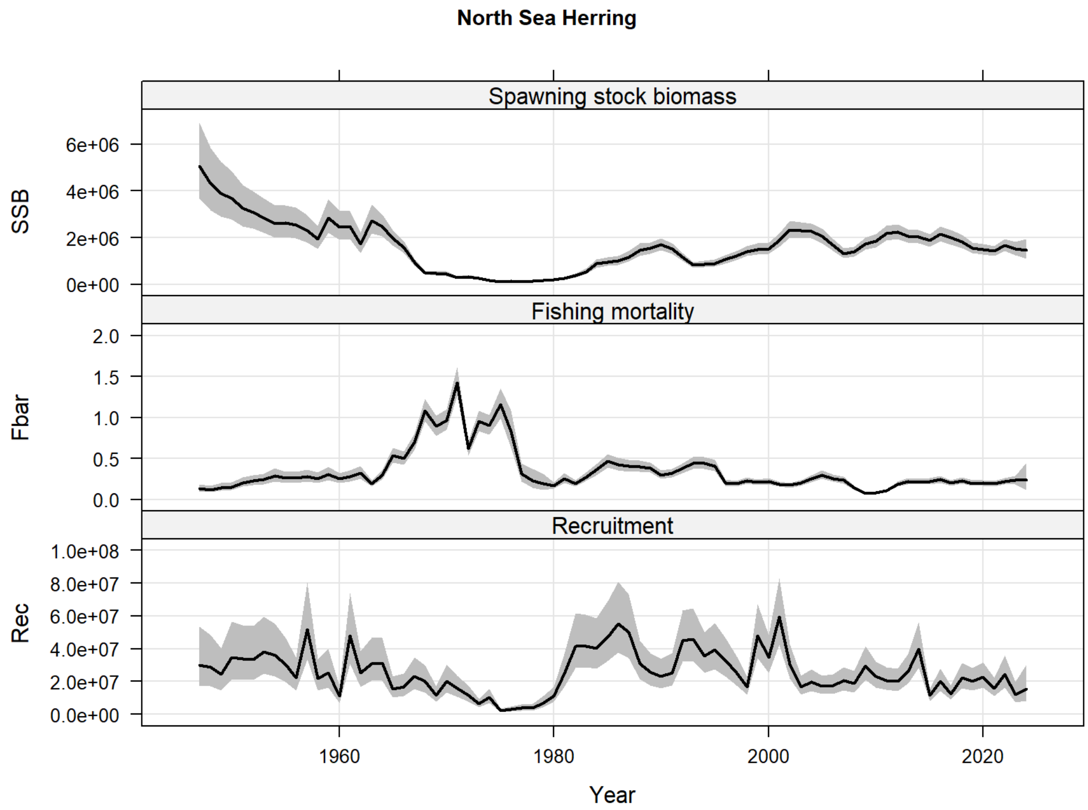


Figure 3.8: Assessment trajectory (SSB, Fbar (ages 2-6) and recruitment).

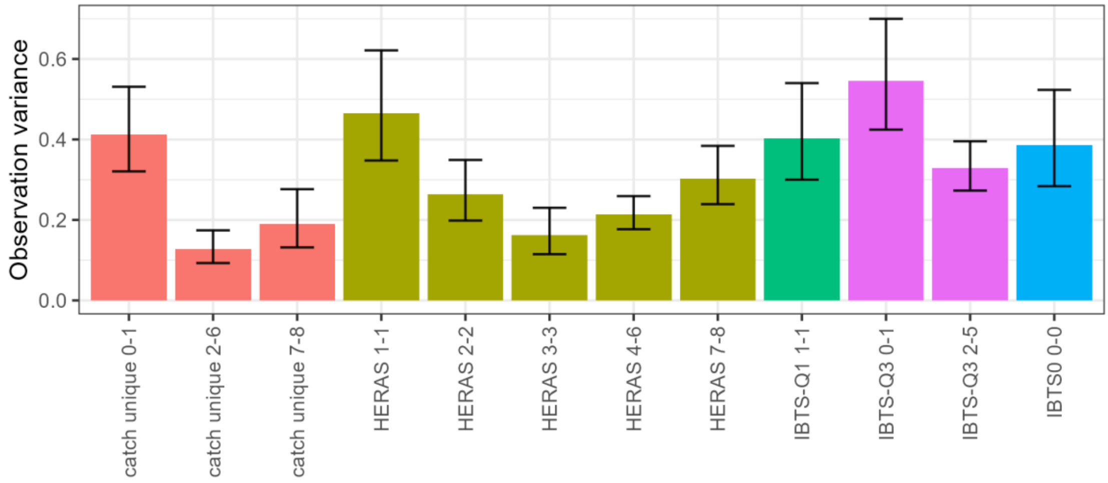


Figure 3.9: Observation variance per data source as estimated by the assessment model.

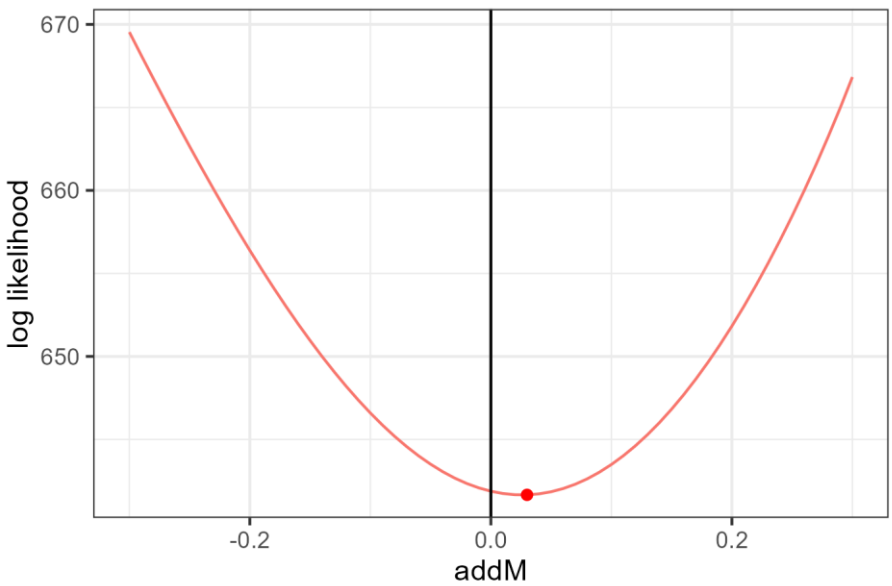


Figure 3.10: Profiling of the stock assessment model over a range of additive rescaling (addM) of natural mortality M.

Table 3.1: NSAS herring reference points (HAWG 2024; ICES (2024a)). Biomass reference points in tonnes.

Framework	Ref.point	Value
MSY approach	MSY Btrigger	1 130 747
MSY approach	FMSY	0.32
PA	Blim	828 874
PA	Bpa	1 049 521
PA	Flim	0.39
PA	Fpa	0.33

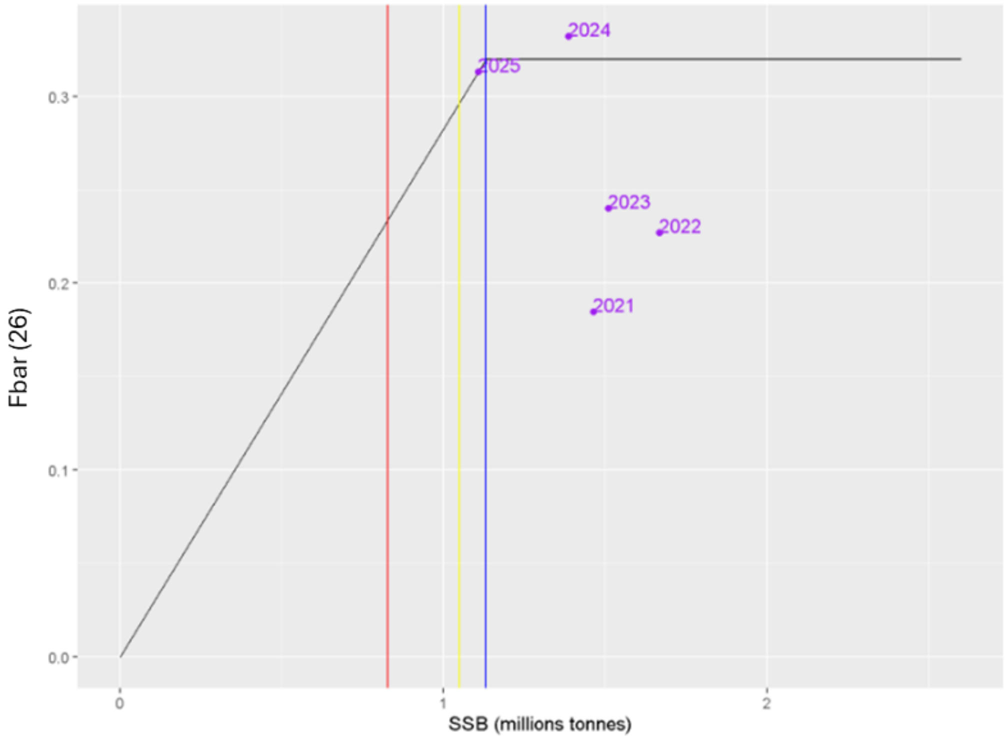


Figure 3.11: The current basis for advice for the NSAS stock is the MSY advice rule, as of HAWG 2024 (ICES (2024a)). The blue vertical line is MSY Btrigger, defining the breakpoint of the hockey-stick, with the plateau corresponding to Fmsy. The yellow vertical line is Bpa. The red vertical line is Blim. Data points show the realised F in the years 2021 – 2023. The fishing pressure in 2024 is calculated in the intermediate year whilst the 2025 fishing pressure is based on forecasting and applying the hockey-stick rule.

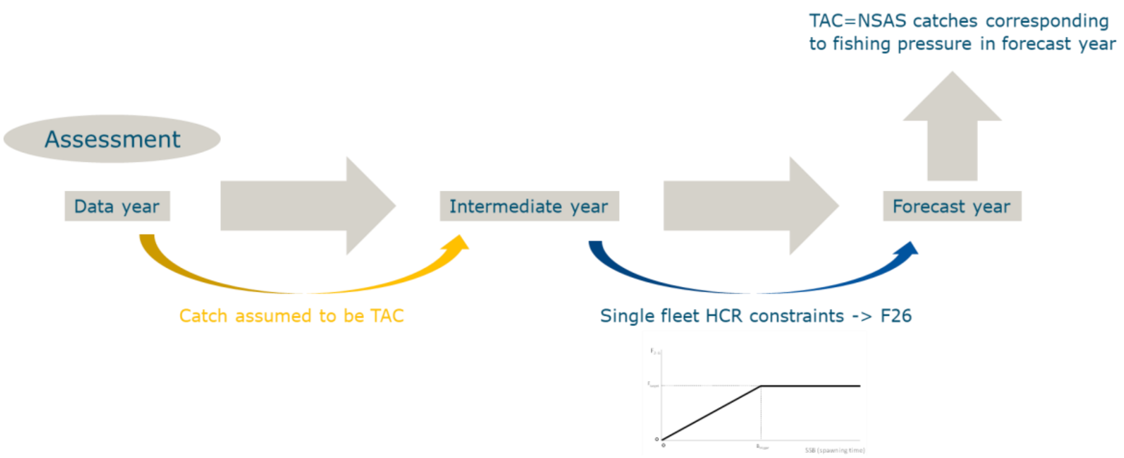


Figure 3.12: Schematic of the TAC decision process.

3.4 Assessment model for the MSE

For the Management Strategy Evaluation (MSE) conducted here, the assessment model configuration of the latest assessment group is used (ICES (2024a)). The M additive scaling value is also taken from HAWG 2024 (ICES (2024a)) as addM=0.02.

The assessment model used for the MSE is fitted without the early larvae index (LAI). This is because the LAI is an index which undergoes specific model fitting and in turn complicates the MSE model implementation. In addition, the LAI only has a marginal impact on the model fit (Figure 3.13), despite being important in the context of the yearly assessment for NSAS, as this index tracks the dynamics of the different stock components. However, in the context of this MSE where model fit is central throughout extensive iterations, the inclusion of the LAI in the assessment model was not deemed a valuable addition. This assessment without the LAI was then used to condition the operating models (OMs) and as a stock assessment model in the management procedure (MP).

The assessment model used to condition the OMs includes the IBTS0 (age 0) and IBTS-Q1 (age 1) indices up to 2024 (corresponding to the intermediate year at HAWG 2024) whilst all other data sources are up to 2023 (corresponding to the data year at HAWG 2024). The inclusion of the IBTS0 and IBTS-Q1 data points for the intermediate year allows the stock assessment model to estimate recruitment in the intermediate year. In the base assessment, as used during the yearly HAWG working group, this estimate of recruitment is further used in the forecast alongside stock estimation in the forecast year using specific assumptions (ICES (2024a)).

The conditioning of OMs is based on the use of all data available. However, for the application of the management procedure in the stock projections, no data for the IBTS0 and IBTS-Q1 indices are used in the intermediate year, which is a deviation from the model currently used during the yearly HAWG working group. More specifically, in the MSE projections, the IBTS-Q1 and IBTS0 indices are not generated in the intermediate year, but kept up to the data year to avoid complexity in the MSE model. Instead of using recruitment estimated by the assessment model in the intermediate year, recruitment is taken as a 10-year average weighted by assessment uncertainty in recruitment, which is the same as the assumption in the forecast and continuation year. Similar to removing the LAI index, the impact of removing intermediate-year data for IBTS-Q1 and IBTS0 has negligible impact on the estimation of SSB (Figure 3.14).

The SAM single fleet model configuration is given in Annex 5. Diagnostics of the single fleet stock assessment model are further shown in Annex 6.

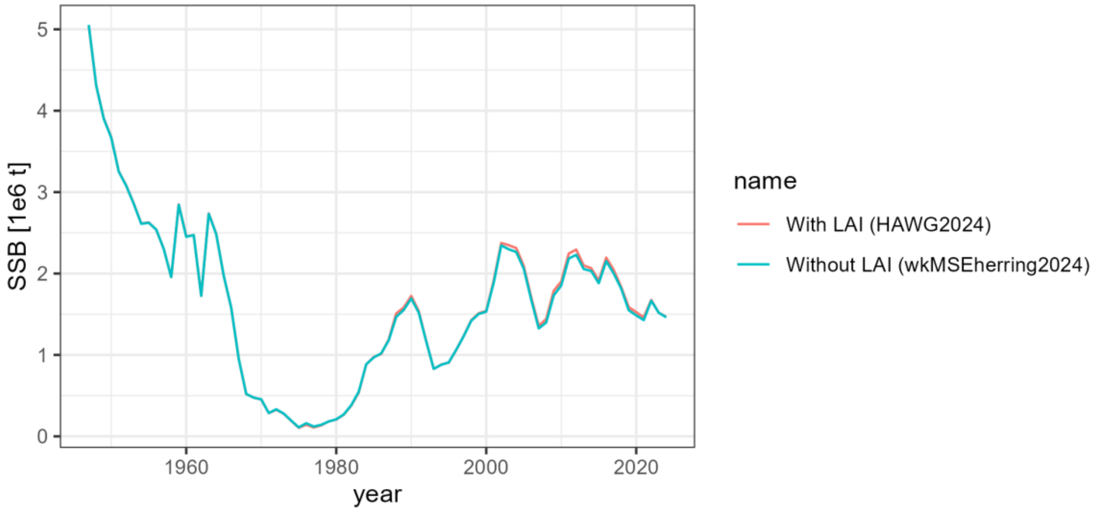


Figure 3.13: Comparison of SSB from the stock assessment models configured for HAWG and WKMSEherring. The difference between the two assessment models is the non-inclusion of the data from the LAI survey (early larvae survey).

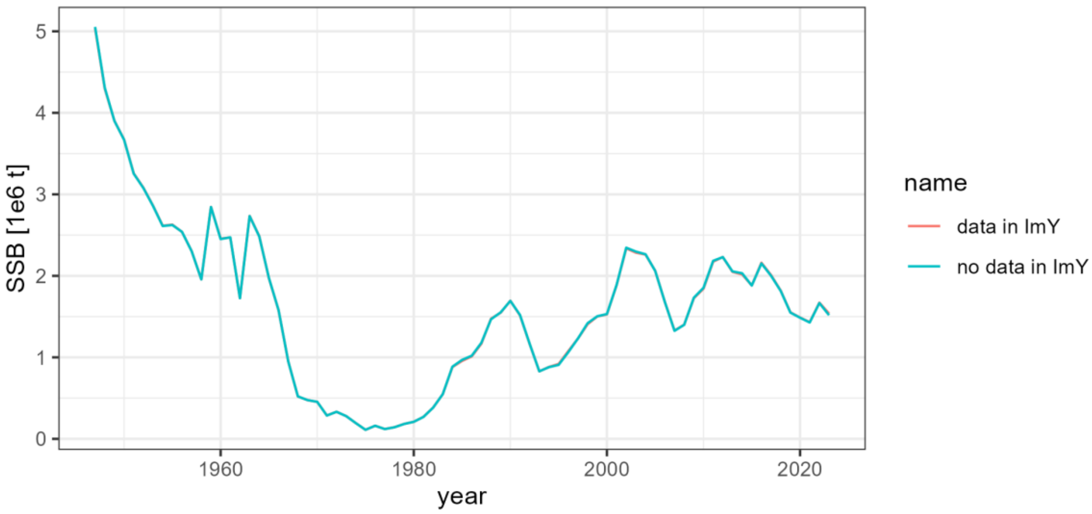


Figure 3.14: Comparison of SSB from the stock assessment models as configured for WKMSEherring with and without data in the intermediate year (ImY). The data in the intermediate year are the IBTS-Q1 and IBTS0 indices.

3.5 MSE request and framework

The MSE follows the framework from WKNSMSE (ICES (2019)) with two main blocks (Figure 3.15): the operating model (OM) and the management procedure (MP). Following the joint request (Annex 1), the management procedure (MP) is for the single fleet, i.e. aggregating catches across fleets harvesting the NSAS stock. In contrast, the operating model is multi-fleet, modelling the dynamic of each fleet separately.

The MSE is broken down into different components (Figure 3.15):

1. The Operating Model (OM) which simulates biology and the fishery system with 4 different fleets A-D
2. The Management Procedure (MP) that uses a set of observations to determine a single catch advice.
3. Within the MP, the estimation method estimates the state of the stock each year. Here, the SAM stock assessment configured as of HAWG 2024 (ICES (2024a)) with the changes described in Section 3.4, is used as the estimator.
4. Within the MP, the decision process follows the procedure described in Section 3.3. The results from the estimator are projected in the forecast year and a Harvest Control Rule (HCR) is applied. Here, the main HCR considered is a hockey-stick (F2-6 vs SSB) with Btrigger and Ftarget as control points (Figure 3.16) that require tuning. For a given year, the HCR provides a target F that is translated into catch (i.e. TAC).
5. In the implementation system, the catch advice is split into the A-D fleets in the OM.
6. The evaluation of the management strategies is done through performance metrics: average SSB, average catch, Inter-Annual Variation (IAV) in SSB and catch, fishing pressure on adults (over ages 2-6), fishing pressure on juveniles (over ages 0-1) and risk of falling below Blim (biological reference point). More specifically, the latter is taken as the maximum probability that SSB is below Blim, which corresponds to ICES Risk3 (ICES (2019)). Performance metrics are evaluated in the short (5 years, 2024-2028), medium (5 years, 2029-2033) and long term (15 years, 2034-2048).

In order to encapsulate a range of uncertainty in the MSE, seven different OMs are considered (Table 3.2). These OMs are broken down into three categories:

1. base category, containing a single OM using rationales relating to previous ICES working groups on NSAS herring.
2. stock recruitment category, containing 3 OMs encompassing a range of uncertainties around stock recruitment.
3. natural mortality category, containing 3 OMs encompassing a range of uncertainties around natural mortality.

This set of seven OMs is the reference set of the MSE, capturing a range of plausible uncertainties about the stock biology. This follows best practice when conducting MSEs, i.e. to consider more than a single OM (Punt et al. (2016)). Results for the MSE are integrated over all seven OMs in the reference set, but performance of management strategies under individual OMs is also reported.

Additional robustness and sensitivity tests are conducted. The runs for the sensitivity tests are limited to the cell in a grid of Ftarget-Btrigger combinations (with the ranges specified in the joint request) that maximises catch with an ICES risk ≤ 5% for the reference set. This was done for practical reasons and should not be considered a recommendation for selecting that cell. These sensitivity tests consider changes in exploitation patterns, juvenile fishing mortality, age shift in selection patterns for the directed fishery, omitting TAC constraints, and implementation of banking and borrowing (scheme 1 as described in WKNSMSE (ICES (2019))). All OMs are projected 25 years (to 2048) into the future with 1000 replicates. A summary of these OMs is given in Table 3.2. The specificities for the stock recruitment OMs are further expanded in Table 3.3.

The seven OMs from the reference set are used for tuning the control points of the HCR (Btrigger/Ftarget). To that aim, they are combined so there is equal weight between the three categories. In practice, the base OM is replicated three times (i.e., 1000 x 3) to achieve a comparable weight to the other two OMs categories (i.e., three OMs for stock recruitment and three OMs for natural mortality). Through combining, a final combined OM of 9000 replicates is constructed. Performance metrics are computed on that basis.

The MSE is conducted using the FLR framework, more specifically the mse package thereof[2].

[2] *https://github.com/flr/mse*

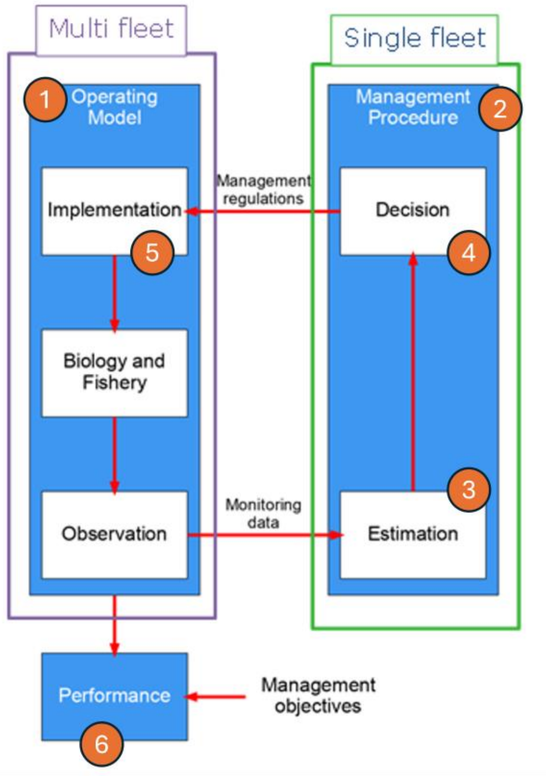


Figure 3.15: Diagram of the MSE framework. The different numbers correspond to the different components of the model, as described in detail in the text. 1: the Operating Model. 2: the Management Procedure. 3: estimation method. 4: decision process. 5: implementation system. 6: evaluation of the management strategies. Adapted from Punt et al. (2016).

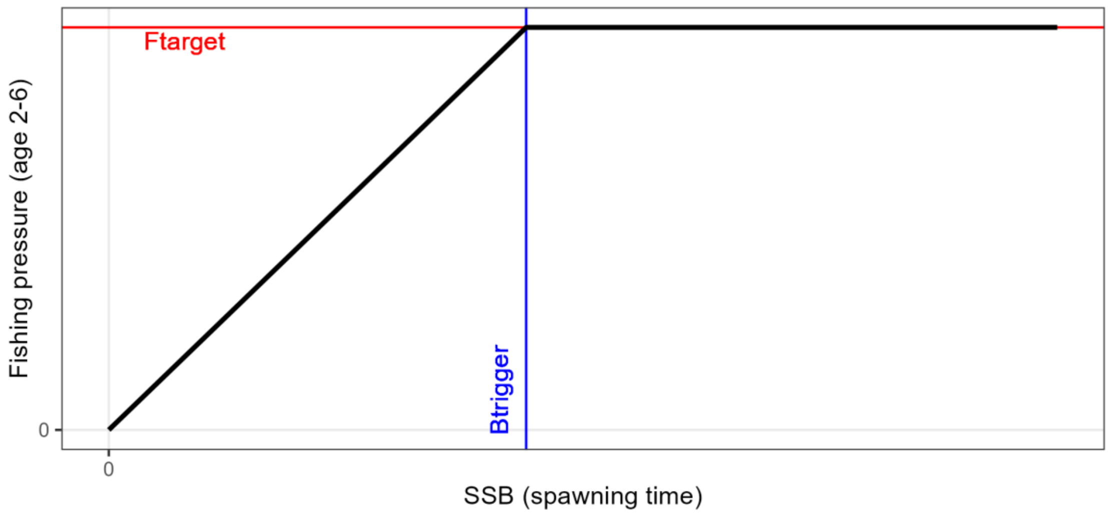


Figure 3.16: Requested Harvest Control Rule (HCR) for North Sea autumn spawning herring. The Ftarget and Btrigger are the control points.

Table 3.2: Summary of considered OMs, either as a reference set (ref), as a sensitivity test (sens) or as a robustness test (rob). Whilst reference sets are used for tuning (Btrigger/Ftarget), the runs for the sensitivity tests are limited to the cell corresponding to maximizing catch with an ICES risk < 5% for the reference set. Short time series refers to the period 2002-2024, while long time series refers to the period 1947-1978, 1991-2024 (i.e. full time series but excluding the recovery period of the stock after its collapse).

Name	Biological variables	Fishing selectivity	Fleet_behavior	Specificities	Type	Group
Base	10 years	Random walk using last 10 years	2022-2023 effort	SR fitted to short time series with steepness based on full time series without recovery period	ref	Base
SR1	10 years	Random walk using last 10 years	2022-2023 effort	SR fitted to short time series	ref	SR
SR2	10 years	Random walk using last 10 years	2022-2023 effort	SR fitted to full time series without recovery period	ref	SR
SR3	10 years	Random walk using last 10 years	2022-2023 effort	SR fitted to full time series without recovery period with depensation. This only concerns Berverton-Holt (BH)	rob	-
SR4	10 years	Random walk using last 10 years	2022-2023 effort	SR fitted to full time series without recovery period with autocorrelation	ref	SR
M1	10 years	Random walk using last 10 years	2022-2023 effort	M replicates using multi-species covariance matrix	ref	M
M2	10 years	Random walk using last 10 years	2022-2023 effort	Upward absolute scaling of M (addM = 0.07)	ref	M
M3	10 years	Random walk using last 10 years	2022-2023 effort	Downward absolute scaling of M (addM = -0.03)	ref	M
SEN1	10 years	resampling 2013-2021	2013-2021 effort	Fishing exploitation resampled from the 2013-2021 time period	sens	-

Name	Biological variables	Fishing selectivity	Fleet_behavior	Specificities	Type	Group
SEN2	10 years	resampling 1998-2003	1998-2003 effort	Fishing exploitation resampled from the 1998-2003 time period	sens	-
SEN3	10 years	resampling 2013-2021	2013-2021 effort	F01 varied over 0-0.1 in steps of 0.025	sens	-
SEN4	10 years	resampling 2013-2021	2013-2021 effort	Shifting selectivity of the A fleet. No fishing from fleets B-D.	sens	-
SEN5	10 years	Random walk using last 10 years	2022-2023 effort	TAC constraints off	sens	-
SEN6	10 years	Random walk using last 10 years	2022-2023 effort	Banking and borrowing with TAC constraints on	sens	-
M4	10 years	Random walk using last 10 years	2022-2023 effort	Downward absolute scaling of M (addM = -0.05)	rob	-
M5	10 years	Random walk using last 10 years	2022-2023 effort	Upward absolute scaling of M (addM = 0.1)	rob	-

Table 3.3: Summary of the conditioning of stock recruitment for the different OMs. The aspects that are changed across OMs are: the type of stock recruitment models, the time period considered, the inclusion of autocorrelation (AR1 process), the inclusion of depensation, any fixing of steepness and the modelling used for the deviance in recruitment. Two periods are considered for the fitting of the stock recruitment. The short time period is 2002-2024, corresponding to the recent recruitment regime for NSAS. The long time period is 1947-1978, 1991-2024, corresponding to the entire time series but excluding the period over which the NSAS stock was recovering from collapse. Recruitment deviance was modelled with either a multivariate lognormal distribution (rlnorm) or a multivariate lognormal distribution with an AR1 autocorrelation process (rlnormar1).

OM	Stock recruitment Model	Time period	AR1 process	depensation	Steepness assumption	deviance
Base	Beverton-Holt	short	no	no	fix S	rlnorm
Base	Segmented Regression	short	no	no	fix a	rlnorm
SR1	Beverton-Holt	short	no	no		rlnorm
SR1	Segmented Regression	short	no	no		rlnorm
SR2	Beverton-Holt	long	no	no		rlnorm
SR2	Segmented Regression	long	no	no		rlnorm
SR3	Beverton-Holt	long	no	yes		rlnorm
SR4	Beverton-Holt	long	yes	no		rlnormar1
SR4	Segmented Regression	long	yes	no		rlnormar1

4 OM construction

The construction of the OMs draws from what was developed during previous NSAS MSE workshops (ICES (2019)).

4.1 Stock replicates

A key part of the MSE is the inclusion of uncertainty in stock numbers and fishing pressure (or fishing mortality) estimates. Here, these uncertainties are introduced in the OM by including parameter estimation error using the variance-covariance matrix derived from the SAM model fit. From the fits of the stock assessment model, 1000 replicates of model parameters are generated, providing 1000 stock replicates. Further details can be found in Fisher (2024) and the WKNSMSE report (ICES (2019)).

Whilst stock numbers and fishing pressure at age is replicate-specific in the historical period, all observations are kept the same. The impact of the number of replicates on the calculation of ICES risk 3 is shown in Figure 4.1. Figure 4.2 shows the uncertainty bandwidth from the stock replicates, together with examples of individual stock replicates.

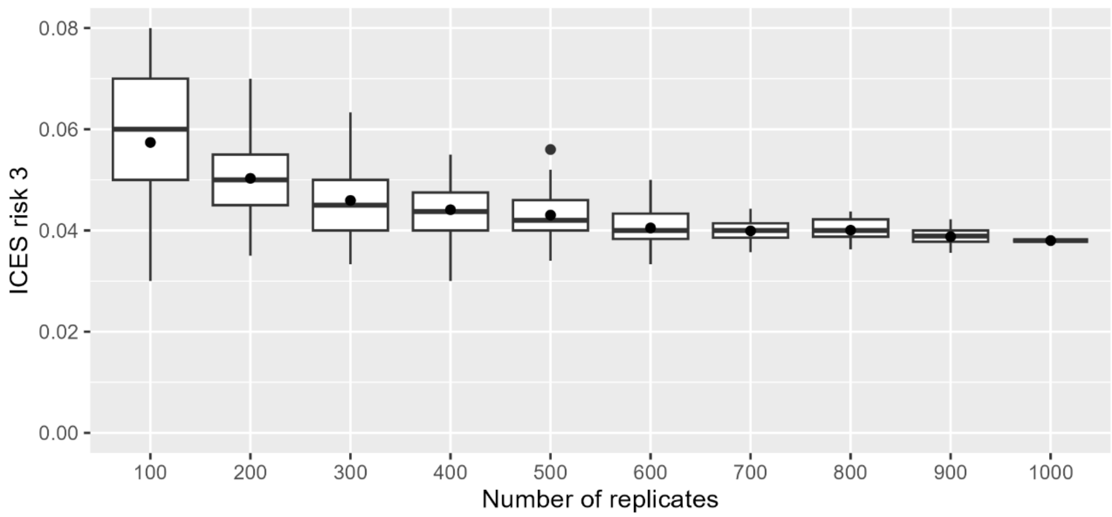


Figure 4.1. Impact of the number of replicates on ICES risk 3, using the base OM. The y-axis gives the distribution of 50 calculations of Risk3, where each calculation uses the number of replicates shown on the x-axis that were re-sampled with replacement from the original 1000 replicates.

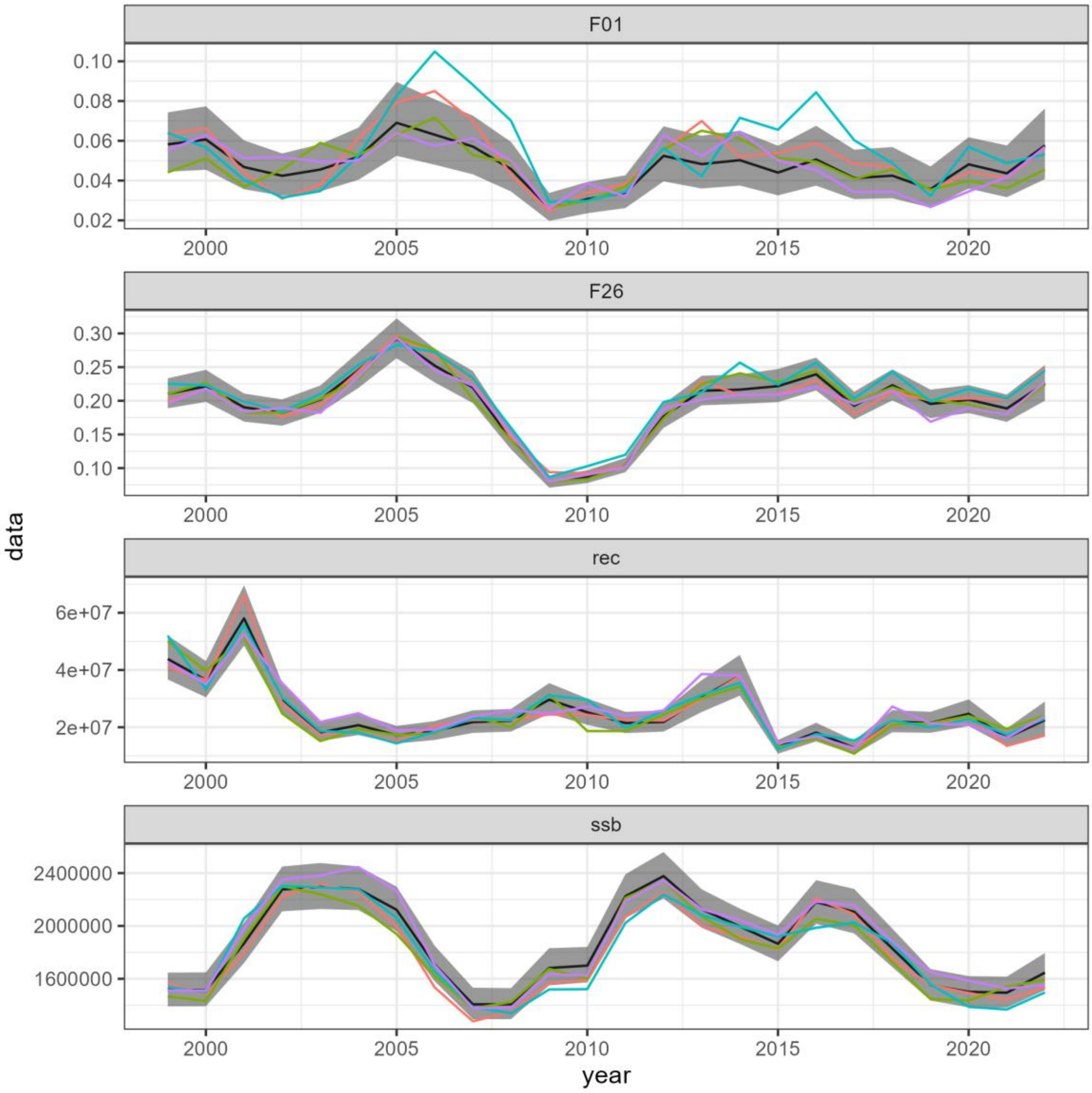


Figure 4.2. Uncertainty bound from stock replicates and examples of individual stock replicates for fishing pressure on juveniles (F01), fishing pressure on adults (F26), recruitment and SSB. Medians are shown by black lines, ribbons are the 95%/5% quantiles and the coloured lines ('worm plots') show four randomly selected individual replicate-specific trajectories. The base OM is used.

4.2 Resampling of biological parameters

In the future projections, stock weights, maturity level and natural mortality are resampled together from the last 10 years (i.e. 2014:2023), i.e. in the low productivity phase for the stock that is estimated to have started around 2002. The resampling is done in blocks of years to maintain a certain level of autocorrelation. This resampling process is performed by:

1. Randomizing a starting year in 2014:2023
2. Randomizing the number of years in a block, with a maximum 10 (in the case 2014 is drawn as the first year with a block of 10 years).

The process results in a series of blocks spanning the projection period (2024-2048). This sampling scheme results in an uneven sampling of individual year (2014-2023), as shown in Figure 4.3. The corresponding maturity level, stock weights and natural mortality are further shown in Figures 4.4-6.

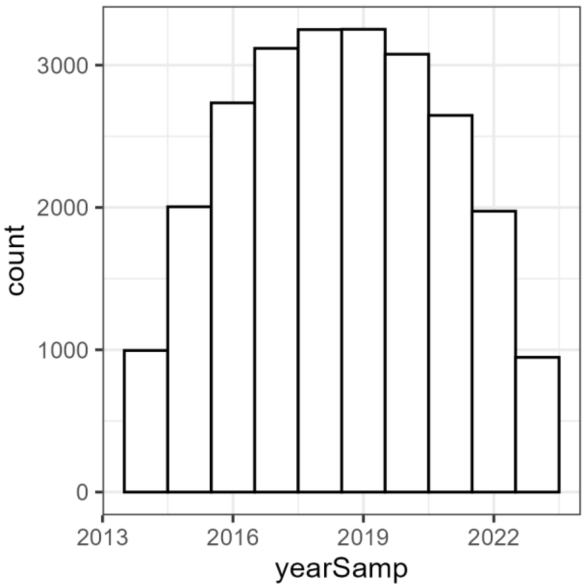


Figure 4.3: Distribution of years for the resampling of biological parameters.

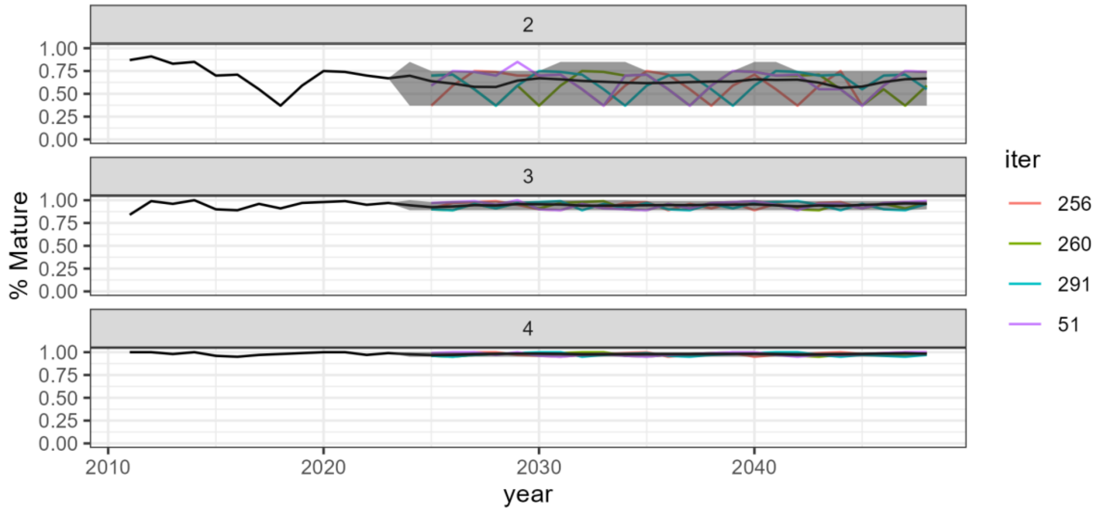


Figure 4.4: Future maturity. The black line is observed maturity for historical years and the median across replicates for the projection period. Ribbons are the 95%/5% quantiles in the projection period. Individually coloured lines are drawn from individual OM replicates.

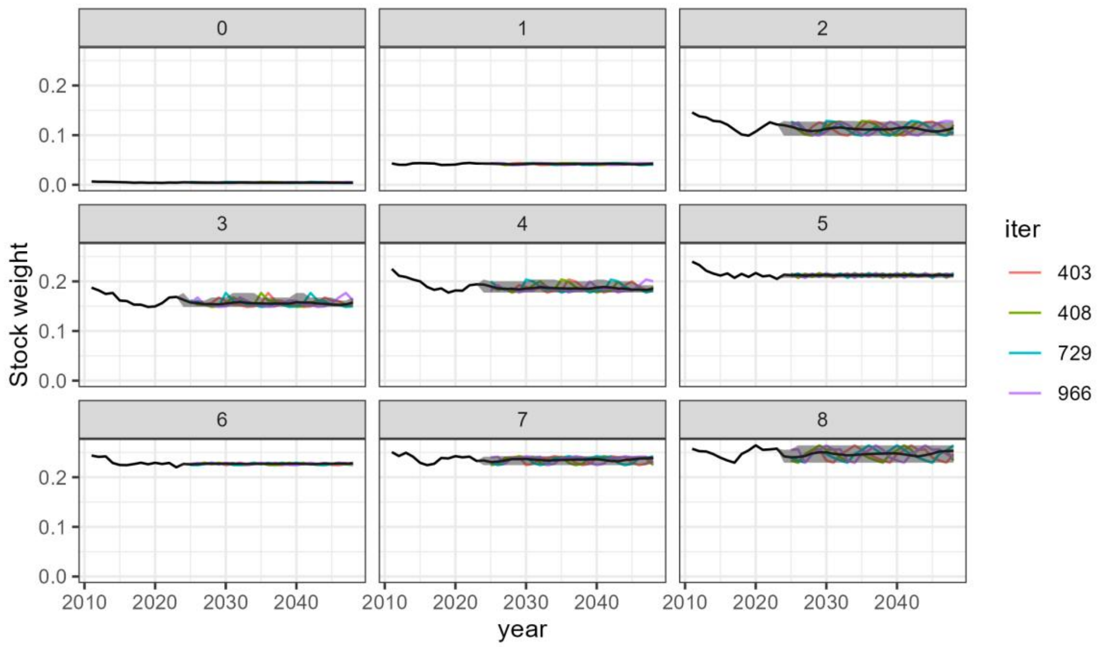


Figure 4.5: Future stock weights. The black line is observed stock weight for historical years and the median across replicates for the projection period. Ribbons are the 95%/5% quantiles in the projection period. Individually coloured lines are drawn from individual OM replicates.

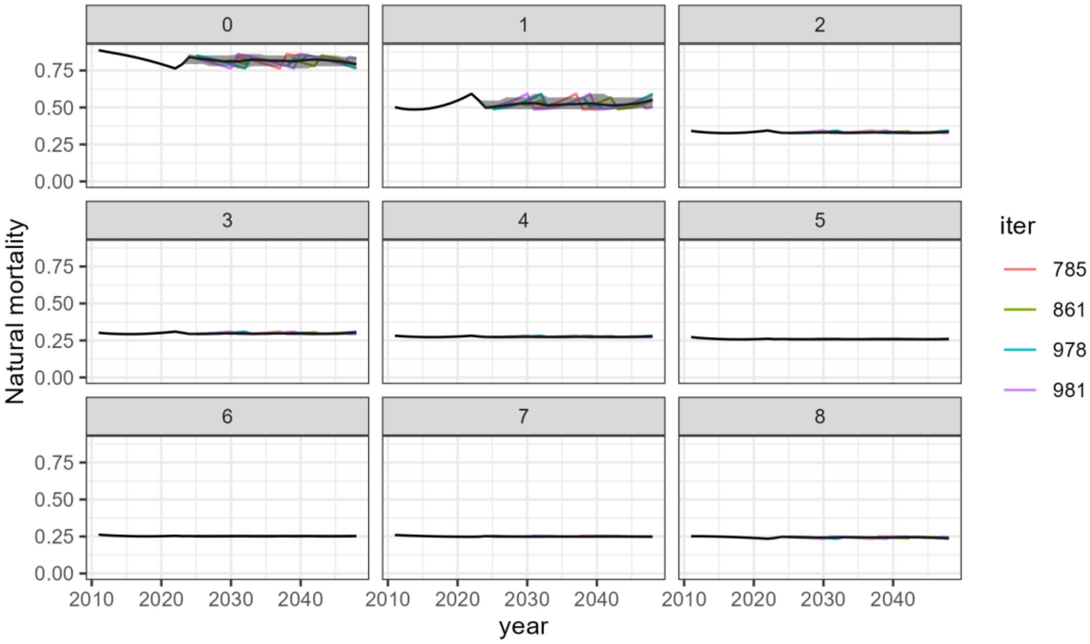


Figure 4.6: Future natural mortality. The black line is natural mortality from the SMS multi-species model for historical years and the median across replicates for the projection period. Ribbons are the 95%/5% quantiles. Individually coloured lines are drawn from individual OM replicates.

4.3 Fleet effort

The implementation system, as depicted in Figure 3.15, distributes the single catch for the forecast year derived from the HCR among the different fleets (A-D). Such a split is based on the yearly fleet fishing effort, which is defined as apical (or maximum) fishing pressure over all ages, i.e.:

$$E_i(y) = \max\big(F_i(a = 0{:}8, y)\big)$$

with $E_i(y)$ is the fishing effort of fleet i in year y. The quantity $F_i(a, y)$ is the fishing pressure at age in year y. The fishing effort by fleet is shown in Figure 4.7. Because fishing effort relates to fishing pressure, it allows one to control the dynamic of fishing fleets whilst retaining the selectivity of a given fleet. In practice, the quantities that will be used in the implementation system are the efforts of fleet B, C and D relative to fleet A. These are calculated as:

$$r_{B/A}(y) = E_B(y)/E_A(y)$$

$$r_{C/A}(y) = E_C(y)/E_A(y)$$

$$r_{D/A}(y) = E_D(y)/E_A(y)$$

These quantities are specific to each replicate. These are plotted across 1000 replicates in the historical period in Figure 4.8. For example, the fishing effort of fleet B in 2022-2023 is about 19% of the fishing effort of fleet A whilst the fishing effort of fleets C and D are marginal in that time period. For the different OMs, the relative fishing effort is resampled from different time periods:

- base and other OMs in the reference set: relative fishing effort taken as the average over 2022-2023
- SEN1: relative fishing effort resampled in a block of years (similar process as described in Section 4.2, but different blocks to those used for biological variables) over the period 2013-2021
- SEN2: relative fishing effort resampled in a block of years (similar process as described in Section 4.2, but different blocks to those used for biological variables) over the period 1998-2003

The differences between these periods for the fishing effort relative to fleet A are exemplified in Figure 4.9 (fleets B, C and D).

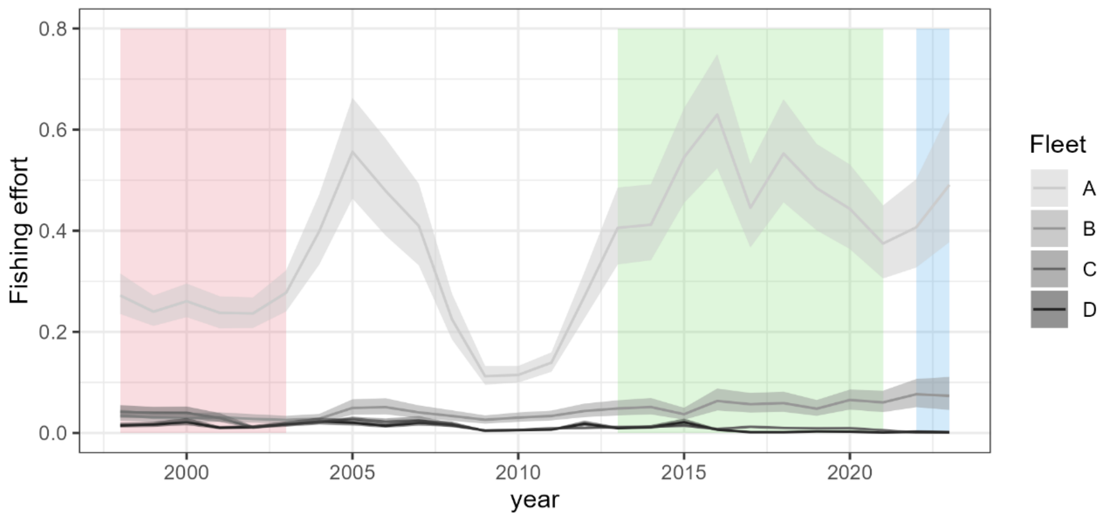


Figure 4.7: Time series of effort over the period 1998-2023 for fleets A-D. The shared areas corresponds to the periods over which the effort for fleets B-D relative to the effort of fleet A is resampled, namely 2022-2023 (blue), 2013-2021 (green) and 1998-2003 (red). Ribbons are the 95th/5th quantiles.

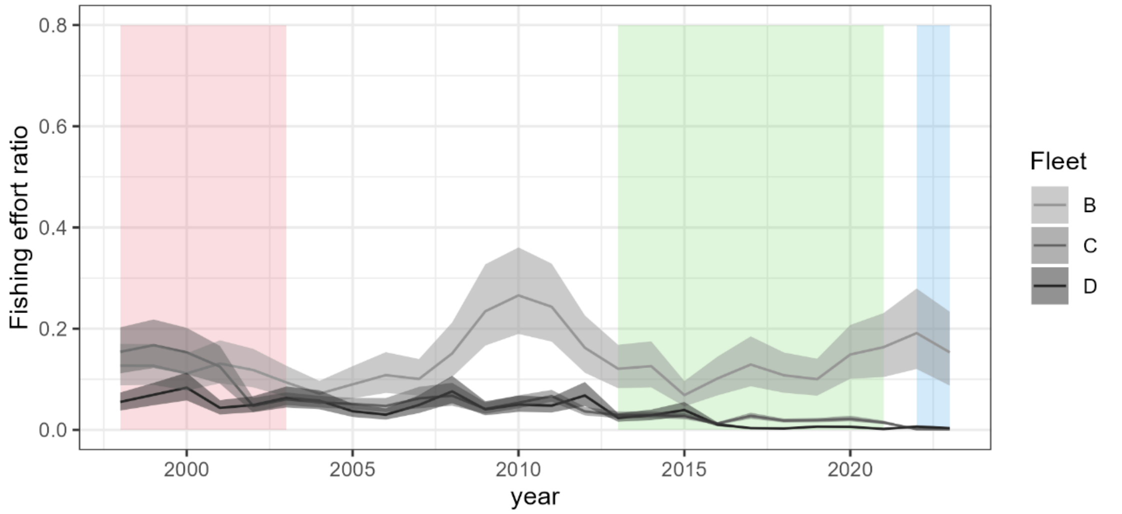


Figure 4.8: Time series of ratios of efforts of fleets B-D relative to fleet A over the period 1998-2023. The shared areas corresponds to the periods over which the effort for fleets B-D relative to the effort of fleet A is resampled, namely 2022-2023 (blue), 2013-2021 (green) and 1998-2003 (red). Ribbons are the 95%/5% quantiles.

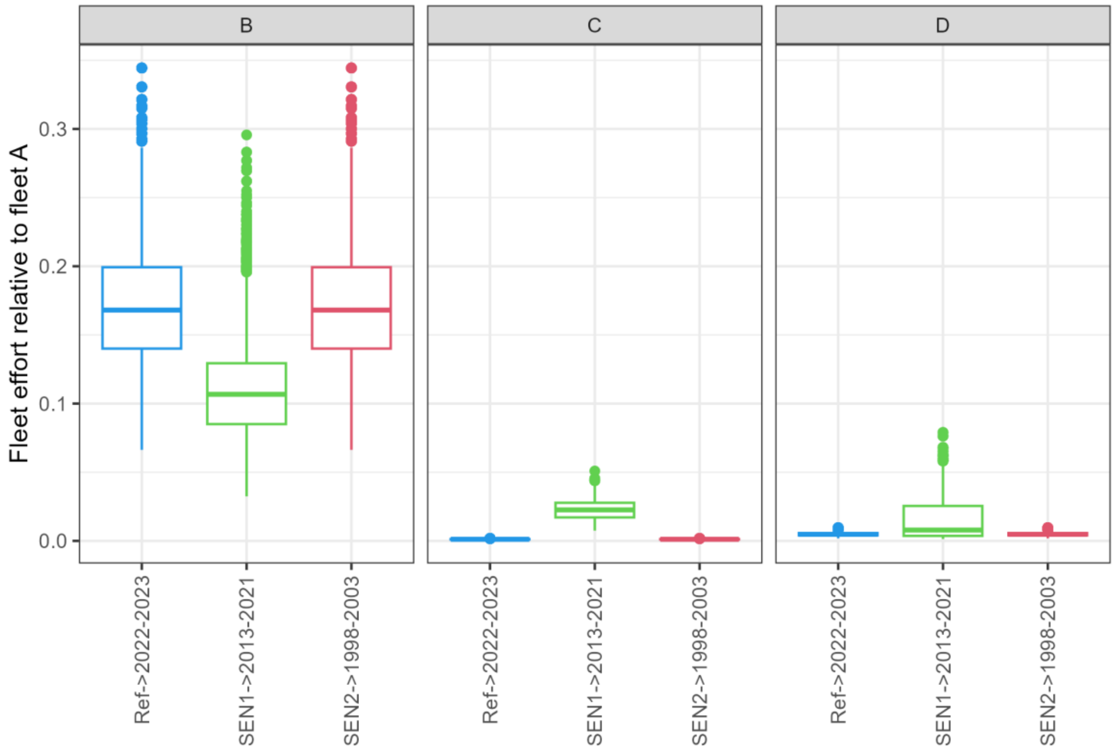


Figure 4.9: Statistics of relative fleet effort across replicates for the time periods over which it is resampled: base and other OMs in the reference set (2022-2023), SEN1 (2013-2021) and SEN2 (1998-2003).

4.4 Fishing selectivity

The future selection patterns are assumed to follow an age-correlated random walk (similar to the design in the SAM assessment). Starting from the 2023 estimated selection pattern, each of the following years' selection is obtained by modelling a change in selection-at-age to the next year as differences in log-transformed F-at-age (log-differences). All log-differences from one year to the next for the projected time-series follow a normal distribution with zero mean and a variance-covariance matrix of log-differences over the years 2014–2023 (from year y to year y+1 within each age). To prevent extreme F-at-age changes, each generated log-difference was checked; a log-difference was kept if it resulted in log-transformed F-at-age within ± 1.96 times the age-specific SD (i.e. 95% CI) of the original 2014-2023 values, and re-generated if not.

The fishing selectivities across base and other OMs in the reference set (Ref), 2013-2021 (SEN1 OM) and 1998-2003 (SEN2 OM) are shown in Figure 4.10. For the A-fleet which is dominating the fishery, it can be observed that the selectivity at age is increasing with age in recent years whilst more dome shaped over the period 1998-2003. The resampling in the different periods translates into different fishing selectivity patterns as shown in Figures 4.11-13.

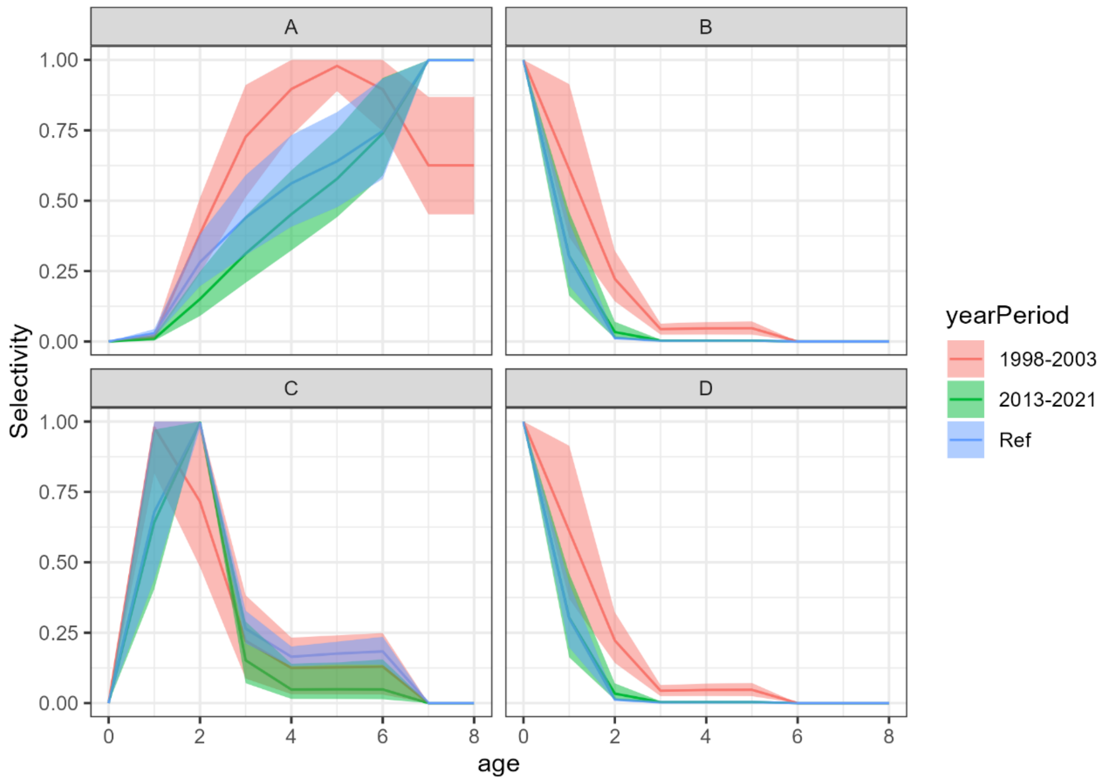


Figure 4.10: Fleet selectivity at age over which the process generating future selectivity is resampled: base and other OMs in the reference set, Ref (random walk using the last 10 years), SEN1 (2013-2021) and SEN2 (1998-2003). Ribbons are the 95th/5th quantile*S*.

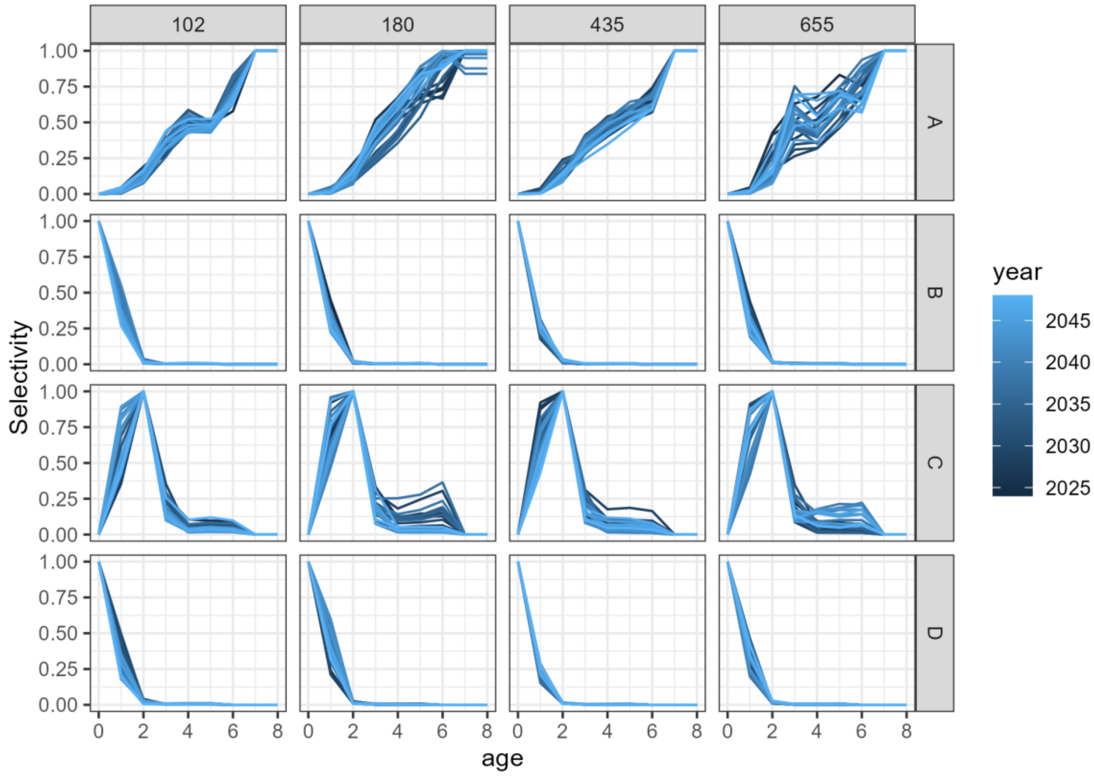


Figure 4.11: Base and other OMs in the reference set, Ref (random walk using the last 10 years). Example of fleet selectivity development in projections for four different replicates.

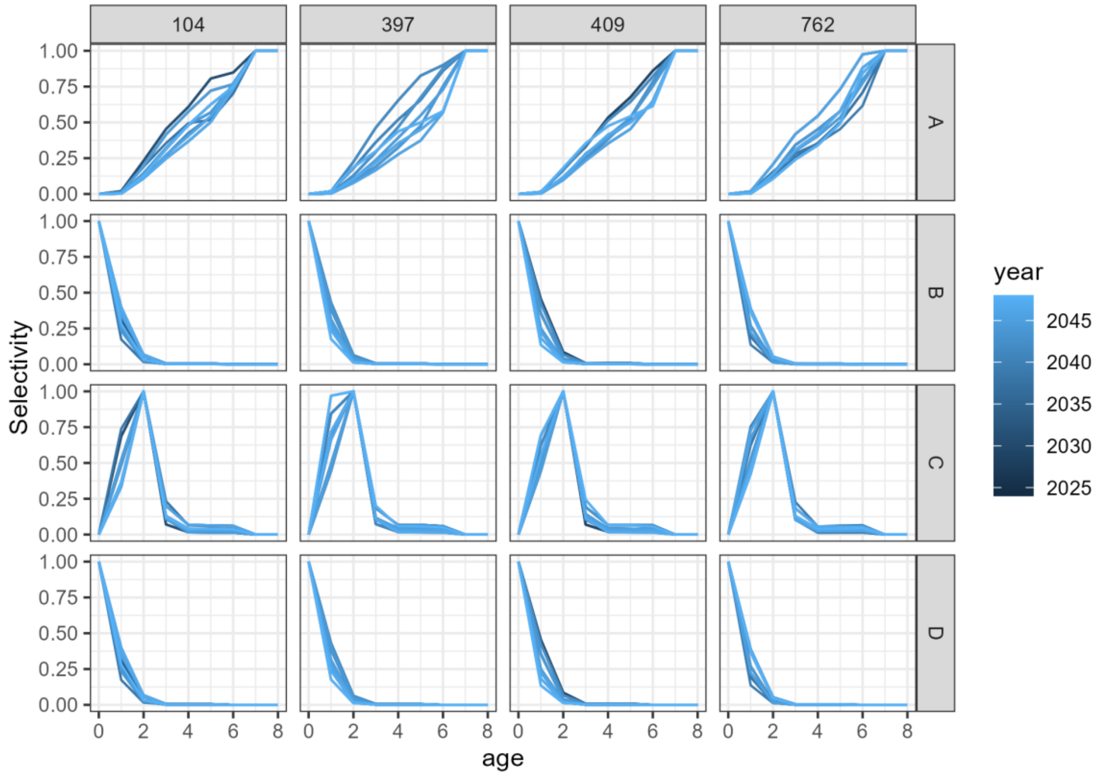


Figure 4.12: SEN1 sensitivity test (exploitation resampled over 2013-2021). Example of fleet selectivity development in projections for four different replicates.

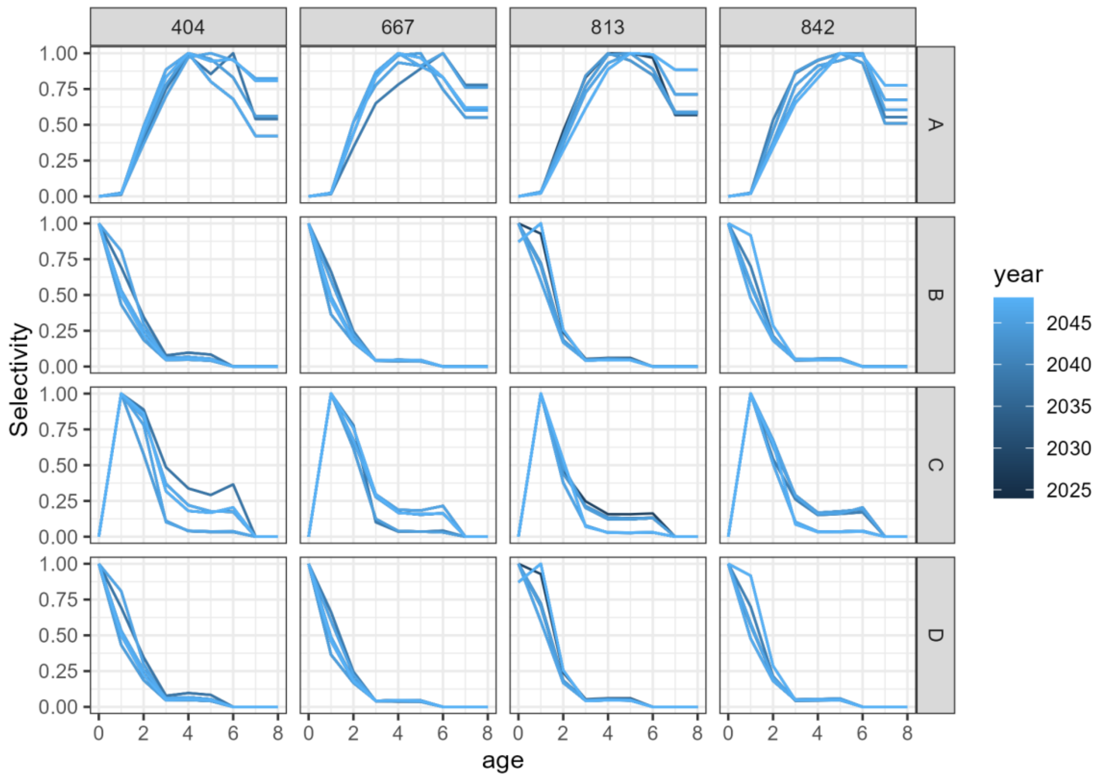


Figure 4.13: SEN2 sensitivity test (exploitation resampled over 1998-2003). Example of fleet selectivity development in projections for four different replicates.

4.5 Natural mortality

The natural mortality is based on the output of the SMS multi-species model from WGSAM (ICES (2024b)) which spans the period 1974-2022. The raw values are first smoothed using a loess smoothing (span of 0.5). For years prior to 1974 and after 2022, a rolling average over the 5 most recent years is used. The resulting natural mortality times series at age are then scaled by profiling the assessment model over a range of additive scaling (Figure 3.10). As of HAWG 2024, the additive scaling is addM = 0.02, a value that is used in all OMs except those considering deviations in natural mortality (Table 3.2, namely M2, M3, M4 and M5). The OMs considering uncertainties in natural mortality deviate in different ways to the procedure above and will be described separately in the following sections.

One specificity of the natural mortality scenarios (M1-5) is that whilst each replicate is conditioned on different natural mortality vectors, the natural mortality taken as observations (i.e. used for fitting the estimator in the management plan) is the one from HAWG. This mimics a model misspecification, simulating the case of having an underlying natural mortality that differs from the one applied in the yearly assessment model used to derive catch advice.

The comparison between the different natural mortalities at age is shown in Figure 4.14.

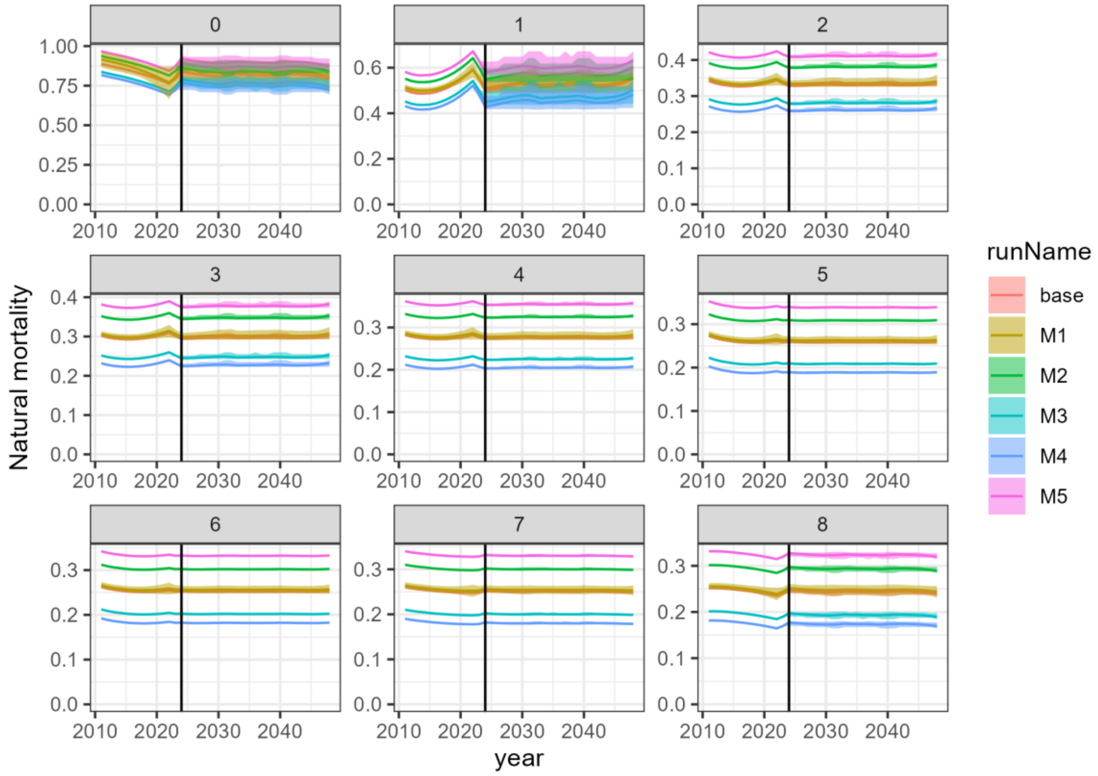


Figure 4.14: Time series of natural mortality at age for the different OMs considered. Ribbons are the 95th/5th quantiles. The black vertical line is the year at which projections are starting.

4.5.1 M1: replicates in predation mortality

OM M1 considers replication in the predation mortality using the variance-covariance matrix from the 2023 SMS multi-species model (ICES (2024b)). A set of 1000 replicates of natural mortality at age are first generated (Figure 4.15). For each replicate, the SAM stock assessment model is profiled over additive M scaling (as shown in Figure 3.10). This profiling procedure results in an addM between 0.01 and 0.03 (Figure 4.16) which is a very narrow range of values over the value of the base assessment model (0.02). Using the optimal additive M scaling for each replicate, the stock assessment is then fitted. Therefore, each replicate benefits from an assessment fit that is used to draw further uncertainty from the variance-covariance matrix derived from the SAM model fit, as described in Section 3.3 and 3.4. The resulting uncertainty over 1000 replicates is like the base OM as shown in Figure 4.17.

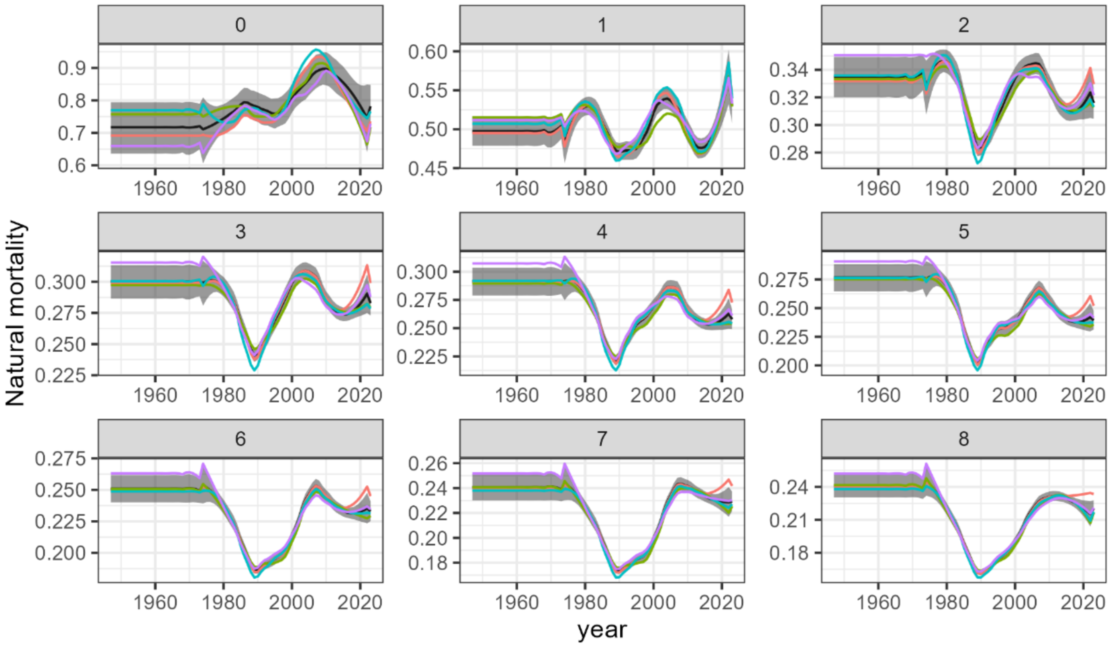


Figure 4.15: Replicates in natural mortality. The graph shows individual replicates together with the 95%/5% quantiles as grey shaded areas.

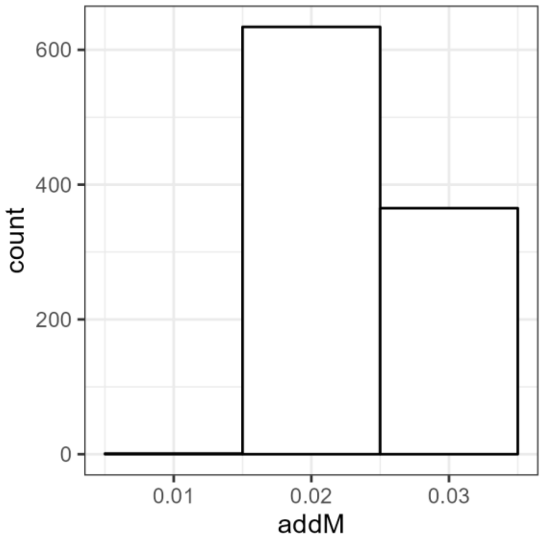


Figure 4.16: Additive M resulting from the profiling of individual assessment models imputed with 1000 replicates of natural mortality.

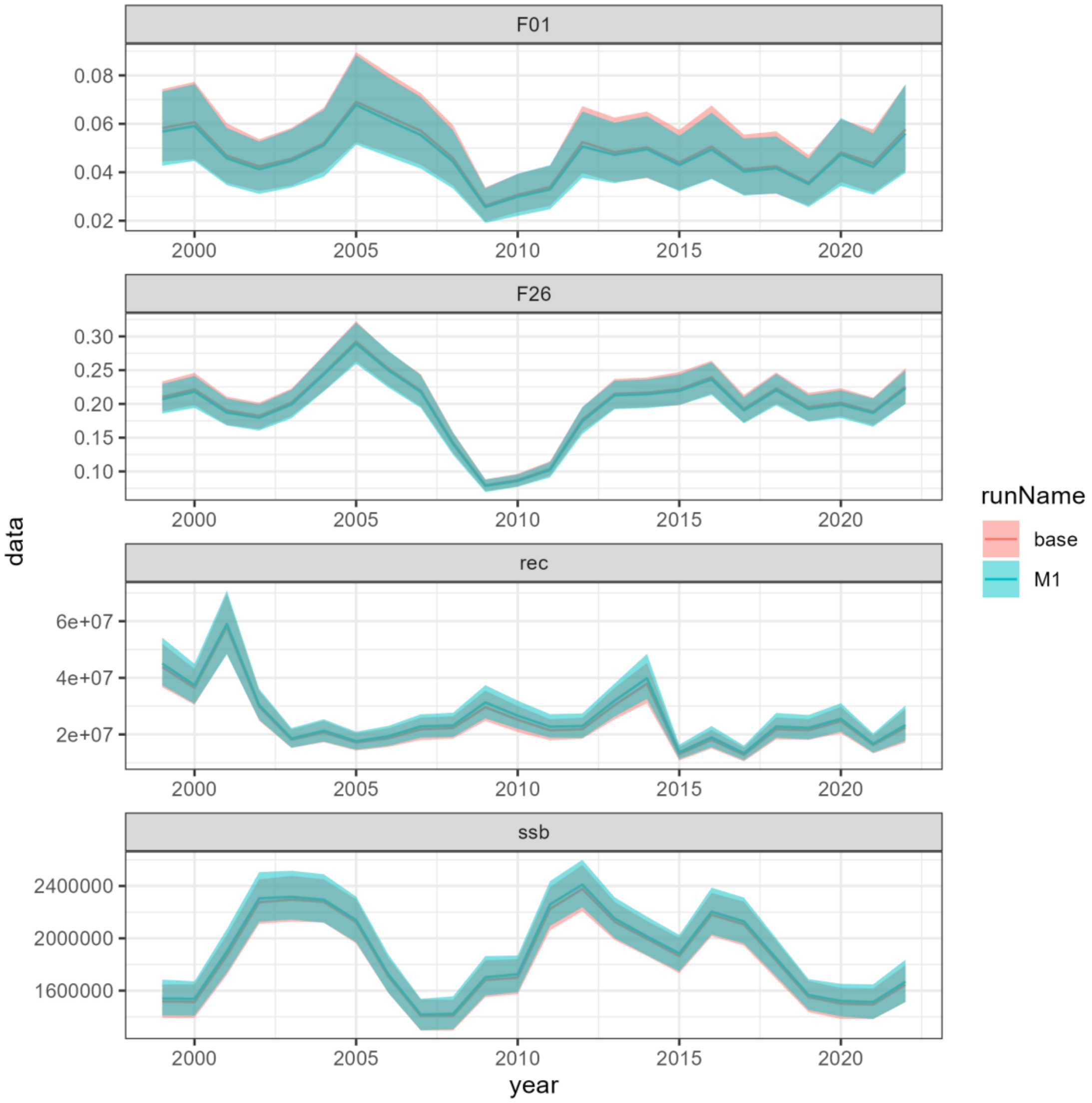


Figure 4.17: Stock trajectories (F01, F26, rec and SSB) over the historical period for the base and M1 OMs. Ribbons are the 95th/5th quantiles.

4.5.2 M2, M3, M4 and M5: scaling of natural mortality

Natural mortality for the NSAS herring assessment is derived from the SMS multi-species model (ICES (2024b)) as described above. These values are updated triennially by the ICES working group on Multi-Species Assessment Methods (ICES (2024b)). Given this 3-year cycle, it is not uncommon that mortality rates shift between updates of SMS assessments. Taking these new values directly as input may cause considerable shifts in the perception of the NSAS herring stock. To alleviate this issue, the profiling method described previously was developed at the 2018 benchmark (ICES (2018)) and 2021 inter-benchmark (ICES (2021)) to smooth the transition between these cycles. This process is, however, based on a statistical fit of the stock assessment model which does not alleviate potential uncertainties in the absolute level of natural mortality. In that context, the M2 and M3 OMs aim at encapsulating this uncertainty in the OM reference set (i.e. cases considered in the tuning of the HCR). More specifically, the M2 OM considers higher natural mortality levels whilst the M3 OM considers lower natural mortality levels. OMs M4 and M5 consider further extremes in the bandwidth of absolute level of natural mortality but only as robustness tests. Here, the methods employed to derive the M2, M3, M4 and M5 OMs are explained.

For M2 and M3, the profiling of the base assessment over stock replicates is employed. This is in line with the estimation uncertainty of all other parameters to generate OMs. Using the variance-covariance matrix derived from the SAM model fit, 1000 stock replicates and associated new sets of observations were generated. The SAM stock model was then fitted to each individual set of observations and the profiling method from the 2021 inter-benchmark (ICES (2021)) was applied, spanning additive scaling in natural mortality from -0.1 to +0.1 in steps of 0.01. This procedure yielded additive scaling in natural mortality for each replicate, in turn providing the distribution of additive scaling across the 1000 replicates. This distribution is shown in Figure 4.18. It can be observed that the additive scaling that is cumulative to the 0.02 from the base assessment is between -0.08 and +0.08 with mean of 0, and the 95% quantile range is 0.05 to 0.05. Using these results, the M2 and M3 OMs are constructed with:

- M2: adding 0.05 to the HAWG 2024 natural mortality vector, leading to a total additive scaling of 0.07.
- M3: subtracting 0.05 to the HAWG2 024 natural mortality vector, leading to a total additive scaling of -0.03.

Drawing uncertainty in additive scaling from assessment stock replicates was deemed more consistent with the OM conditioning for this MSE exercise. For this reason, M2 and M3 OMs were included in the reference sets. However, an alternative way to draw uncertainty in additive scaling is to use the profiling of the base assessment. From this profile, a 95% confidence interval can be drawn for the range of parameters for which the log-likelihood lies within 1.92 of the maximum log-likelihood value. This is exemplified in Figure 4.19 (black horizontal line as the 1.92 level offset). The corresponding natural mortality time series are used for the M4 and M5 OMs that are only considered as robustness tests. More specifically:

- M4: total additive scaling of -0.05
- M5: total additive scaling of 0.1

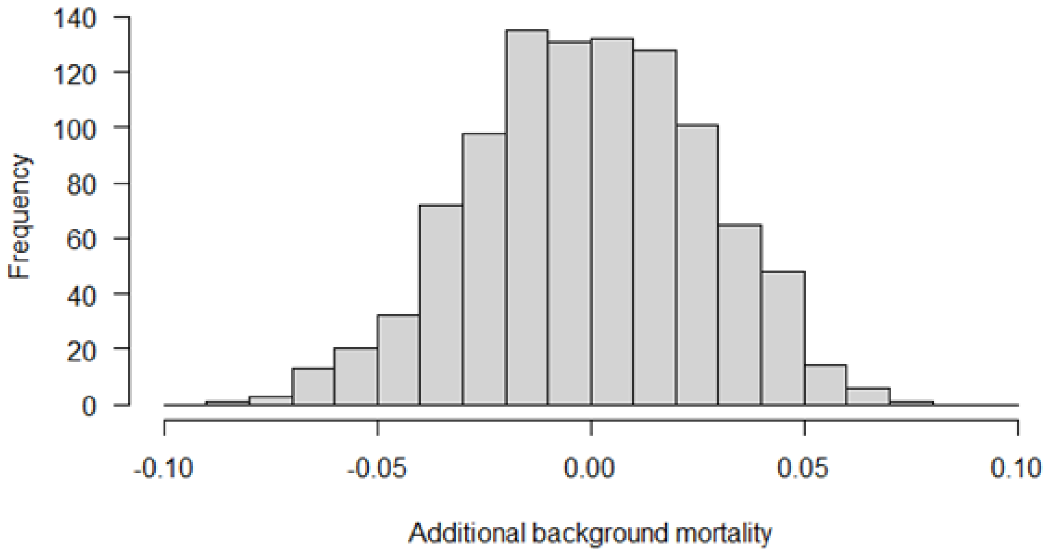


Figure 4.18: Estimated additional background mortality for 1000 replicates by scanning over additive M within each replicate.

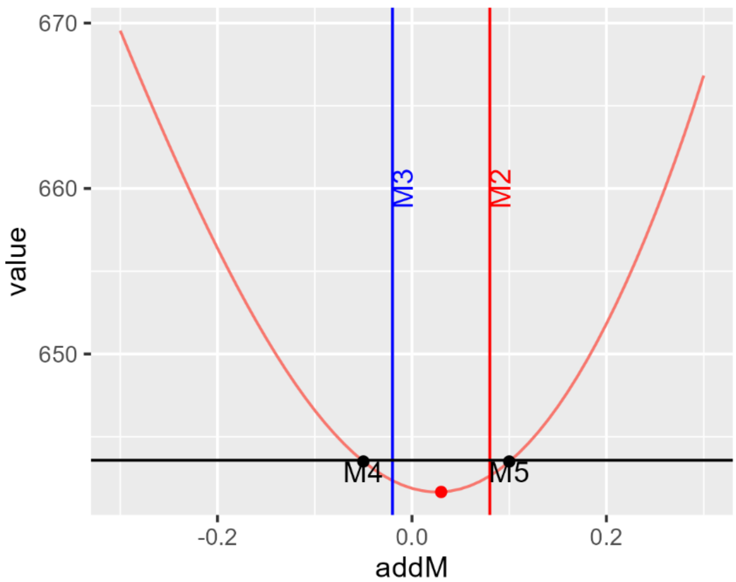


Figure 4.19: Stock assessment model profiling over different level of additive M scaling and corresponding natural mortality OMs. The OMs from the reference set are M2 (red vertical line, addM = 0.07) and M3 (blue vertical line, addM = -0.03) that have their additive rescaling derived from the distribution of assessment profiling over 1000 stock replicates (Figure 4.1). The OMs M4 and M5 are part of the robustness test and are derived as an offset of 1.92 points in log likelihood from the optimum (red point).

4.6 Recruitment, SR function and deviances

Simulating recruitment into the future is a key aspect of the MSE exercise. Assumptions about recruitment often drive the trajectory of the stock and, in turn, the ability to harvest it at higher or lower fishing mortalities. For this MSE, stock recruitment relationships are derived for each replicate, after these are drawn from the stock assessment model variance covariance matrix (Section 4.1). Here, five different bases of stock recruitment underpin the different OMs (Table 3.3). More specifically:

1. Base: rationale relating to previous ICES working group on NSAS herring, i.e. recovery dynamics from the full period and productivity from recent period.
   - Fit using a cropped time series (2002-2024), fixing steepness based on the full time series (1947-2024 excluding 1979-1990).
   - No depensation parameter
2. SR1: what if recovery dynamics is also inferred from the recent time period?
   - Fit using a cropped time series (2002-2024)
   - No depensation parameter
3. SR2: what if productivity is inferred from the full period?
   - Fit using the full time series (1947-2024 excluding 1979-1990)
   - No depensation parameter
4. SR3: what if depensation is taking place?
   - Fit using the full time series (1947-2024 excluding 1979-1990)
   - Inclusion of depensation parameter
5. SR4: what if autocorrelation is playing an important role?
   - Fit using the full time series (1947-2024 excluding 1979-1990), using autocorrelation as a parameter
   - No depensation parameter

For all OMs except SR3, a 50%/50% mix of stock recruitment functions between segmented regression (Segreg) and Beverton and Holt (BH) is used. For SR3, because of the poor statistical fit for the segmented regression function including depensation, only the BH function is considered (i.e. 500 replicates). The fits between the different recruitment functions are shown in Figure 4.20. Expectedly, the fits over the long time series lead to higher productivity and steepness.

The recruitment functions are used to generate recruitment in projection years. On top of these, deviances are applied as an additional multiplier. These deviances are predefined for each replicate. These are generated using a log normal distribution with sd estimated ad hoc from the fit of the recruitment function for each replicate. The deviances for the different OMs are shown in Figure 4.21.

For SR4, however, these deviances are correlated in time. For the BH SR relationship, autocorrelation is estimated directly when fitting the SR model to the stock-recruit pairs (and therefore affecting the fit of the relationship as well). For the Segreg SR relationship, autocorrelation is calculated from the residuals of the model fit. For both models, auto-correlated deviances are generated from an auto-correlated log-normal distribution with a mean of zero and sd as estimated from the residuals.

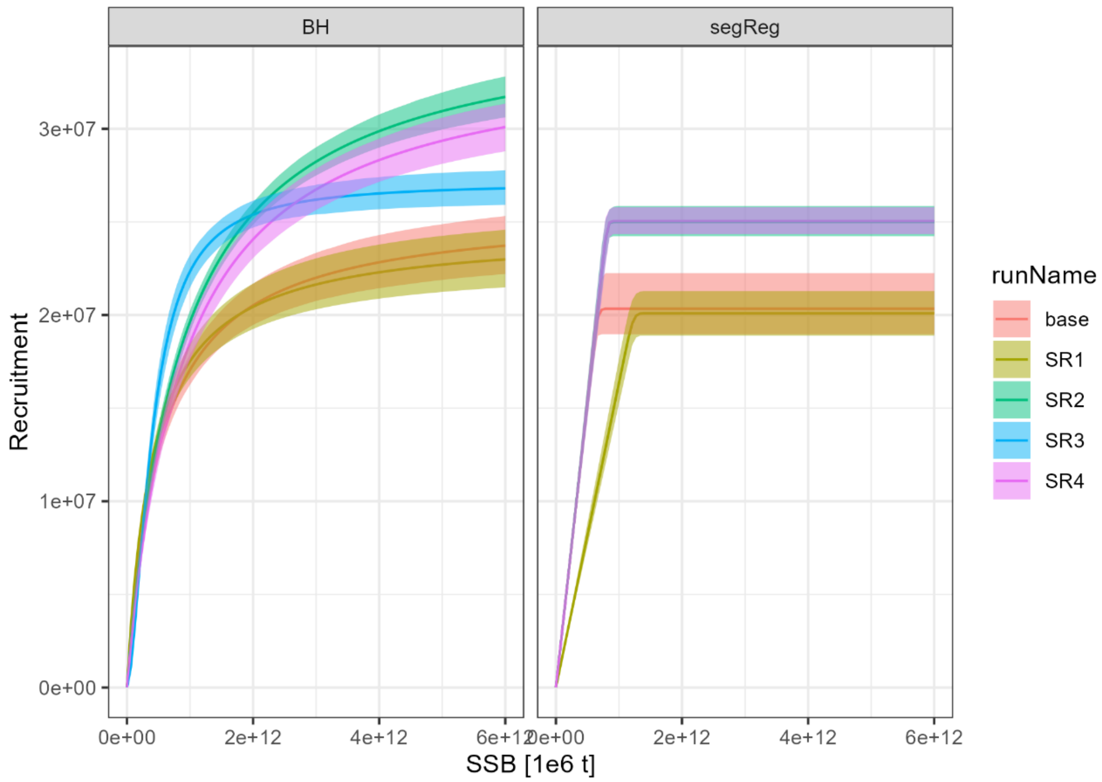


Figure 4.20: Stock recruitment functions for the different OMs considered. All OMs except SR3 is a 50% mix of Segmented Regression (segReg facet) and Beverton-Holt (BH facet). Ribbons are the 95th/5th quantiles over 500 replicates.

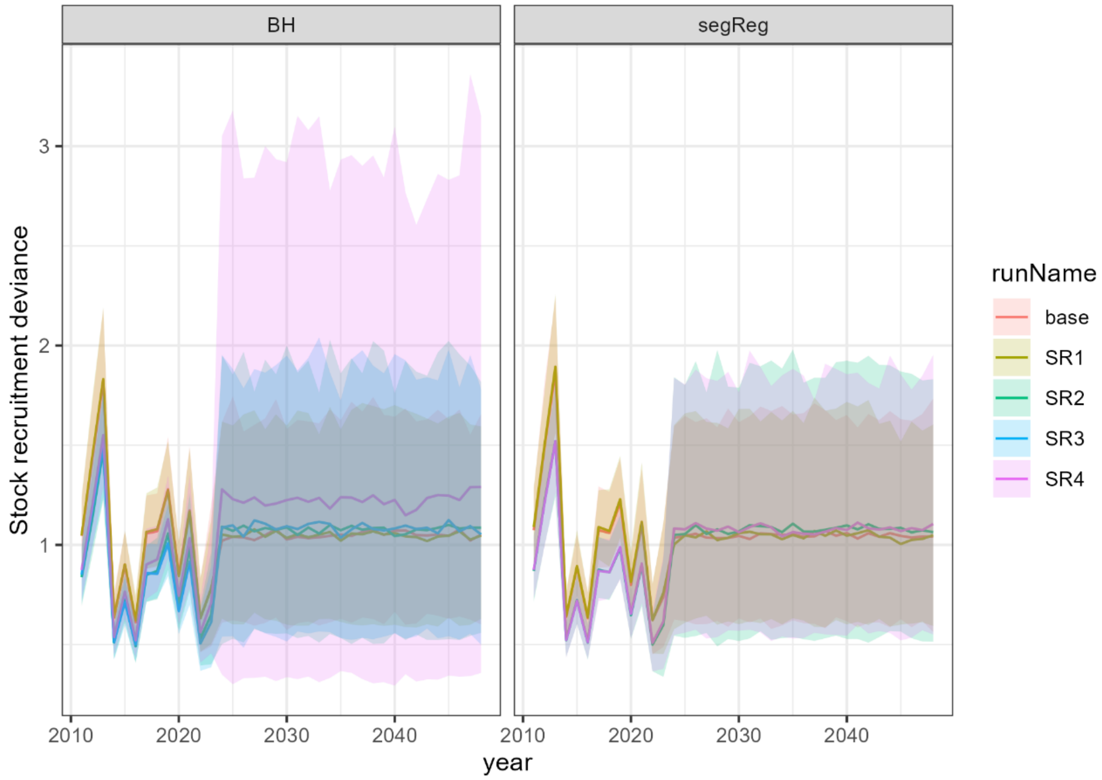


Figure 4.21: Recruitment deviance for the different OMs considered. All OMs except SR3 is a 50% mix of Segmented Regression (segReg facet) and Beverton-Holt (BH facet). Ribbons are the 95th/5th quantiles.

4.7 Process error

The SAM model estimates process error on the log-transformed numbers-at-age. Process error can be interpreted as departures not explained by fishing or the assumed natural mortality at each age (e.g. migration). New process errors must be generated in OM projections, because other OM values estimated by or derived from SAM are conditioned on the process error estimates. Practically speaking, OM projections without process error will result in new numbers-at-age that are higher and lower than they should be. Process errors are assumed to follow a normal distribution with zero mean and age-specific standard deviations output by SAM. These standard deviations are used with a truncated normal distribution (with truncation at ± 3sd) to generate new process errors for OM projections. The equation for incorporating these process errors is shown in Section 4.8.4. The level of process error for the different ages is presented in Figure 4.22.

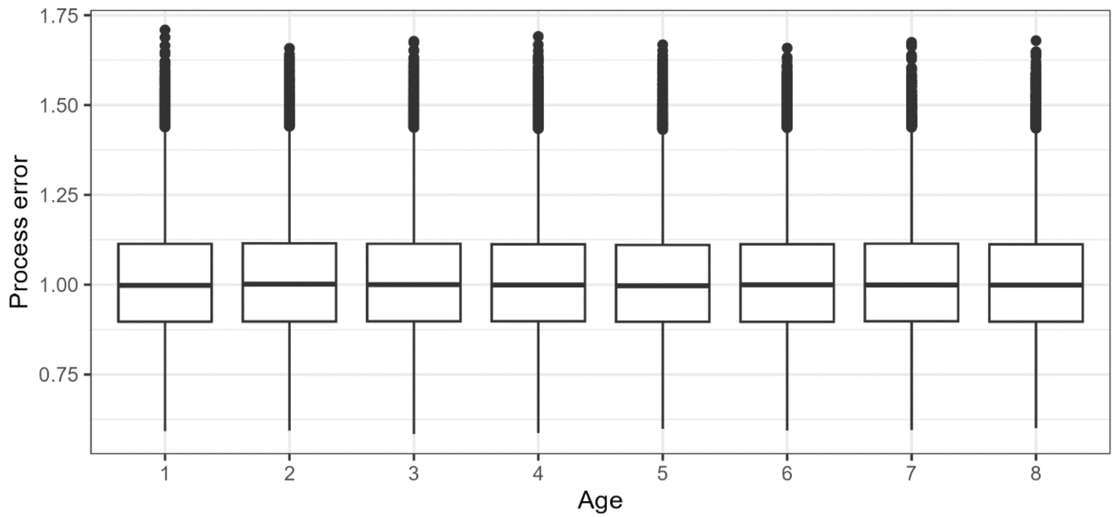


Figure 4.22. Process error deviance applied to stock numbers at age in the projection period.

4.8 OM Projection

The OM projections are done using the FLasher package (Scott and Mosqueira (2023a,b)) from the FLR framework (Kell et al. (2007)). In order to describe the process underlying the projection, the documentation from the FLasher package will be used, more specifically the FLasher_reference vignette (Scott and Mosqueira (2023a)).

4.8.1 Calculating the fishing mortality

Fishing effort and fishing selectivity (defining the exploitation pattern) drives the catches and biological stock abundance through fishing mortality. As such, it is a key metric of the model, as explained in Section 4.3. Fishing mortality (F) is an age-structured metric that represents the impact of fishing on the stock. The fishing mortality imposed on the fish stock from a catch amount is known as the partial fishing mortality. This is because in FLasher, fishing mortality can be split across ages into different stocks and different fleets. For NSAS herring, the fish stock is fished by multiple fleets (A-D). The partial fishing mortality is calculated as:

$$pF_{a,c} = Sel_{a,c}E_c$$

where a is age, and c the index relating to the fleet realising the catches. The quantity pF is the partial fishing mortality, Sel is the fleet selectivity and E is the fishing effort. The effort E is defined as presented in Section 4.3 (maximum fishing pressure over all ages, or apical F). The total fishing mortality imposed on a stock from all the fisheries is the sum of the partial fishing mortalities:

$$F_a = pF_{a,c_A} + pF_{a,c_B} + pF_{a,c_C} + +pF_{a,c_D}$$

4.8.2 Projecting the fisheries

In each time step of the projection, the landings and discards numbers are updated. In the case of NSAS herring, discard is set to 0 as bycatch in fisheries other than fleet B and D is considered to be negligible. Catch numbers at age in each time step are calculated using the Baranov equation. For NSAS herring, the stock is fished by fleets A-D. The catch of each fleet is the partial catch. The total catch equals the sum of the partial catches from that stock:

$$C_a = pC_{a,c_A} + pC_{a,c_B} + pC_{a,c_C} + pC_{a,c_D}$$

The partial catch is given as:

$$pC_{a,c} = \left(pF_{a,c}/Z_a\right) \times (1 - e^{-Z_a}) \times N_a$$

with Z the total mortality of the stock and N the abundance at age. The total mortality is given as:

$$Z_a = F_a + M_a$$

with M the natural mortality at age.

4.8.3 Projecting the stock

The biological stock is projected one timestep at a time by the FLasher package. The method calculates the survivors from the previous timestep and places them in the current timestep. Recruitment in the current timestep is calculated using the stock recruitment function. More specifically, the projection sequence is as:

1. Calculate total mortality (Z) on the stock in the previous timestep (see above).
2. Calculate survivors (S) from the previous timestep.
3. Calculate recruitment for the current timestep.
4. Place survivors and recruitment in the appropriate age classes in the current timestep.

4.8.4 Calculating survivors

Abundance at age in a given time step is at the start of a timestep, i.e. before any fishing or natural mortality occurs. The survivors S are the abundances at age at the end of a timestep in a given year y, calculated as:

$$S_{a,y} = N_{a,y} \cdot e^{-Z_{a,y}} \cdot e^{\varepsilon_{a,y}}$$

where $N_{a,y}$ is the population abundance at the start of the timestep and $\varepsilon_{a,y}$ is the age- and year-specific process error on survival (Section 4.7). The survivors are put into the abundances in the next timestep. The age group that survivors are placed in depends on the timestep and the timing of recruitment. For the MSE, the recruitment is taken at age 0 and timesteps are of 1 year. The plus group is age 8. As the projection progresses by 1 year, the survivors are placed in the next age group.

4.8.5 Recruitment

Recruitment is one of the most important biological processes as it drives the dynamics and productivity of the stock. It is also a source of great uncertainty and can be strongly affected by external drivers such as environmental conditions. In Flasher, calculating recruitment has two main stages:

- Calculating the spawning stock biomass (SSB)
- Calculating the recruitment from the SSB: $R = f(\text{SSB})$

SSB in year y is calculated as:

$$SSB_y = \sum_{a=1}^{A} N_{a,y} \times \text{Mat}_{a,y} \times Wt_{a,y} \times exp\left(-Z\text{prespwn}_{a,y}\right)$$

Where A is the plusgroup in the stock, Mat is the proportion mature, Wt is the mean weights at age and Zpreswn is the amount of mortality (fishing and natural) experienced by the stock before spawning occurs in the current timestep. The timing of spawning for NSAS herring is 0.67 (1st September).

In the MSE, recruitment is modelled as a single unit, meaning a single recruitment event takes place. It is considered appropriate for the MSE modelling purpose though in practice NSAS herring has four different spawning components with spawning timings in both autumn and winter. The recruitment functions considered here are described in Section 4.6. The calculated recruitment is inserted into the abundance in the same timestep meaning that recruitment is calculated at the start of the timestep.

The final step in calculating recruitment is the application of deviances. This is an additional multiplier applied to the calculated recruitment. The deviances are used to introduce further variability in recruitment.

4.8.6 Projection targets

Generally, projections are controlled by targets using FLasher (Scott et al. (2023a,b)). Targets cannot be specified by age and are aggregated over all ages. However, the targets are specific to each stock replicate. Targets can be set to different components of the model, e.g. the entire stock itself, or a specific fishery. The catch and fishery targets are specific to each fleet. Furthermore, F targets also require the age range over which to calculate it. These targets can be defined as relative to another fleet (e.g. effort of B-fleet relative to fleet A). See next section for specifics to the NSAS herring MSE.

4.8.7 Evaluation of targets

To solve the projection, the FLasher package attempts to find the fishing effort values in the appropriate timestep to hit the desired target:

$$e = t - \hat{t}$$

Where t is the target defined, $\hat{t}$ is the state of the operating model at a given level of fishing effort (where we are) and the error e is the difference between the two. FLasher attempts to find the fishing efforts so that $\hat{t}$ minimizes e. The projection targets for this MSE are described in Section 4.9, and ensure that the single TAC from the MP is converted into fleet-based catches in the OM, respecting the relevant fleet effort and fishing selectivity scenarios (Sections 4.3 and 4.4 respectively).

4.9 Implementation system

To implement the single TAC into four different fleets, the implementation system is undergoing an optimization process based on different targets. This is done to constrain fluctuations in behaviour of the four fleets. This process is undertaken using the Flasher package from the FLR framework, as described in the sections above.

Important to the projections are the following:

- Catch quota is what is implemented by managers in practice. Therefore, the catch should correspond to the catch quota from the management strategy.
- The distribution of the catch quota in term of fishing pressure between the different fleets should mimic selected past periods (reference set as 2022-2023, SEN1 as 2013-2021, SEN2 as 1998-2003). The period over

which fleet behaviour is resampled defines selectivity at age and fishing effort (Sections 4.3 and 4.4).

Under these two principles, projections are made using the following four projection targets:

1. $C_{\mathrm{OM}}(y) = \mathrm{TAC}(y)$
2. $E_B(y) = r_{B/A} \times E_A(y)$
3. $E_C(y) = r_{C/A} \times E_A(y)$
4. $E_D(y) = r_{D/A} \times E_A(y)$

with E_i the effort of fleet i as defined in Sections 4.3 and 4.8.1. The quantity $r_{i/A}$ is the ratio of effort i relative to fleet A (see Figure 4.7 and 4.8). These are sampled over selected periods (Table 3.2). The projection based on fishing effort relative to fleet A links fishing pressure on juveniles to fishing pressure on adults (Figure 4.23). This is because a decrease or increase of fishing pressure is implemented in fleet A (targeting adults) and the relative decrease or increase in effort of this fleet is followed by other fleets (e.g. fleet B and D with a bycatch of juveniles).

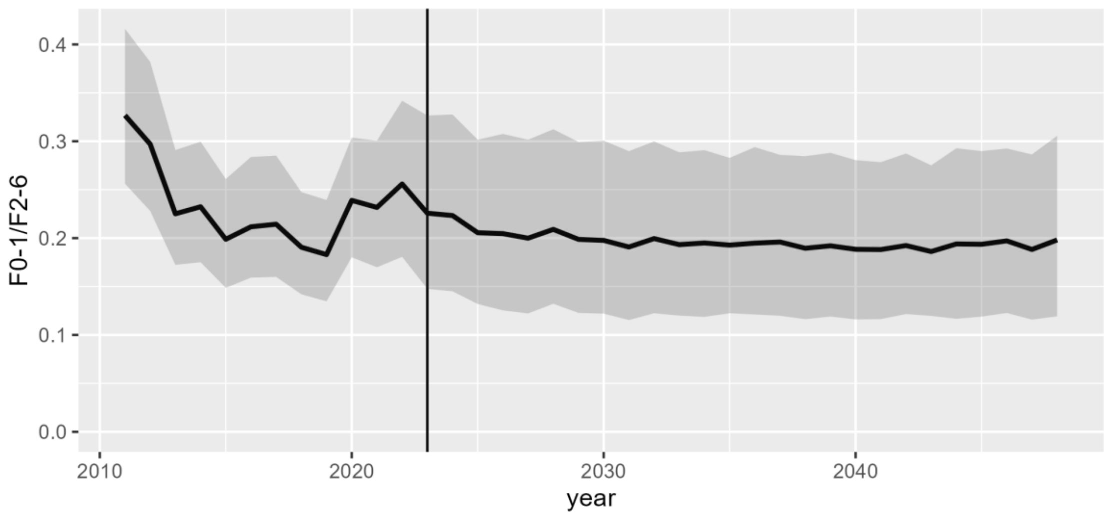


Figure 4.23: Yearly ratio of F0-1 to F2-6 for projections under F=0.2 constraint. This quantity remains independent from the choice of fishing pressure on adults. Ribbons are the 95th/5th quantiles.

4.10 Observation Error Model (OEM)

In each projection timestep, catch and survey data are generated. The catch at age is generated directly from the OM by applying the deviances to each catch at age:

$$C_{\mathrm{obs}}(a,y) = C_{\mathrm{OM}}(a,y) \times \mathrm{dev}_C(a,y)$$

with dev_C being catch at age deviance.

The survey indices are generated as:

$$I_{a,y,i} = q_{a,i} \times N_{a,y} \times e^{-t_i \times Z_{a,y}} \times \mathrm{dev}_{Ii}(a,y)$$

with $I_{a,y,i}$ and $q_{a,i}$ the survey index numbers of survey i at age a in year y and catchability of survey i at age a, respectively. The quantity t_i is the timing of the survey and dev_{Ii} is the index at age deviance.

The residuals for catch at age and survey at age are drawn from a log normal distribution, using standard deviations estimated by the SAM assessment model. This is specific to each data source, each OM and each stock replicate. The resulting residuals on catch at age are shown in Figure 4.24. The resulting index at age residuals are shown in Figure 4.25.

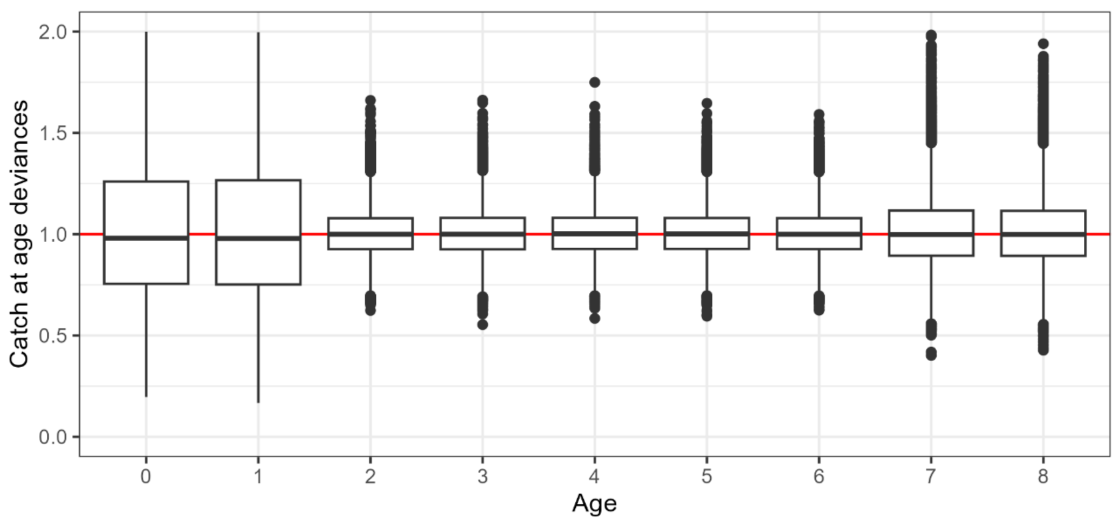


Figure 4.24: Distribution of deviances in catch at age for the projected years.

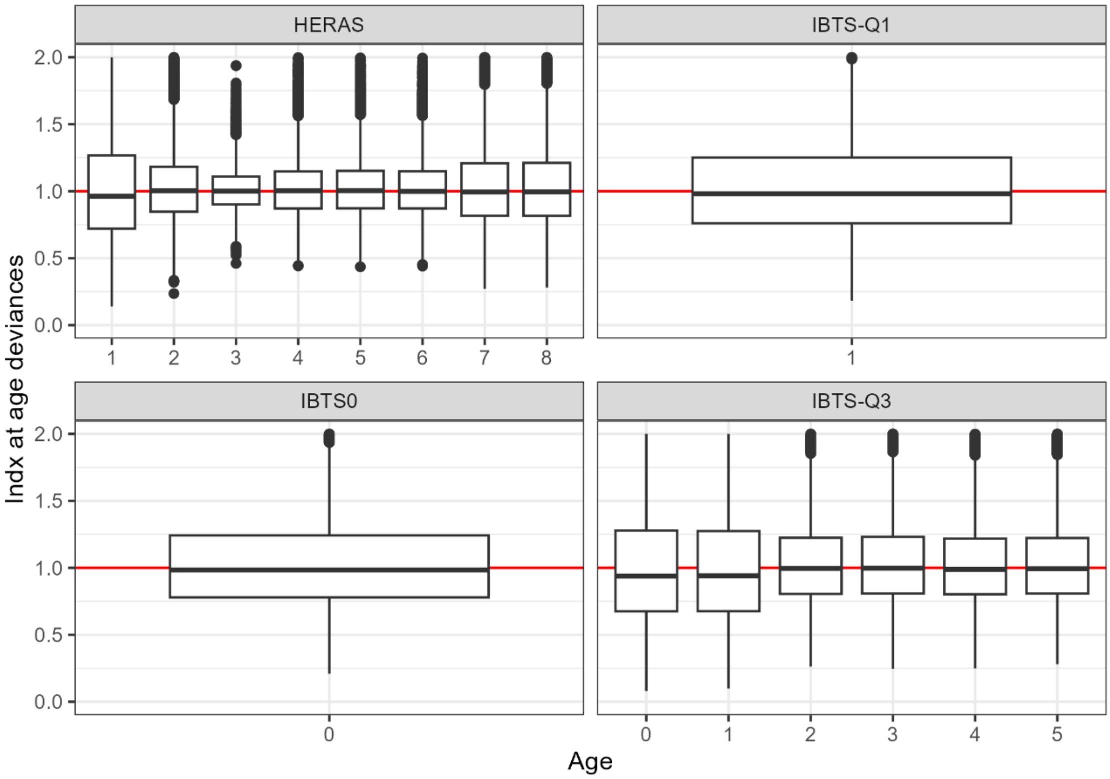


Figure 4.25: Distribution of deviances in index at age for the projected years.

4.11 Biological reference point

The biological reference point Blim is of importance for the MSE because the risk performance metrics are calculated as the probability of being below Blim. The derivation of Blim is done similarly to previous derivation during ICES workshops and working groups on NSAS, i.e. WKPELA (ICES (2018)), IBPNSHerring 2021 (ICES (2021)), HAWG 2024 (ICES (2024a)). More specifically, a segmented regression stock recruitment function is first fitted. Then, Blim is taken as the breakpoint of this fit, i.e. the *b* parameter. For NSAS, the time series used to fit the recruitment function spans 1947-2024 excluding years over which stock recovery is taking place (1979-1990). In the MSE, the Blim values are specific to each replicate as each replicate exemplifies a unique set of stock recruitment pairs. The distribution of Blim values is shown in Figure 4.26 for the different OMs. As anticipated, the distribution of Blim values is the same for the base, SR1, SR2, SR3 and SR4 OMs, because the conditioning of the replicates is the same. However, changes in Blim are induced when altering natural mortality, i.e. for M1, M2, M3, M4 and M5. In Figure 4.26, one can observe a shift of the breakpoint to either side of the base OM scenario when natural mortality is scaled up (M2 and M5) or scaled down

(M3 and M4). This aspect is exemplified in Figure 4.27 by comparing the Blim segmented regression fitting for the base, M2 and M3 OMs.

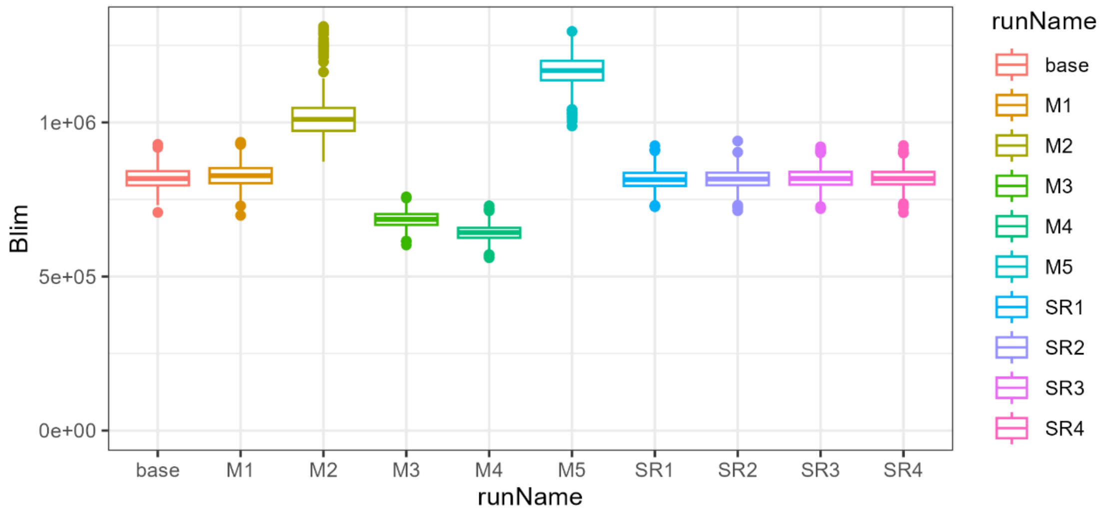


Figure 4.26: Distribution of the biological reference point Blim across replicates for all OMs considered.

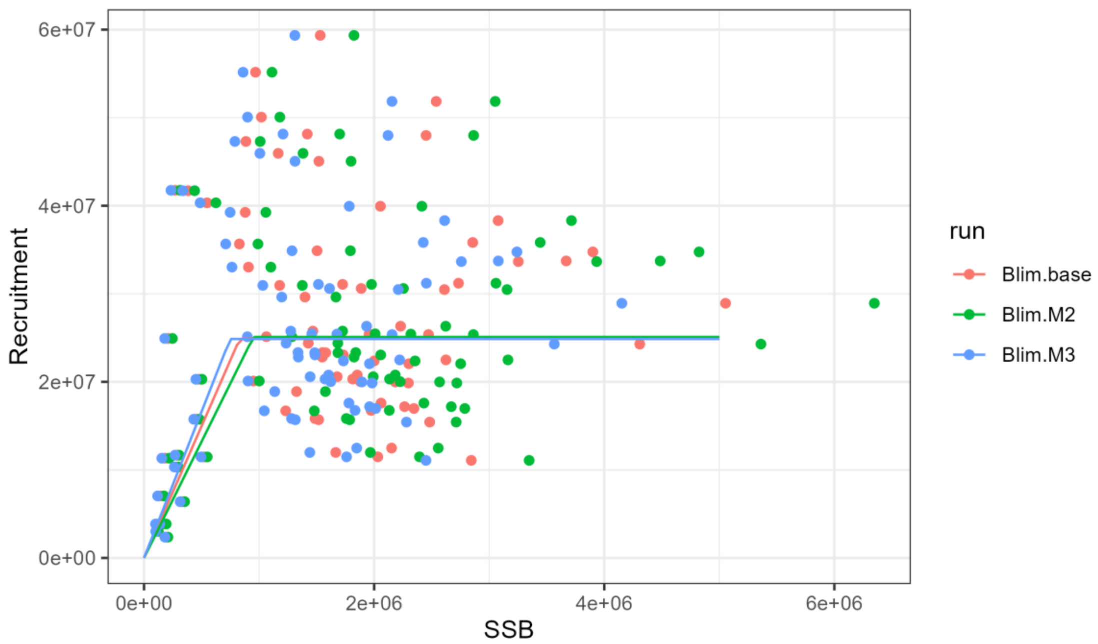


Figure 4.27: Contrast in the derivation of the Blim biological reference point for the base, M2 and M3 OMs. Using a segmented regression fit, Blim is taken as the breakpoint.

5 Management Procedure (MP) and tuning

The management procedure is model-based and uses the stock assessment model described in Section 3.4 as the Estimation model (Figure 3.15). It should be noted that even though the models that are used to construct the OMs are similar in structure to the stock assessment model used as the Estimation model, they are independent of one another in the MSE and do not necessarily share the same assumptions (e.g. biological parameters, recruitment models, etc.), and only the monitoring data from the OM is passed onto the Estimation model (Figure 3.15). In the management procedure, stock status (more specifically SSB), is estimated using this Estimation model. From this stock status, the management procedure consists of the following rules:

1. A Harvest Control Rule (HCR) with a fishing mortality equal to the target F when SSB is at or above Btrigger. In the case that the SSB is forecasted to be less than Btrigger at spawning time in the year for which the TAC is to be set, the TAC shall be fixed consistently with a fishing mortality that is given by: F=Ftarget×SSB/Btrigger. This HCR is shown in Figure 3.16.
2. A constraint on the inter-annual variation of TAC is applied when the HCR would lead to a TAC that deviates by more than 20% below or 25% above the TAC of the preceding year. In such a case, the TAC is respectively set as 20% below or 25% above the TAC of the preceding year. The TAC constraint only applies if the SSB at spawning time in the year for which the TAC is to be set is higher or equal to Btrigger.
3. A 10% banking and borrowing mechanism is allowed. Banking and borrowing should be suspended when SSB is below Btrigger. The impact of this is only tested under one scenario in SEN6.

The combination of Ftarget and Btrigger defines the HCR. It is important to note that Ftarget and Btrigger are control points for the HCR, based on management objectives and decisions, and should not be confused with reference points for the stock (e.g. Fmsy and MSYBtrigger or Bpa when MSYBtrigger is not estimated). The process of evaluating the different combinations of Ftarget-Btrigger is the tuning. This evaluation is done against performance metrics defined in a given time horizon.

5.1 Performance metrics

For this MSE, performance metrics are computed yearly or over specific time horizons. More specifically:

- Yearly stats (acronyms encountered in the different figures in parentheses):
  - CV of catch per year (cvC) [Note this is an annual CV rather than interannual variability given for IAV below]
  - Mean catch per year (Cy)
  - Probability that spawner biomass is below Blim (PBlim)
  - CV of SSB per year (cvSB)
  - Mean SSB per year (SBy)
  - Mean of fishing pressure on juveniles (age 0-1) per year (Fjuv.y)
  - Mean of fishing pressure on adults (age 2-6) per year (Fadult.y)
- Metrics over the entire period (2024-2048), short (5 years, 2024-2028), medium (5 years, 2029-2033) and long term (15 years, 2034-2048):
  - ICES Risk3, max probability that spawner biomass is below Blim
  - Average percentage interannual change (or average interannual variability) in catch (IAV), calculated as follows:

    $$\text{IAV} = \text{average over years } (y) \text{ and replicates } (i) \text{ of } \left|\frac{C_{y+1}^{i}}{C_{y}^{i}} - 1\right|$$
  - Mean catch over years
  - Average interannual variability in SSB (see above definition for IAV in catch)
  - Mean SSB over years
  - Mean fishing pressure on juveniles (age 0-1) over years
  - Mean fishing pressure on adults (age 2-6) over years

In relation to the LTMS request, performance criteria in the short, medium and long term are: average SSB, average yield, Indicator for year to year variability in SSB and yield and risk of SSB falling below Blim. Furthermore, the long-term goal requested for the combination of Ftarget-Btrigger control points are: 1) maximising yield, 2) minimising the risk of falling below Blim and 3) achieving stability of catches.

5.2 Fixed F OM projections

Prior to running the MP, simple projections under constant fishing pressure are conducted: 1) F=0 and 2) F=0.2.

The trajectories for F=0 are shown in Figure 5.1. Expectedly, the OMs that generate the highest recruitment due to their stock recruitment relationship are SR2 and SR4. This is because these are fitted on the long time series which leads to higher productivity overall. The M2 scenario (positive additive scaling in natural mortality) also exemplifies high recruitment. These high recruitment scenarios translate in high SSB levels. Important to note is the high degree of variability of the SR4 scenario due to the inclusion of autocorrelation in recruitment deviances.

The annual risk under an F=0.2 constraint is shown in Figure 5.2. The corresponding yield in different time periods is shown in Figure 5.3. Prior to the application of any MP, these projections give first insights into the behaviour under each individual OM. For example, it is clear that the SR4 scenario is prone to higher risk because of large extremes induced by its large variability in recruitment, especially for the BH stock recruitment function which was associated with high autocorrelation. This aspect is exemplified in the stock recruitment pairs Figure 5.4 which displays extremes in SSB and recruitment levels compared to other scenarios.

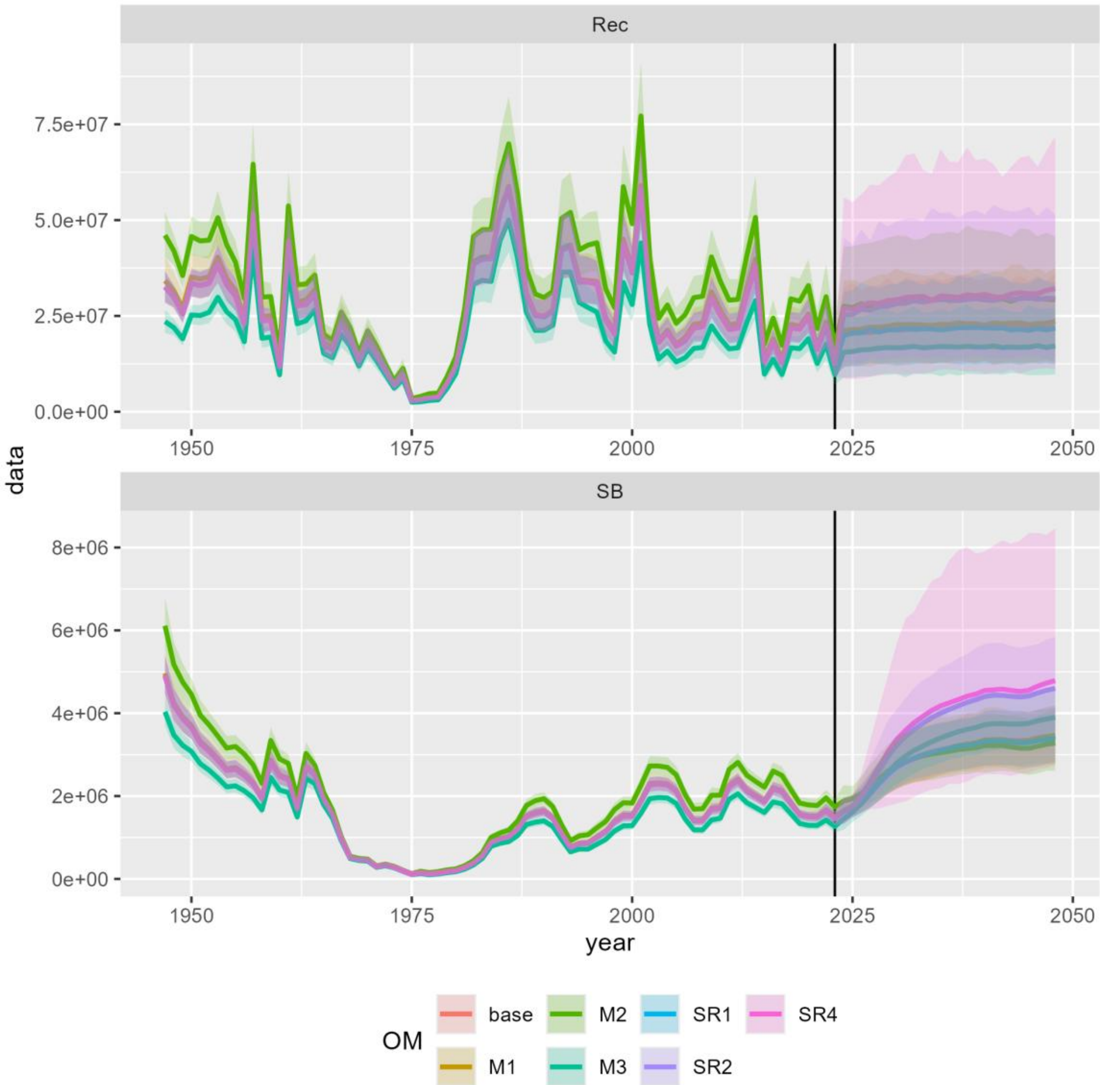


Figure 5.1: Comparison OMs under no fishing pressure. Ribbons are the 95th/5th quantiles.

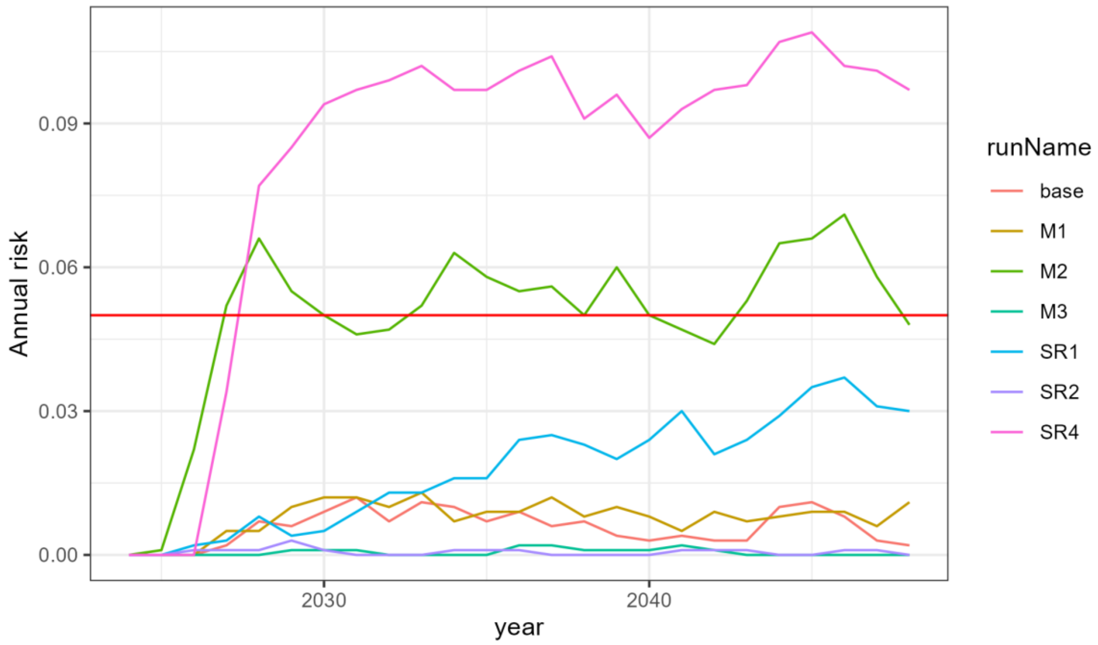


Figure 5.2: Annual risk under F=0.2 constraint for all OMs. The red horizontal line is the 5% ICES Risk3 threshold.

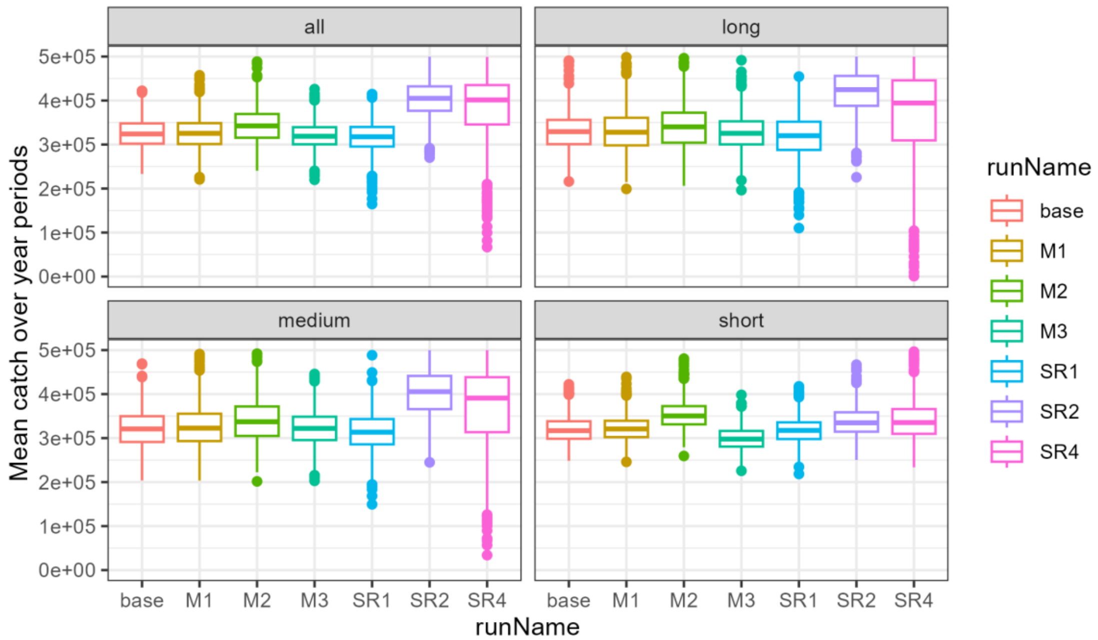


Figure 5.3: Catch at different time horizon under F=0.2 constraint for all OMs.

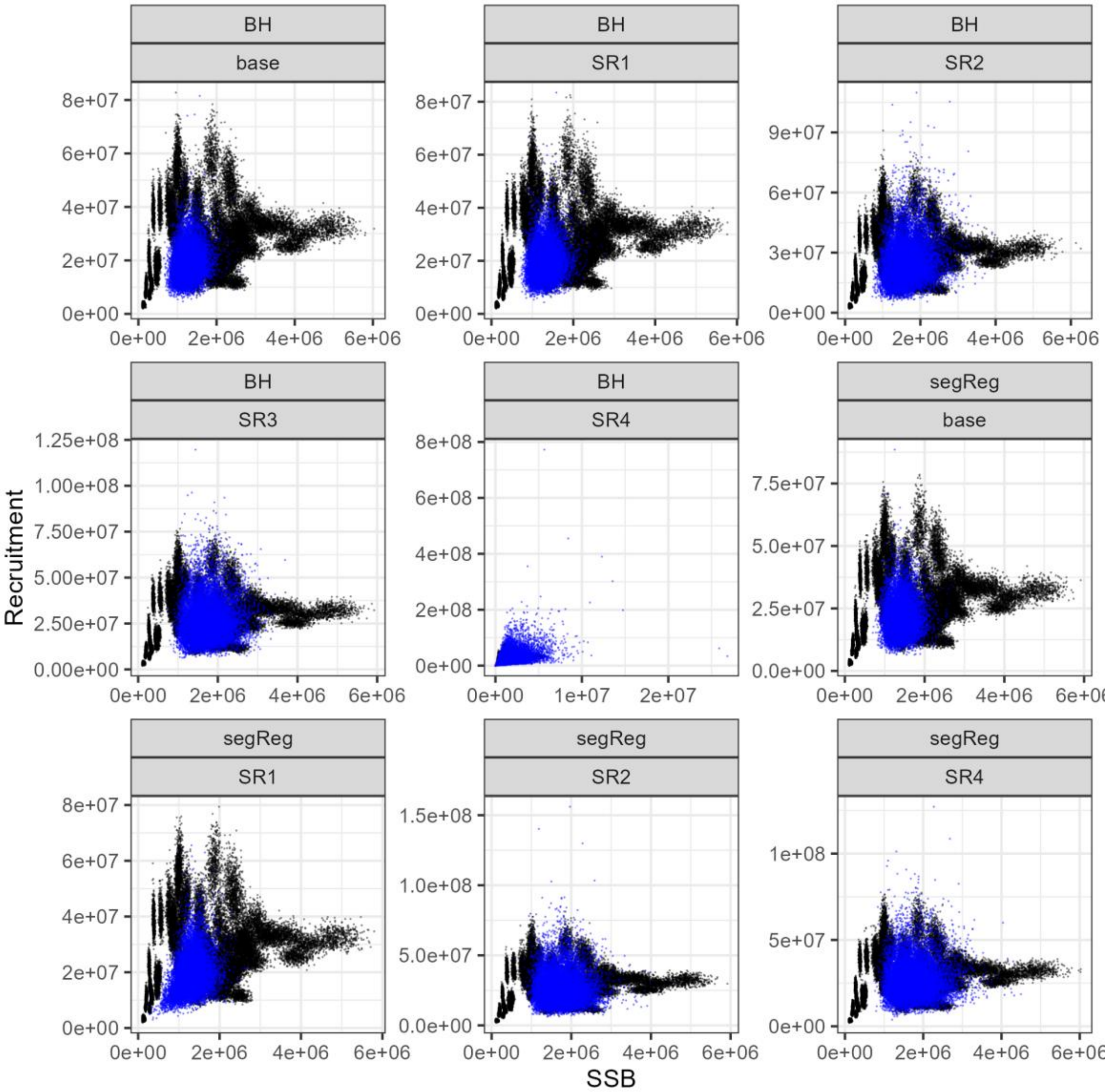


Figure 5.4: Stock recruitment pairs for projections under F=0.2 constraint. Both the historical (black dots) and projection period (blue dots) are plotted for the different types of stock recruitment scenarios. Note the different scales on the y-axis.

5.3 Description of MP building blocks

5.3.1 Decision process

The catch advice for NSAS herring is taking place yearly, with a data lag of 1 year and a management lag of 1 year. The decision process consists of the following steps (Figure 3.12):

1. The assessment model (i.e. the estimator) is run in the intermediate year, with input data up to the data year because of the 1 year data lag.
2. The stock estimations are projected in the intermediate year with a catch constraint using the already known TAC for the intermediate year.

3. The stock is projected in the forecast year with a fishing pressure constraint. This target fishing pressure is calculated using SSB in the forecast year, at spawning time. The HCR determining the fishing pressure is a hockey stick type rule with Ftarget and Btrigger as control points (see Figure 3.16). It is important to note that taking SSB in the forecast year is complicated because fishing pressure in the forecast year is derived from the SSB in the forecast year based on the HCR, which implies that an optimization process is taking place to find the right fishing pressure.
4. TAC is derived from catches corresponding to fishing pressure in the forecast year.
5. A further TAC constraint is applied. More specifically, if the TAC deviates by more than 20% below or 25% above the TAC of the preceding year, the TAC is restricted to this respective limit. The TAC constraint shall not apply if the SSB at spawning time in the year for which the TAC is to be set is less or equal to the Btrigger control point.

The TAC is the quantity implemented by managers. Here, no implementation error is added (i.e. no overshoot). The first year in the projection is conditioned on the catches in 2023 and an already known TAC in 2024, determined at HAWG 2024 (ICES (2024a)).

5.3.2 Stock estimator

It is important to recall that in the decision process, the estimator is run with data up to the data year and catch quota is determined for the forecast year, i.e. 2 years ahead of the data year, as explained in previous sections. Using the results from the estimator, two projections of one year are performed prior to applying the HCR shown in Figure 3.16.

The estimator used here is the SAM stock assessment model, configured at HAWG. The convergence of the model is tracked for each iteration and projected year. Two types of non-convergence are captured:

1. Failing in model convergence leading to no model output
2. Hessian not positive definite, leading to estimates without SD

When case 1. occurred, the stock replicate was dropped entirely. When case 2. was encountered, model estimates were obtained and used. It only prohibited the use of the SDs of recruitment estimates which are used for estimating recruitment in the forecast as a weighted average over 10 years using uncertainty in recruitment as weighting factor. When no uncertainty in recruitment was available, an average with equal weighting was applied.

In addition to the implementation of the estimator with the SAM model, a shortcut was also built. This shortcut replicates the true population estimate and applies a deviance on SSB estimates in the forecast year. This deviance is drawn from a log normal distribution using an sdlog. For example, runs with the shortcut set with sdlog=0 corresponds to a perfect stock assessment in the data year undergoing the same decision process described above. This shortcut was used to troubleshoot the R code and run basic tests (e.g. Figure 5.8). All results presented further for the MP include the stock assessment model as the stock estimator.

5.4 Estimation model performance and convergence

Overall, model convergence was not an issue. Across all OMs, no replicate exemplified a SAM model run that did not converge (convergence case 1, as listed in the previous section, see Figure 5.5 bottom panel). However, for some replicates the hessian was not positive definite, which did not allow the derivation of model uncertainties. This case happened on single years over a marginal amount of replicates (Figure 5.5 top panel). Moreover, the occurrence of this case did not prohibit the computation of subsequent projections.

Using the NSAS herring standard stock assessment model as the estimator is computationally intensive for the MSE model. However, the inclusion is deemed necessary because of the difficulty in characterising the estimation bias of the assessment within the management procedure. The discrepancy between the underlying OM's SSB and the estimation from the estimator is shown in Figure 5.6(a). When taken across replicates, it is clear that the estimator yields accurate estimation of the population. However, there is a contrasting mix of fits across stock replicates. This is something to be expected, because the level of estimation bias of the stock assessment model in the terminal year is a mix of several factors such as the trends in stock trajectory or whether the stock is at low or high levels. This is shown in Figure 5.6(b) for different replicates. In general, the maximum assessment bias was in the order of 20% (Figure 5.6(c)). The inclusion of the stock assessment model in the management procedure compared with perfect knowledge of the stock in the data year generally induces a substantial increase of the risk (Figure 5.7).

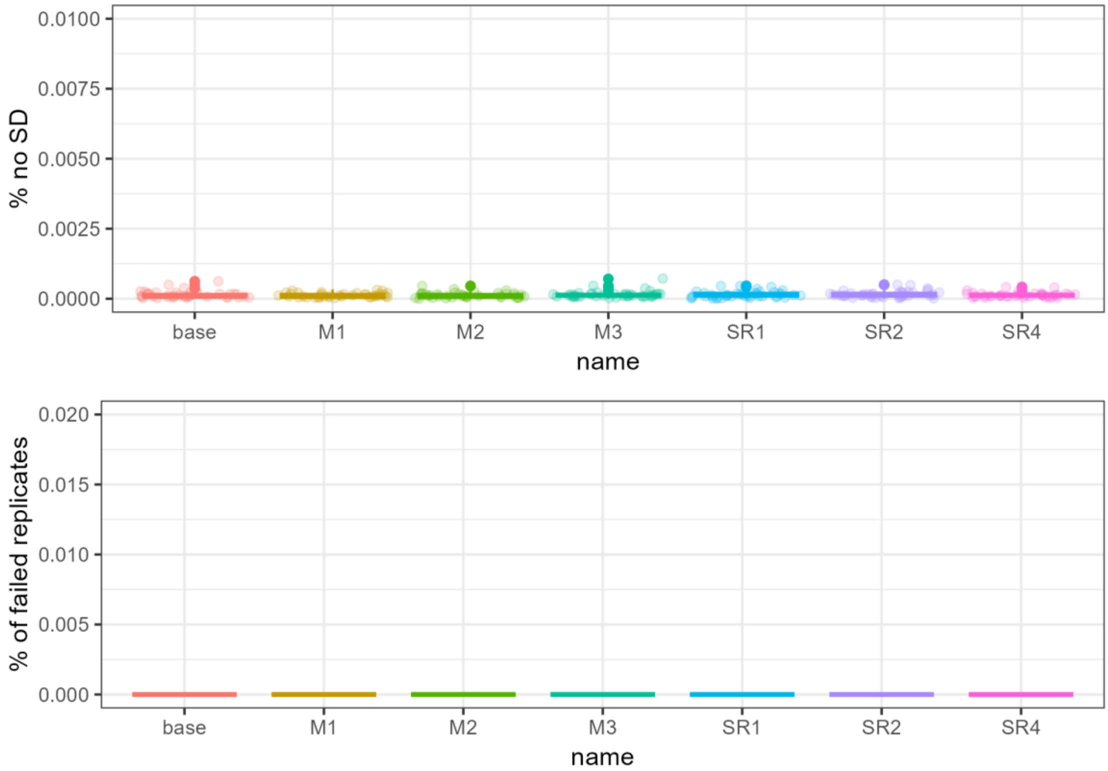


Figure 5.5: Model convergence across OMs and all 41 grid cells covered (see Figure 5.8). The top panel shows the % of individual years across all replicates, OM and grid cells where the Hessian was not positive definite, i.e. no var-covar matrix. The bottom panel shows the % of replicates where the model did not converge.

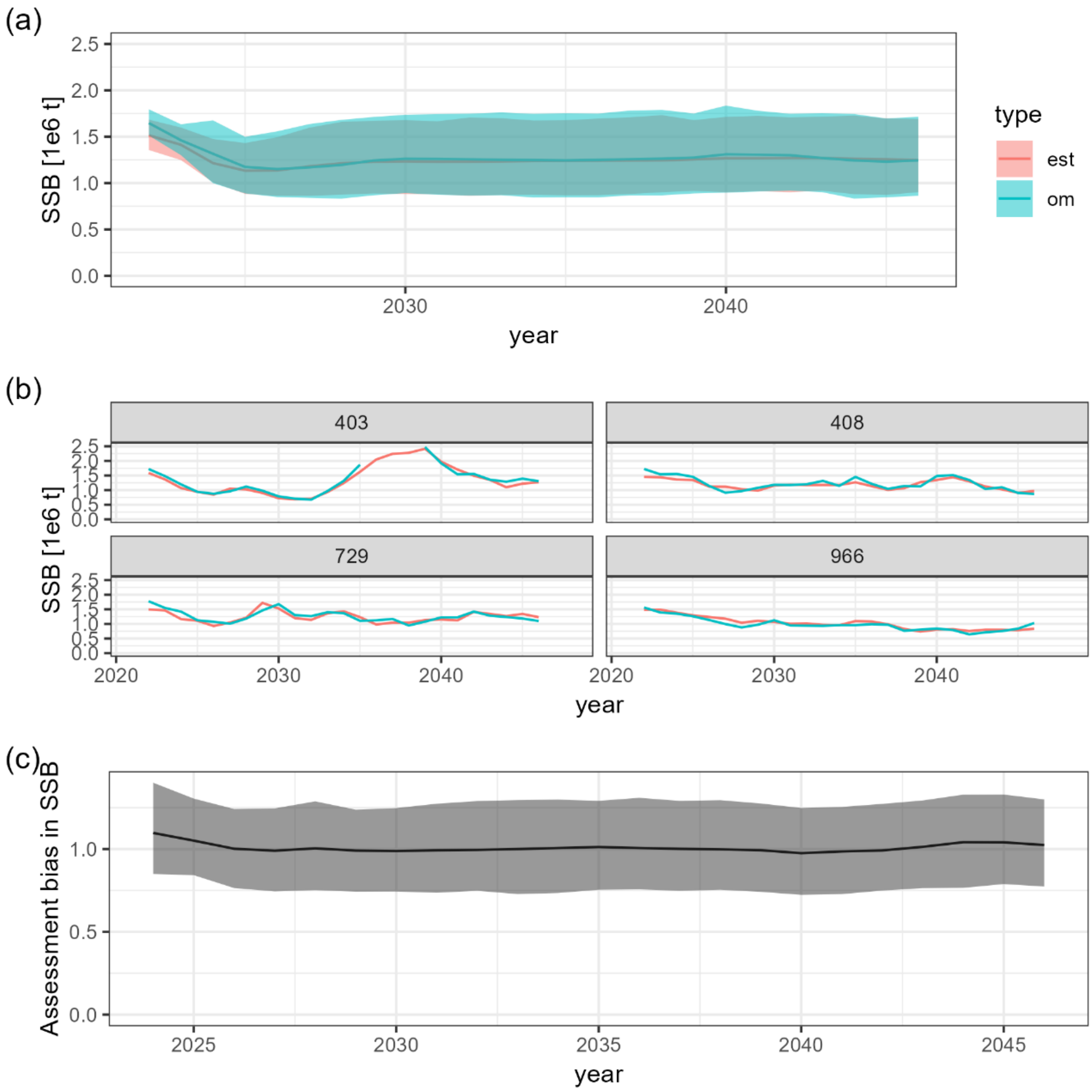


Figure 5.6: Performance of the SAM stock assessment model in SSB. (a) Comparison of SSB from the OM and as estimated by the SAM stock assessment model in projected years. (b) Comparison of SSB from the OM and as estimated by the SAM stock assessment model for four randomly drawn replicates. (c) Bias in assessment estimation, computed as SSB from the estimated divided by the OM SSB. In all panels, the comparison is made for SSB in the terminal year of the projection. The base OM is used for the projections with the following control points: Btrigger=1.7e6 and Ftarget=0.34. Ribbons in (a) and (c) are the 95th/5th quantiles.

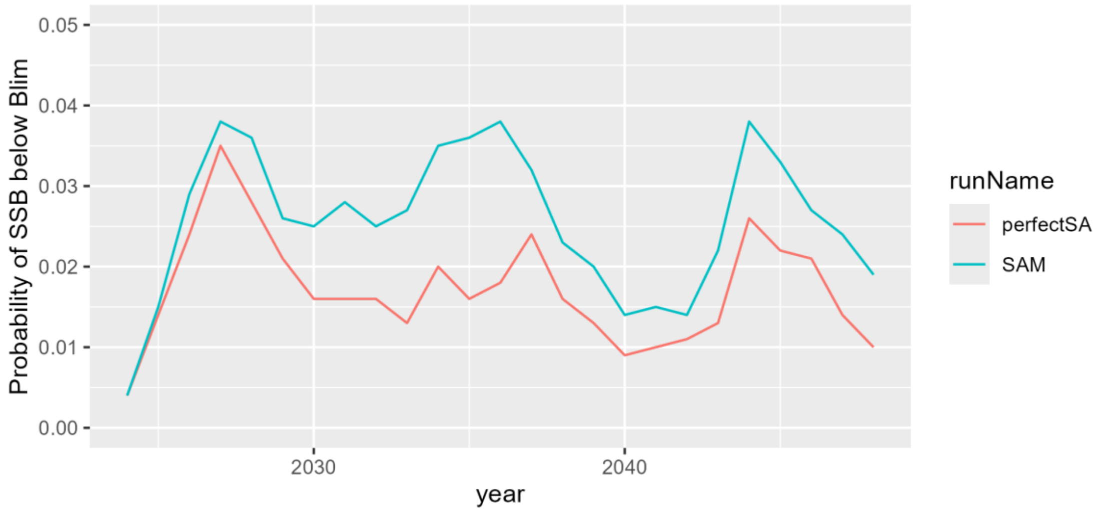


Figure 5.7: Yearly comparison of annual risk between perfect stock estimates and estimates from the stock assessment model in the data year. The control points used are: Btrigger=1.6e6 and Ftarget=0.3.

5.5 HCR tuning and MP performances across OMs

For more detailed results, please see Annex 4.

Tuning process and strategy

Following the MSE request, the tuning of the HCR shown in Figure 3.16 is done over a range of Ftarget (0.18 – 0.39) (with the higest Ftarget being Flim, Table 3.1) and Btrigger ((0.8 – 1.7) × 1e6) control points.

All seven OMs are run for a given grid cell. Performance metrics are calculated for each OM individually. Then, the OMs are combined so there is equal weight between three categories: base, stock recruitment and natural mortality (see Section 3.5). More specifically, the base OM is replicated three times to 3000 replicates and the OMs of the stock recruitment and natural mortality are not replicated, consisting of 1000 replicates per OM. Through combining, a final OM of 9000 replicates is constructed. Performance metrics are computed on that basis.

The tuning for the MSE was computationally intensive because of: 1) the seven different OMs in the reference set, 2) the full-feedback mechanism over a 25-year projection period (i.e. inclusion of the stock assessment model as stock estimator applied each year; Figure 3.15) and 3) the large number of stock replicates (1000). These aspects made a full coverage of all Ftarget-Btrigger control point combinations in the range specified by the request unfeasible. Instead, a strategic set of 41 control points were computed, covering the essential regions of the Ftarget-Btrigger grid. More specifically, the combinations covered focused on: 1) maximizing of the yield together with complying with an ICES risk 3 lower than

5% and 2) resolving the 5% ICES risk 3 boundary. The cells covered are shown in Figure 5.8.

In order to provide results for all the combinations of Ftarget-Btrigger, an interpolation was used. To that aim, the DiceKriging R package (Roustant et al. (2012)) was used to create kriging models with an exponential covariance structure. The performance of this interpolation method was tested through a dedicated bootstrap approach. From the 41 cells covered, a random draw of 8-34 cells was performed. Using these randomly drawn cells, the interpolation was computed. The quality of the interpolation for each performance metric was then calculated using the cells covered but not randomly drawn, taking the mean for each performance metric. The randomization was replicated 100 times per number of randomly drawn cells. The results are shown in Figure 5.9 for ICES risk 3. It can be observed that the results of the interpolation quickly converge towards an accurate estimation (i.e. mean deviation of 1). Other performance metrics not shown here exemplify higher convergence and lower deviations.

Btrigger-Ftarget grid results

The results of the 41 grid cells are presented in Figure 5.10. The coverage of the MSE runs clearly identify the 5% risk boundary. Along the 5% risk boundary, mean long-term catch increases with Btrigger (with a corresponding increase in Ftarget). However, the increase in Btrigger from the lowest to highest value in the grid simultaneously increases interannual variation in catch and decreases SSB. The set of control points that maximizes the yield together with complying to an ICES risk 3 lower than 5% is: Btrigger=1.7e6 coupled with Ftarget=0.34. The stock trajectory for the combined OM is shown in Figure 5.11 whilst the trajectory of each individual OM in the reference set is presented in Annex 3. This set of control points is used further in this document to present results. However, it is important to note that this set of control point does not make effective use of the HCR (Figure 3.16) because of its high trigger point that leads to a management persistently on the slope of the HCR. This aspect, together with increased inter-annual catch variability warrant caution and justifies a combination of Ftarget-Btrigger at a lower Btrigger along the 5% risk boundary.

Fishing pressure on juveniles (in the order of F01=0.06 for the base scenario) is relatively high for Btrigger=1.7e6 coupled with Ftarget=0.34. This is because fishing effort for the B fleet (which is the main fleet targeting juveniles) is scaled together with the A fleet. It is important to note that this scaling is inherent to the modelling and does not reflect any mechanism in management.

Performance metrics at Btrigger=1.7e6-Ftarget=0.34

Performance metrics across OMs are shown in Figure 5.12, exemplifying the contrast between the different OMs. For example, SR2 is the OM with the lowest risk which is a result of its recruitment conditioning that leads to higher recruitment overall (Figure 5.1). In contrast, the SR1 OM has an increased risk due to lower recruitment. The OM with the highest risk is the SR4 OM, also providing the highest SSB. This is because of the large recruitment deviances for this scenario (Figure 4.21). When considering the natural mortality OMs, higher natural mortality (M2) is linked with higher risk and higher SSB and conversely for lower natural mortality (M3). This is a result of SSB levels coupled with Blim estimates (Figure 4.26) which is the reference point risk is dependent on. The effect is further exemplified in Figure 5.13 which shows that SSB relative to Blim is lower at higher natural mortality (M2).

The performance metrics in the long term are presented in Figure 5.14 and follow the trends observed in annual statistics (Figure 5.12). In term of risk, the SR4 OM is a clear outlier due to the deviances in recruitment. When combining OMs in the different categories, there is large spread in catch and SSB largely due to the replicates from the SR4 OM (Figure 5.14(e-f)). The large risk associated with the SR4 OM is compensated by other OMs (Figure 5.14(a)). This high risk is due to the replicates fitted with a Beverton-Holt relationship which exemplified large level of autocorrelation. The ensemble OM approach taken here aims at encapsulating a large range of uncertainties, assuming that the dynamic of the stock in the future lies in this uncertainty and could be represented by any of the OMs. In that context, in order to safeguard against an SR4 type scenario in the future, which would require a more precautionary management, it is important to define indicators (e.g. measuring the level of autocorrelation in recruitment). Such indicators can be derived from the dynamics of the SR4 OM projections. However, within this MSE workshop, this aspect could not be explored in depth but should be considered in upcoming working groups on NSAS herring.

Performance metrics across Btrigger-Ftarget combinations and choice of management strategy

So far, the performance metrics presented integrate results across the seven different OMs of the reference set and their combination (Table 3.2), but only over the cells covered with computation. Using interpolation, the whole range of Ftarget-Btrigger is covered (Figure 5.15). The 5% ICES risk 3 boundaries are further presented in Figure 5.16. OMs that are less prone to risk (e.g. SR2) result in a boundary at high Btrigger-Ftarget combinations and conversely for OMs more prone to risk such as SR1 or SR4. The large contrast between individual OMs is a result of the different assumptions and reflects the level of uncertainty that is being covered in this MSE.

The 5% risk boundary spans a range of expected outcomes in terms of catch, interannual catch variability and spawning biomass (Figure 5.17, Table 5.1-3).

Across this boundary (compare MS1-10 in Table 5.1), a 4% increase in expected catch comes at the cost of an 87% increase in interannual catch variability and an 11% decrease in spawning biomass (and likely a comparable decline in mean catch rates).

Long-term catch increases steadily along the 5% risk boundary Figure 5.18(a), leading to a maximum catch along this boundary located at Btrigger=1.7e6 t which is the edge of the range of Btrigger values considered. However, the increase in catch is relatively small (4%) when compared to other metrics. This is because of the small changes in realised fishing pressure (Figure 5.18(b)). At high Btrigger values, catches are slightly higher (Figure 5.19(a)). However, the discrepancy between the realised fishing pressure and the Ftarget control point is larger as Btrigger increases along the 5% risk boundary. This is because the management is consistently on the slope of the HCR (Figure 3.16), and TAC constraints are suspended below Btrigger. In contrast, the change in long-term inter-annual variability in catch is substantial (87% increase along the 5% risk boundary), as shown in Figure 5.19(b). In short, several management strategies are possible for a similar level of precaution (less than 5% risk) but the stability in long-term catch should be central in the choice of Btrigger-Ftarget. This is because inter-annual catch variability varies substantially (87% along the 5% risk boundary) and offers improvements at low Btrigger, e.g. as opposed to mean catch which exemplifies only a 4% increase along the 5% risk boundary. The different performance metrics are shown in the long (Table 5.1), medium (Table 5.2) and short term (Table 5.3).

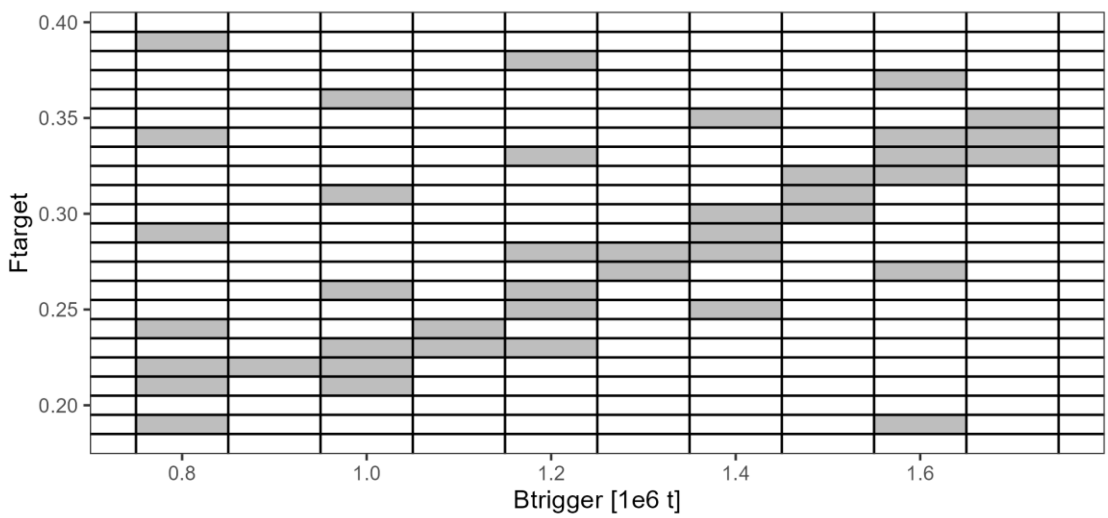


Figure 5.8: Combinations of Ftarget-Btrigger covered. The cells that have been covered are grey shaded.

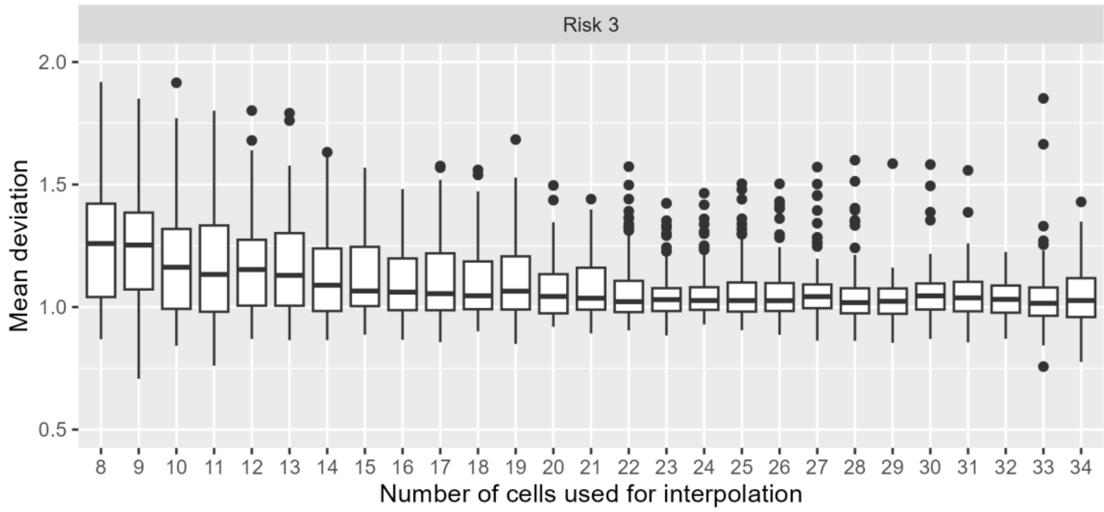


Figure 5.9: Convergence of the kriging models used to interpolate performance metrics across all combinations of Ftarget-Btrigger.

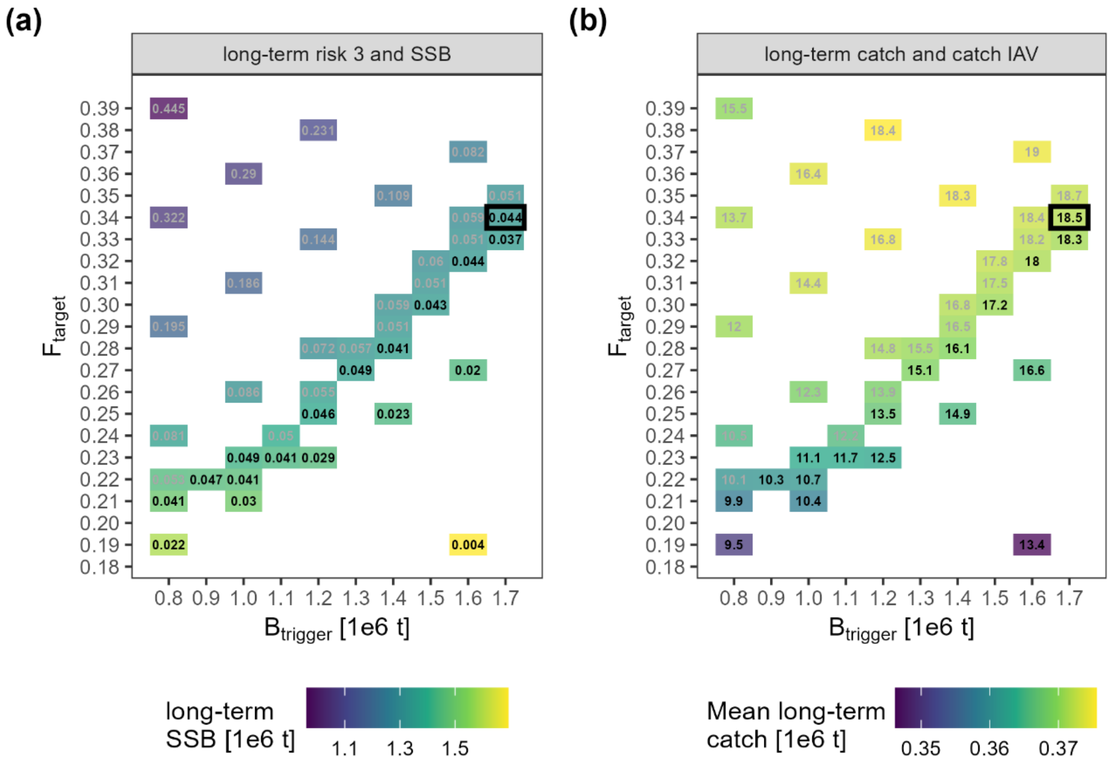


Figure 5.10: Combinations of Ftarget and Btrigger integrated over all operating models. (a) The colours indicate SSB in the long-term, and the values in the cells indicate the associated ICES risk 3. Risk 3 with values greater than 0.05 in grey text. Important to note that risk 3 values are rounded, and if grey are in fact >0.05 unrounded. (b) The colours indicate mean long-term catch, and the values in the cells indicate the associated inter-annual variability in catch, with grey text indicating cases where risk3>0.05. The cell with a thick black border indicates the Ftarget and Btrigger combination that maximises long-term catch while being precautionary (risk3≤0.05).

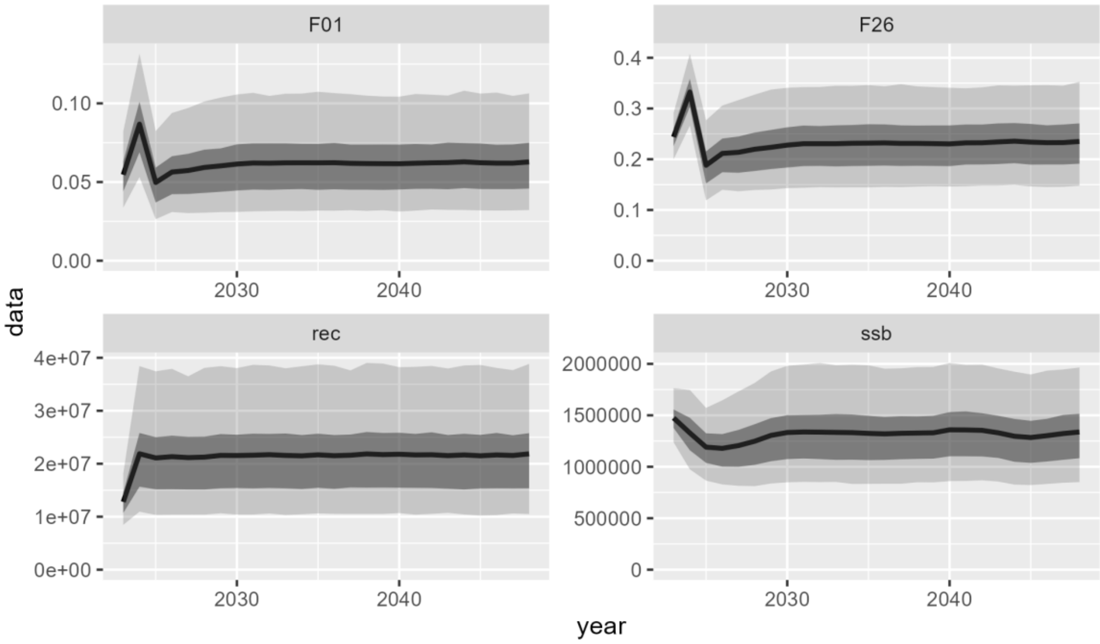


Figure 5.11: Stock trajectory of the combined reference set OM with HCR control points Btrigger=1.7e6-Ftarget=0.34. The trajectories of individual OMs are given in Annex 3. The medians are shown by the black lines and the dark grey shared ribbons are the 75th/25th quantiles and the light grey shaded ribbons are the 95th/5th quantiles.

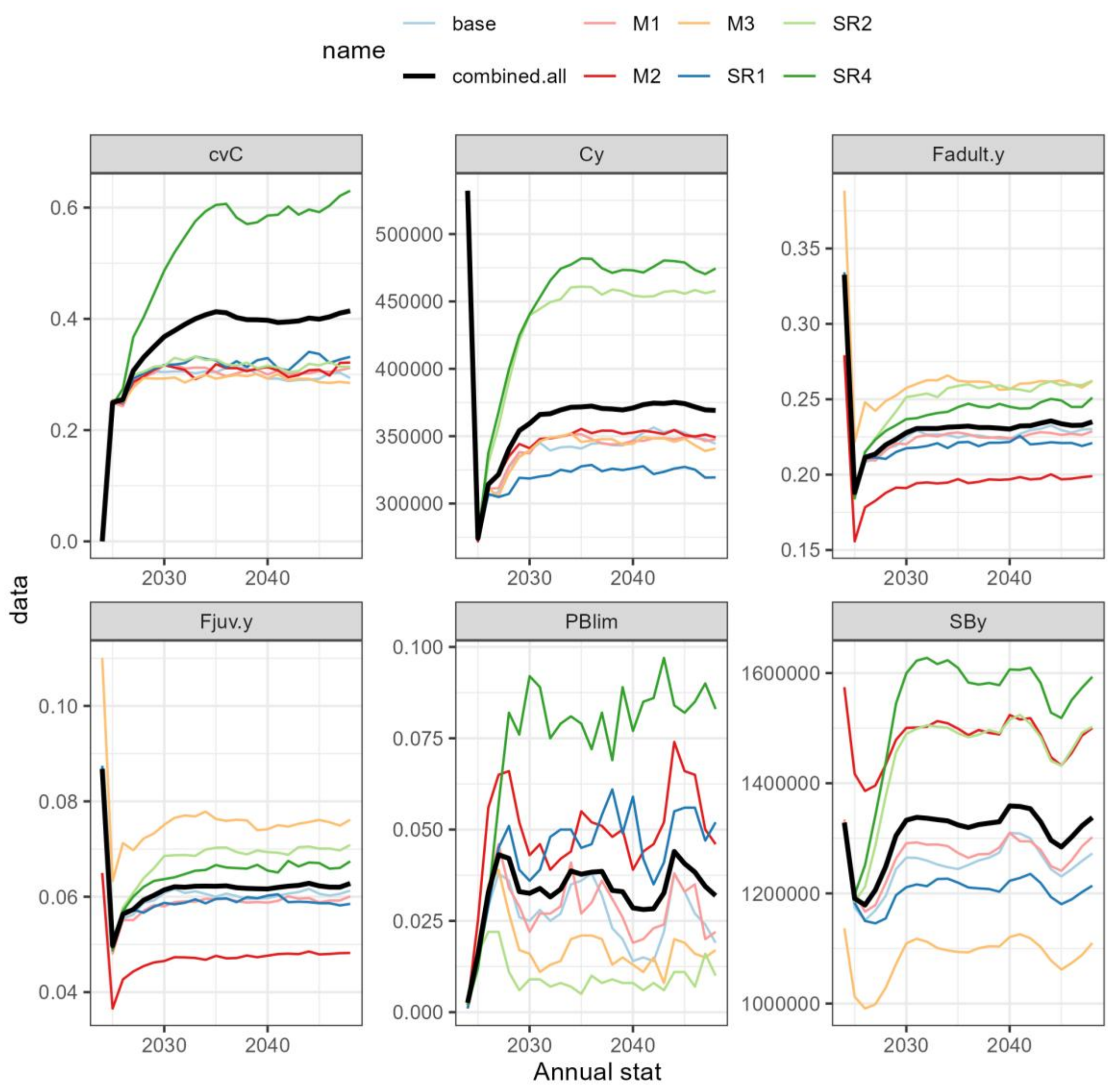


Figure 5.12: Annual performance metrics across individual OMs. Performance metrics presented are CV of catch per year (cvC), mean catch per year (Cy), fishing pressure over ages 2-6 per year (Fadult.y), fishing pressure over ages 0-1 (Fjuv.y), probability that SSB is below Blim (PBlim) and mean SSB per year (SBy). The thick black line is the result of combining OMs. The set of HCR control points used is: Btrigger=1.7e6 and Ftarget=0.34.

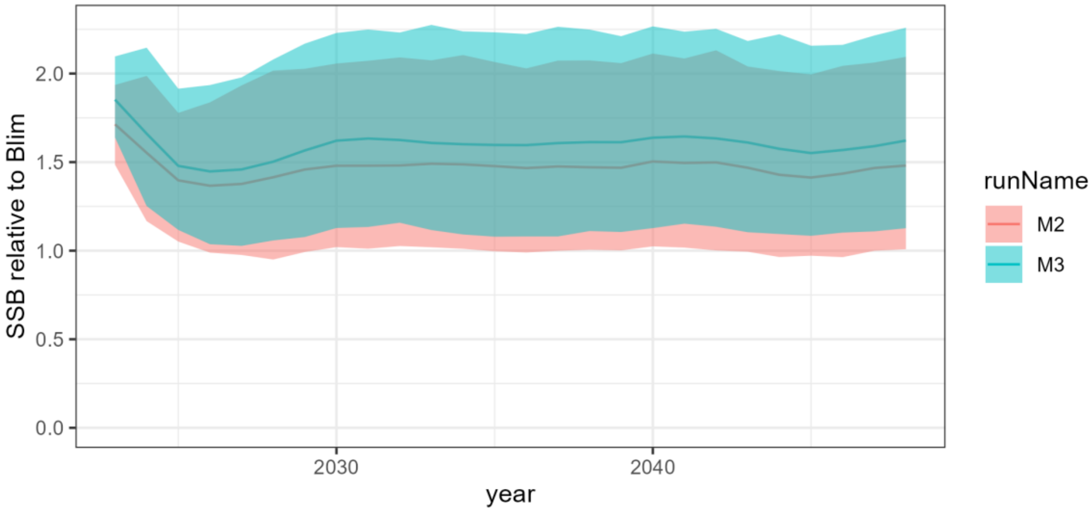


Figure 5.13: SSB level relative to Blim for the M2 and M3 OMs. The set of HCR control points used is: Btrigger=1.7e6 and Ftarget=0.34. Ribbons are the 95th/5th quantiles.

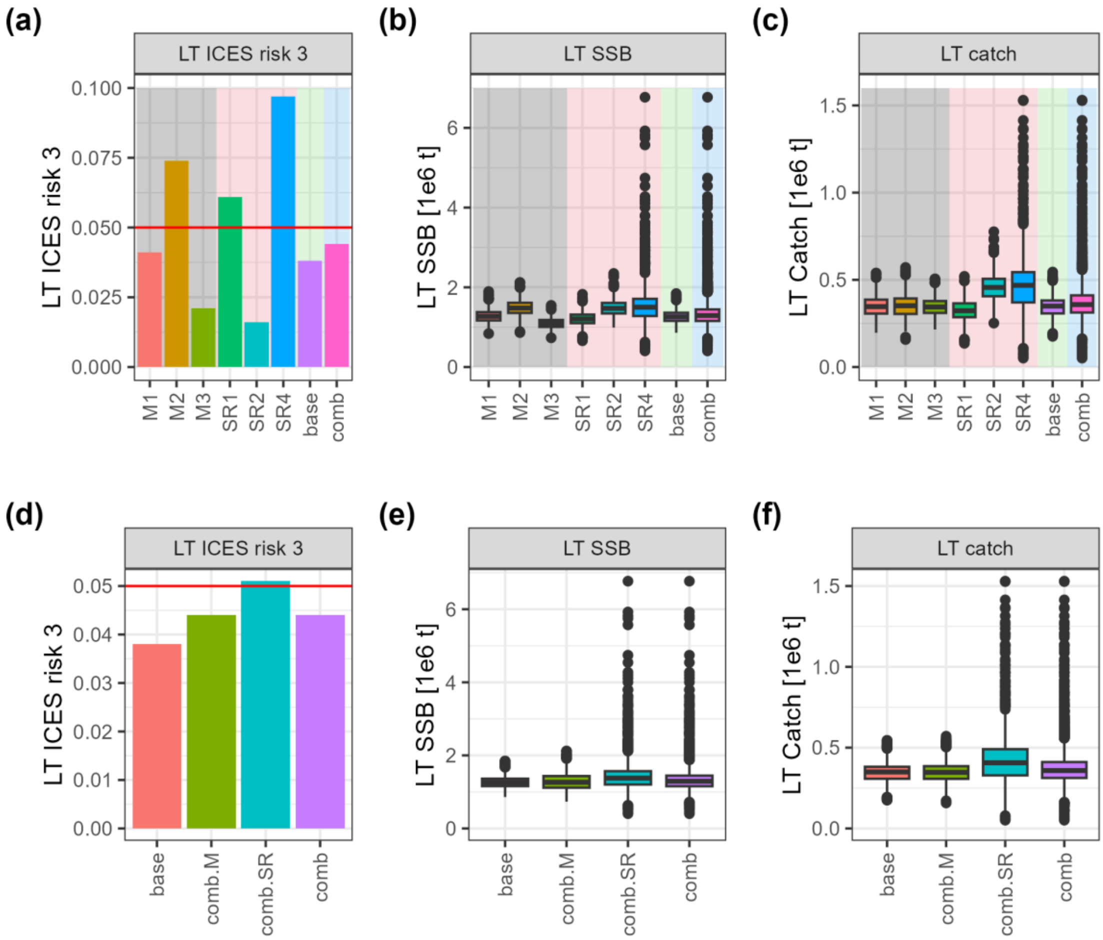


Figure 5.14: Performance metrics in the long term (2034-2048). The performance metrics presented are Long-term ICES risk 3 (a and d), long-term mean SSB (b and e) and long-term mean catch (c and f). The top graphs (a-c) consider OMs in the reference set. The OMs considering uncertainties in natural mortalities (M1, M2, M3) are depicted by the grey shaded area. The OMs considering uncertainties in stock recruitment (SR1, SR2, SR4) are depicted by the red shaded area. The bottom graphs (d-f) are the performance metrics for the different OM categories. The set of HCR control points used is: Btrigger=1.7e6 and Ftarget=0.34.

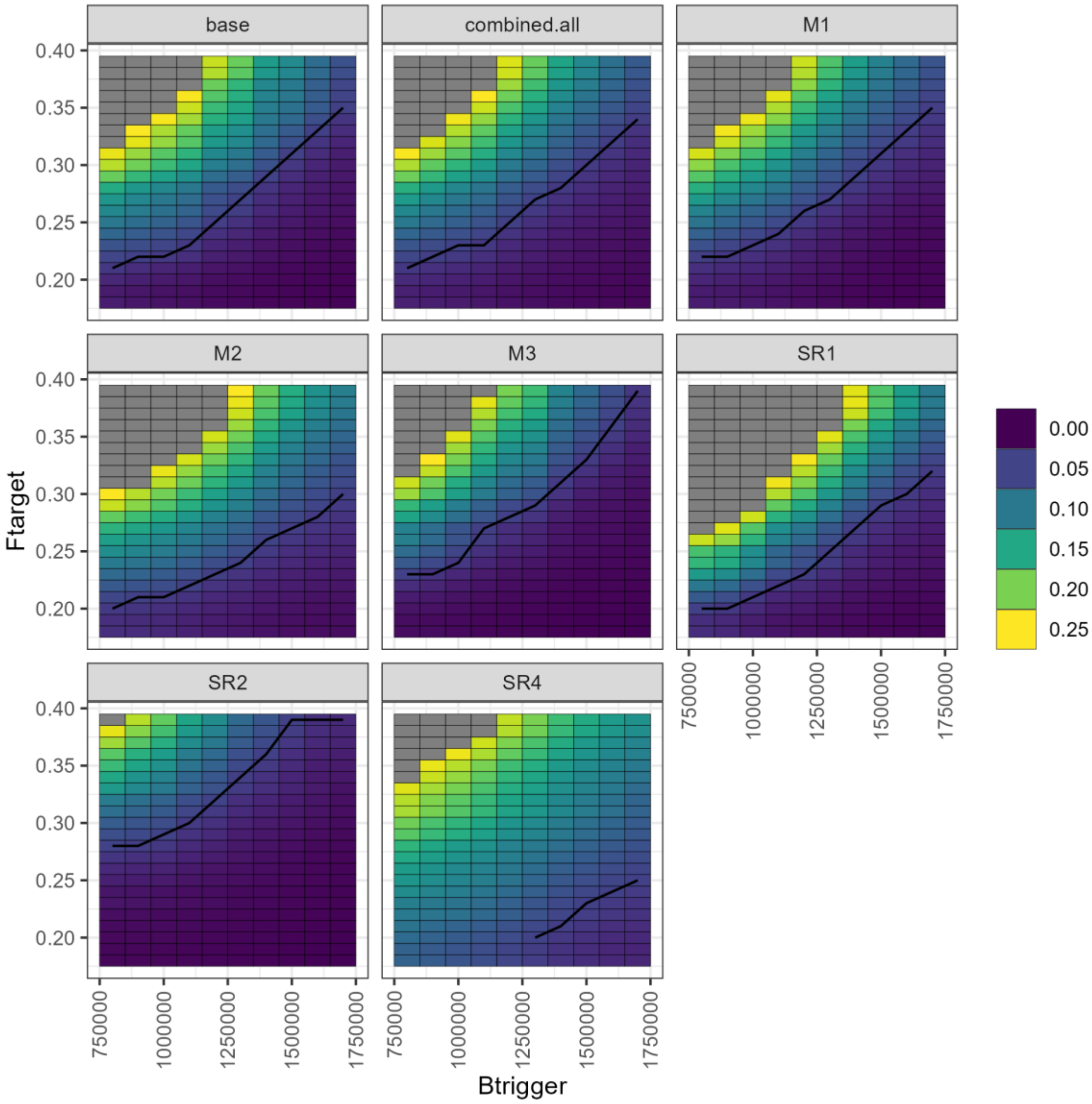


Figure 5.15: ICES long-term risk 3 for all combinations of Btrigger-Ftarget using interpolation. The results are presented for all OMs considered in the reference set and for the final combination of OMs (combined.all). The grey shaded areas corresponds to risk3>0.25. Black lines in each facet is the 5% risk 3 boundary.

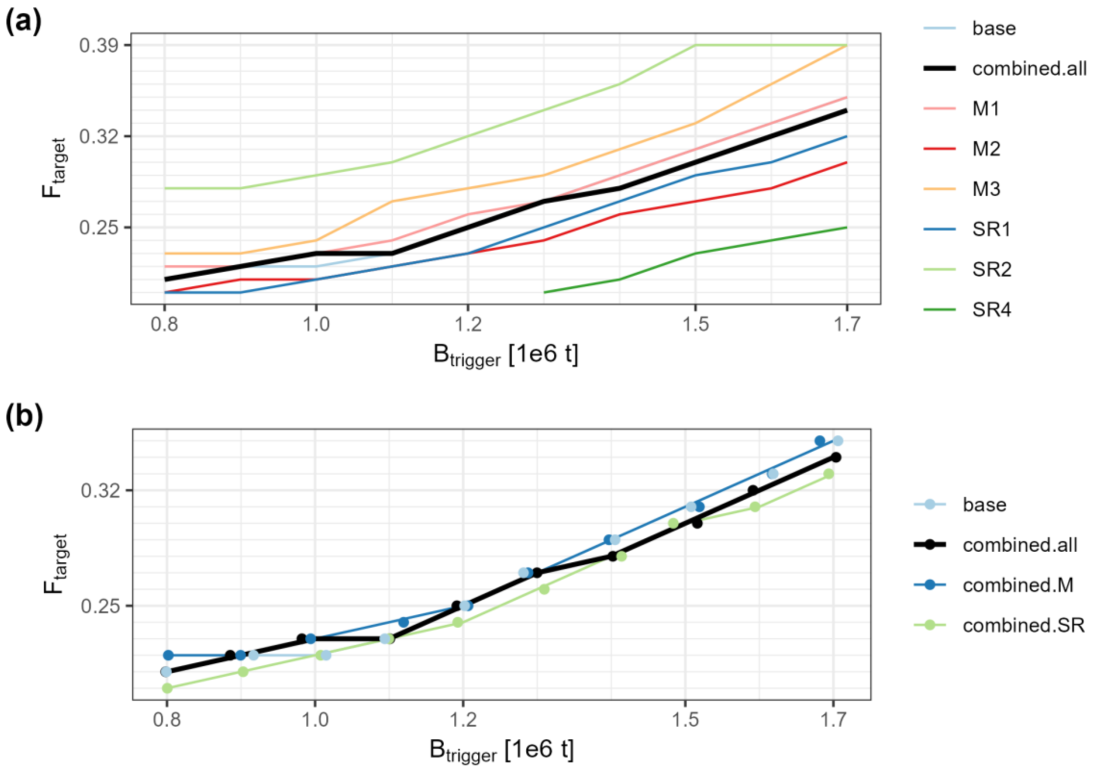


Figure 5.16: 5% ICES long-term risk 3 boundaries using interpolation over all combinations of Btrigger-Ftarget. (a) OMs from the reference set. (b) OM combining categories. Note the jitter to exemplify cases where the 5% risk boundary overlap. The black thick line in both plots is the final combination of OMs.

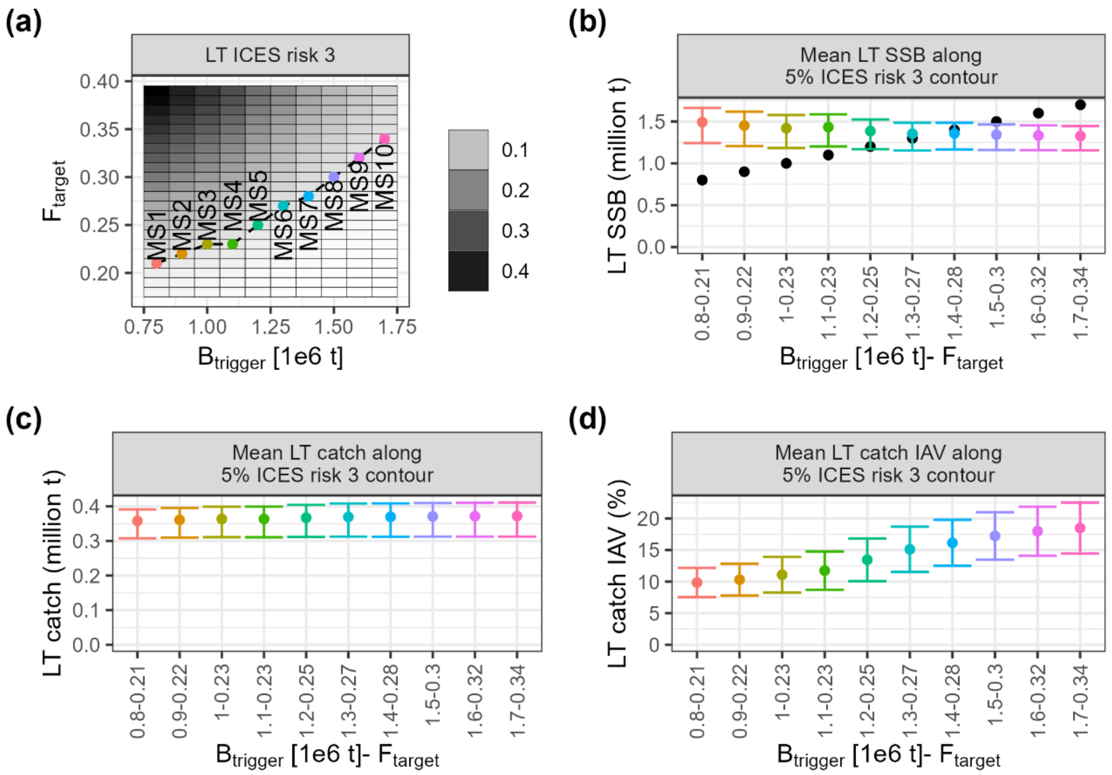


Figure 5.17: Selected performance statistics for the long-term for precautionary cells along the 5% risk 3 boundary of the combined OM for a range of Btrigger values (0.8-1.7 million t) and respective Ftarget (0.21-

0.34). (a) precautionary cells selected. The text associated with each cell (MS1-10) denotes the label for each management strategy. The remaining plots (b-d) indicate long-term performance statistics for these selected cells, integrated over the seven reference set OMs, showing mean with error bars indicating 25-75th quantiles. (b) Long-term SSB, with black dots corresponding to Btrigger levels in each case. (c) Long-term catch. (d) Long-term interannual variability in catch.

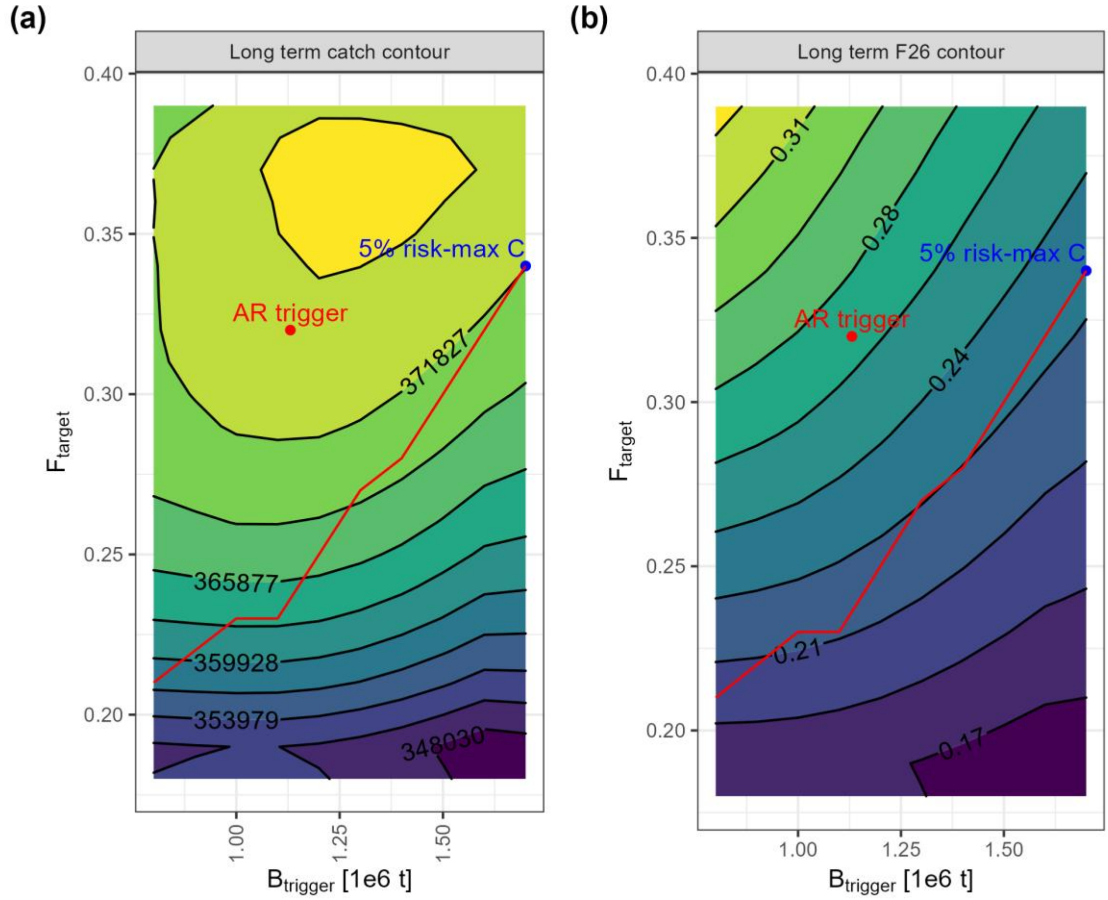


Figure 5.18: Long-term catch and realised fishing pressure contours for the combined OM. (a) Long-term catch. (b) Long-term realised fishing pressure over ages 2-6 (adults). The red line denotes the 5% ICES risk 3 boundary. The red dot is the set of control points for the ICES advice rule currently in place for the management of NSAS herring. The blue dot is the set of control points that maximizes catch whilst complying with an ICES risk 3 less than 5%.

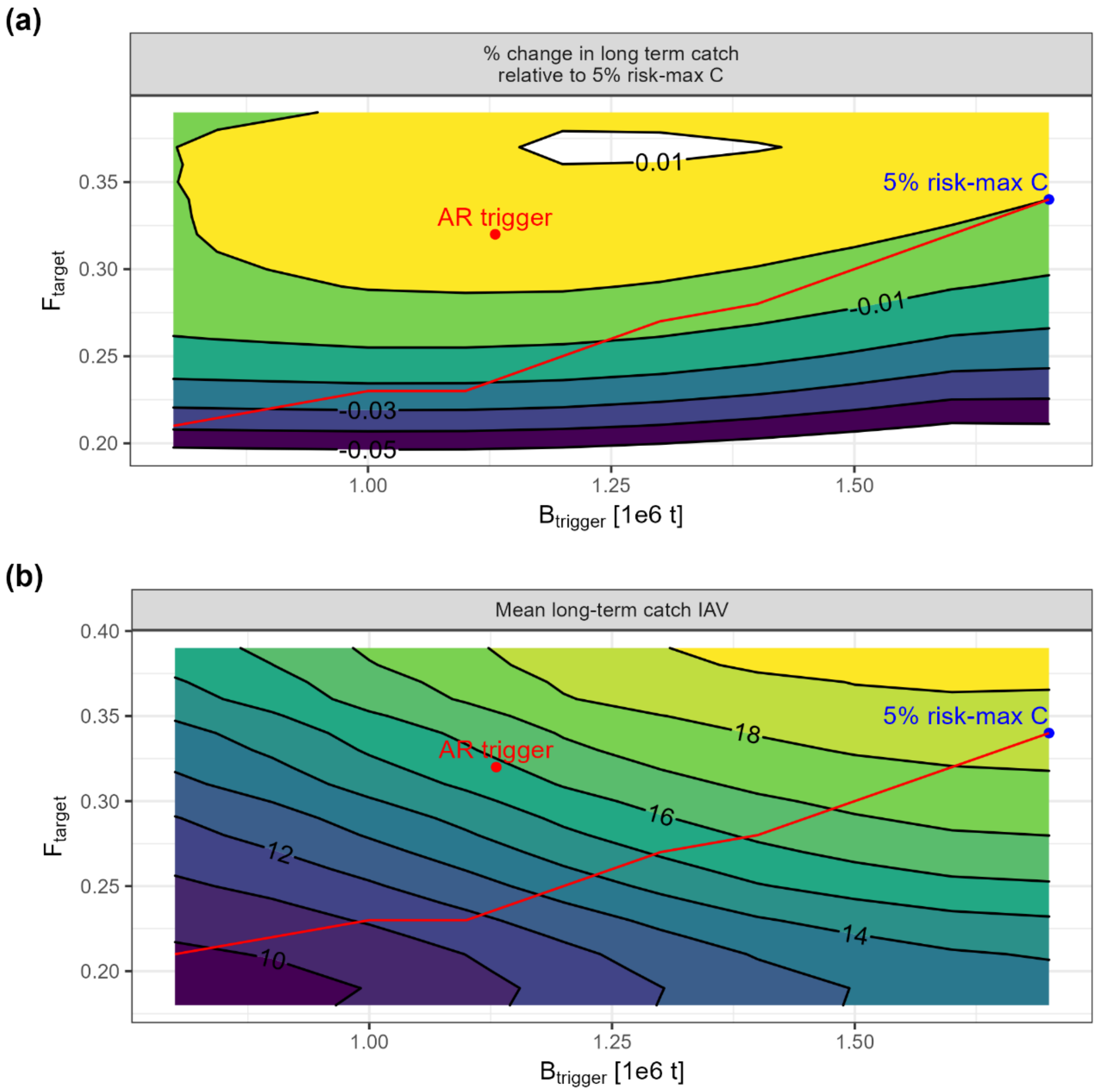


Figure 5.19: Change in long-term catch and long-term catch Inter-Annual Variability (IAV) for the combined OM. (a) % change in long-term catch, taken as relative to the set of control points that maximizes catch whilst complying with an ICES risk 3 of less than 5%. (b) Long-term catch IAV. The red line denotes the 5% ICES risk 3 boundary. The red dot is the set of control points for the ICES advice rule currently in place for the management of NSAS herring. The blue dot is the set of control points that maximizes catch whilst complying with an ICES risk 3 less than 5%.

Table 5.1. Performance statistics in the long-term (years 11-25), integrated over all operating models. The selected management strategies are those precautionary cells along the risk3=0.05 boundary in Figure 5.17. IAV in SSB and catch are given in %.

MS	$B_{trigger}$	F_{target}	risk3	SSB	C	IAV-SSB	IAV-C	F_{2-6}	F_{0-1}
1	800000	0.21	0.041	1492399	357762	10.6	9.9	0.20	0.054
2	900000	0.22	0.047	1451542	360838	10.7	10.3	0.21	0.056
3	1000000	0.23	0.049	1419420	363512	10.8	11.1	0.21	0.057
4	1100000	0.23	0.041	1434036	363505	10.8	11.7	0.21	0.057
5	1200000	0.25	0.046	1385804	367098	11.0	13.5	0.22	0.059
6	1300000	0.27	0.049	1354463	369373	11.1	15.1	0.23	0.061
7	1400000	0.28	0.041	1358071	369702	11.1	16.1	0.23	0.061
8	1500000	0.30	0.043	1343138	370733	11.1	17.2	0.23	0.061
9	1600000	0.32	0.044	1333049	371450	11.1	18.0	0.23	0.062
10	1700000	0.34	0.044	1326810	371888	11.1	18.5	0.23	0.062

Table 5.2. Performance statistics in the medium-term (years 6-10), integrated over all operating models. See caption to Table 5.1 for further details.

MS	$B_{trigger}$	F_{target}	risk3	SSB	C	IAV-SSB	IAV-C	F_{2-6}	F_{0-1}
1	800000	0.21	0.039	1453132	338810	9.1	8.6	0.20	0.053
2	900000	0.22	0.041	1420054	343336	9.2	9.1	0.20	0.055
3	1000000	0.23	0.042	1396541	347451	9.3	10.0	0.21	0.057
4	1100000	0.23	0.033	1413909	347291	9.2	10.6	0.21	0.056
5	1200000	0.25	0.036	1376644	353769	9.4	12.1	0.22	0.059
6	1300000	0.27	0.037	1351763	357854	9.4	13.6	0.22	0.060
7	1400000	0.28	0.033	1357475	358532	9.4	14.4	0.22	0.060
8	1500000	0.30	0.033	1344512	360724	9.4	15.3	0.23	0.061
9	1600000	0.32	0.034	1335235	362201	9.4	15.9	0.23	0.061
10	1700000	0.34	0.034	1329249	363135	9.4	16.3	0.23	0.062

Table 5.3. Performance statistics in the short-term (years 1-5), integrated over all operating models. See caption to Table 5.1 for further details.

MS	$B_{trigger}$	F_{target}	risk3	SSB	C	IAV-SSB	IAV-C	F_{2-6}	F_{0-1}
1	800000	0.21	0.047	1262043	337764	9.4	22.9	0.22	0.058
2	900000	0.22	0.051	1249596	344030	9.5	22.4	0.22	0.060
3	1000000	0.23	0.052	1241355	348055	9.6	22.5	0.23	0.061
4	1100000	0.23	0.042	1250122	343718	9.5	23.8	0.22	0.059
5	1200000	0.25	0.047	1238496	349889	9.6	24.3	0.23	0.061
6	1300000	0.27	0.049	1231377	354084	9.6	25.0	0.23	0.062
7	1400000	0.28	0.043	1236381	352205	9.5	26.1	0.23	0.061
8	1500000	0.30	0.043	1233391	354239	9.5	26.5	0.23	0.061
9	1600000	0.32	0.043	1231612	355672	9.5	26.7	0.23	0.062
10	1700000	0.34	0.043	1230536	356542	9.5	26.8	0.23	0.062

6 Sensitivity and robustness tests

6.1 Sensitivity to exploitation pattern scenarios

All seven operating models in the reference set have been conditioned with a fishing selectivity modelled through a random walk using the last 10 years (2014-2023) and relative fishing effort amongst fleets for the period 2022-2023, given the current reduction in fishing pressure in Division 3.a (through TAC transfer mechanism, see Section 3.2). In addition, NSAS herring experiences fishing pressure on juveniles from bycatch fisheries (e.g. for sprat and Norway pout). The various exploitation pattern scenarios (listed as 1-4 in the request, Annex 1) have been designed to explore sensitivity to these aspects.

6.1.1 SEN1 and SEN2: contrasts in exploitation patterns

SEN1 and SEN2 exploitation pattern scenarios explore periods where either the exploitation patterns themselves or the relative fishing effort among fleets is different compared to the reference set. More specifically, the alternative periods are: 1998-2003 and 2013-2021. The change in exploitation pattern concerns fleet-wise fishing selectivity and fishing effort. For fleet selectivity, it is important to note that in addition to different time periods, the sampling scheme is different. For the reference set period 2022-2023 (i.e. underpinning all OMs in the reference and robustness sets, See Table 3.1), fishing selectivity in projected years is constructed through a random walk process using the last 10 years (2014-2023), keeping consistency with the estimations from the SAM stock assessment model. For the alternative periods (1998-2003 and 2013-2021), fishing selectivity is resampled using the block approach described in Section 4.2, but using separate blocks from the biological variables.

Figure 6.1 illustrates the difference in fishing effort and fishing selectivity among the different periods. The annual performance metrics are shown in Figure 6.2 and performances in the long term in Figure 6.3. These plots illustrate the sensitivity for the base OM only. The reference set conditions lead to higher risk and more variable fishing mortality on ages 0-1, but generally the remaining performance statistics are similar in both level and spread. This provides assurance that assuming current fishing exploitation conditions is reasonable and assuming other time periods for fishing exploitation leads to lower risk (i.e. assuming current fishing exploitation conditions is more precautionary).

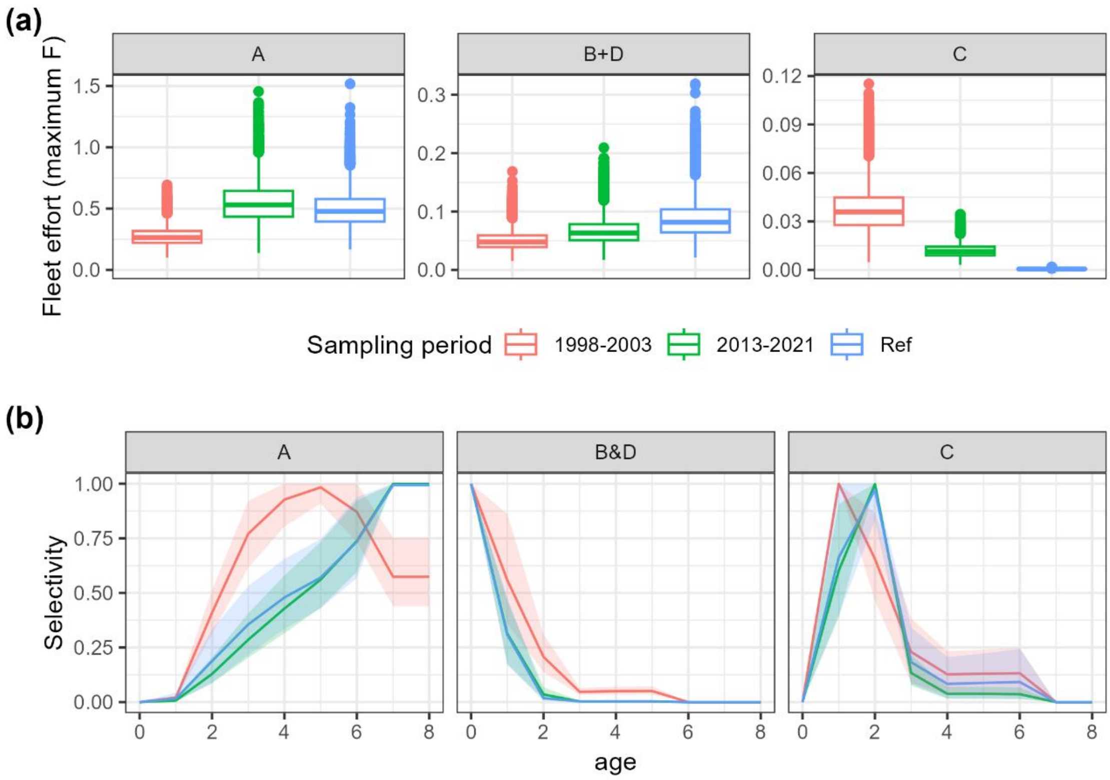


Figure 6.1: Assumptions underpinning the different exploitation pattern scenarios. For these scenarios, both (a) fleet specific fishing effort and (b) fishing selectivity varied. For the base case and other OMs in the reference set (labelled as "Ref"), fishing effort was sampled from 2022-2023 and fishing selectivity was modelled using a random walk based on the period 2014-2023. Alternative exploitation pattern scenarios sampled both fishing effort and selectivity over time periods 2013-2021 and 1998-2003. Fishing effort per fleet was taken as yearly maximum fishing pressure (apical F).

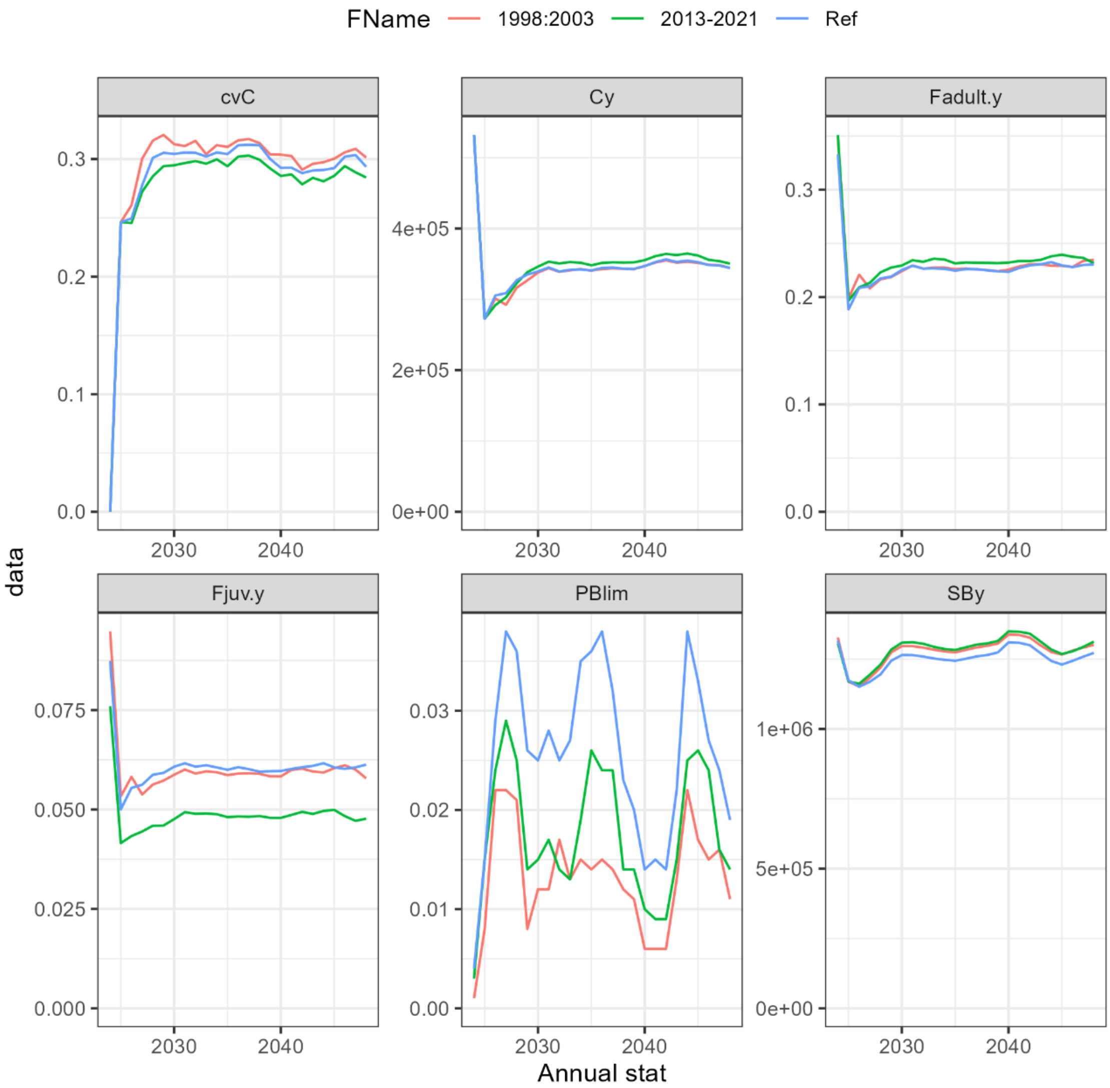


Figure 6.2: Annual stats over grid cell Btrigger=1.7e6-Ftarget=0.34 for the base OM and the base OM with exploitation patterns resampled from the 2013-2021 and 1998-2003 time periods. The yearly statistics are: CV in catch (cvC), catch (Cy), fishing pressure on adults (over age 2-6; Fadult.y), fishing pressure on juveniles (over age 0-1; Fjuv.y), probability of falling below Blim (PBlim) and SSB (SBy).

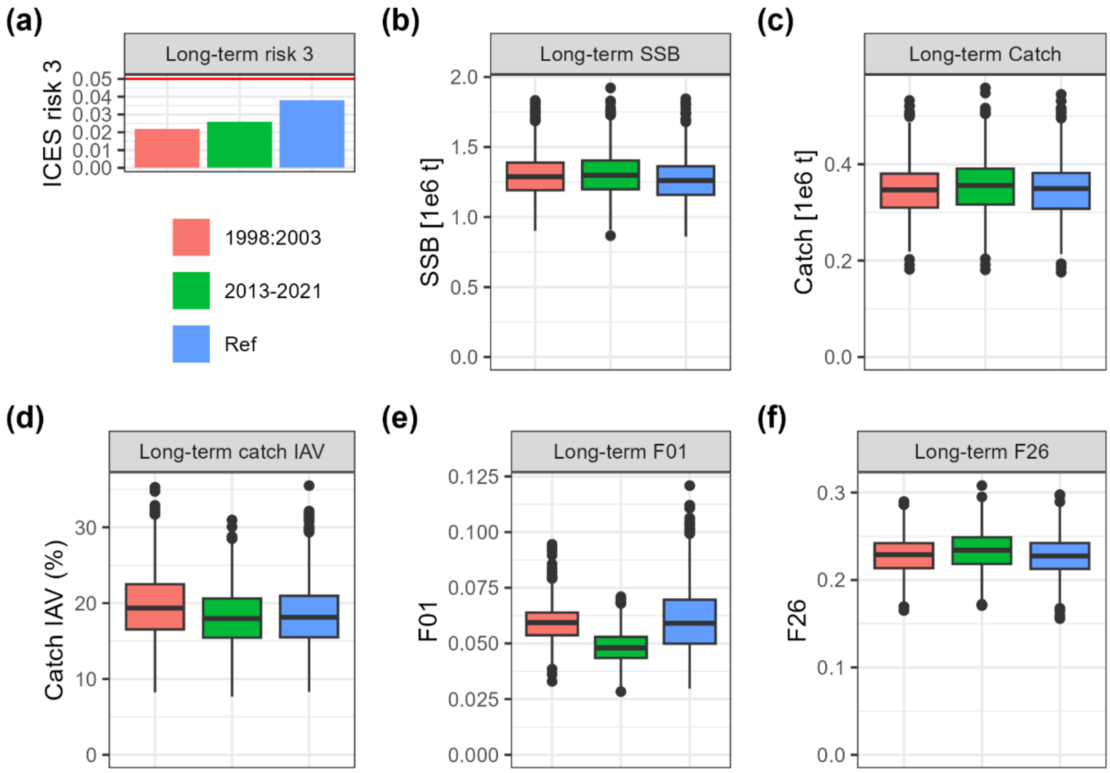


Figure 6.3: Comparison of long-term performance metrics for the three exploitation patterns considered: base case and other OMs in the reference set (labelled as "Ref"), 2013-2021 and 1998-2003. (a) Long-term ICES risk 3. (b) Long-term mean SSB. (c) Long-term mean catch. (d) Long-term mean catch Inter-Annual Variability (IAV). (e) Long-term juvenile fishing pressure (over ages 0-1). (f) Long-term fishing pressure on adults (over ages 2-6). The comparisons are for the baseline operating model only, and for Btrigger=1.7 million tonnes, and Ftarget=0.34.

6.1.2 SEN3: varying levels of juvenile fishing pressure

In the reference set OMs, fishing pressure on juveniles is driven by fishing effort of fleet B in the period 2022-2023. This aspect is a modelling assumption, drawing from historical data. This is because under the management procedure requested here, there is no mechanism to control fishing pressure on juveniles. Therefore, juvenile fishing pressure for the OMs in the reference set should not be interpreted as predictions. In that context, the SEN3 exploitation pattern scenario explores the impact of varying levels of fishing mortality on ages 0-1, more specifically spanning a range from zero to 0.1 in 0.025 steps. Results are presented in Figure 6.4 and 6.5 for annual and long-term performance metrics respectively. Increasing fishing mortality on ages 0-1 has a clear negative impact on risk3, SSB and catch. The calculations given here illustrate the impact of different levels of juvenile fishing pressure to inform potential management mechanisms.

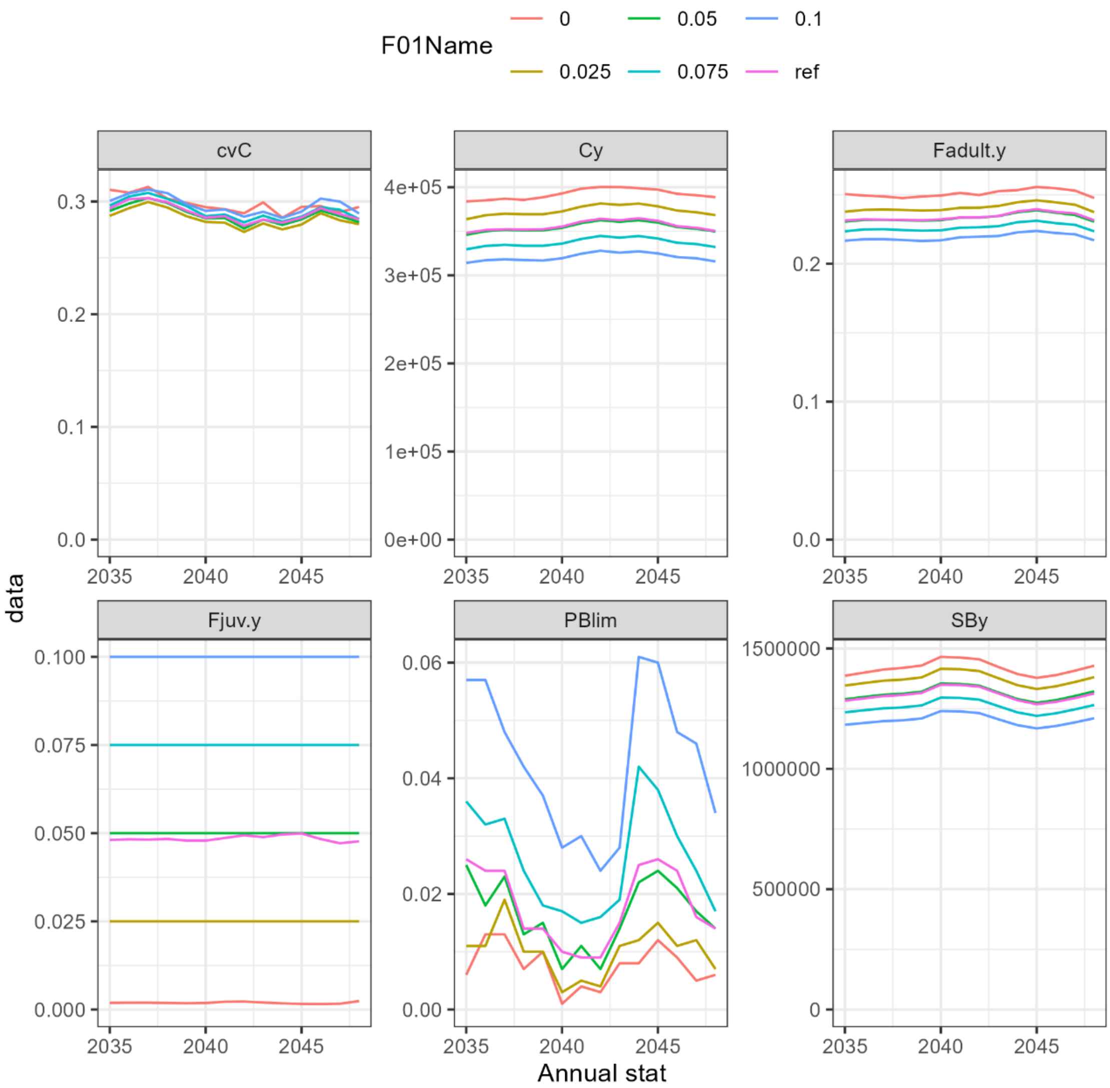


Figure 6.4: Annual stats over grid cell Btrigger=1.7e6-Ftarget=0.34 for the base OM (ref) and the base OM with juvenile fishing pressure F01=0,0.025,0.05,0.075,0.1. The yearly statistics are: CV in catch (cvC), catch (Cy), fishing pressure on adults (over age 2-6; Fadult.y), fishing pressure on juveniles (over age 0-1; Fjuv.y), probability of falling below Blim (PBlim) and SSB (SBy).

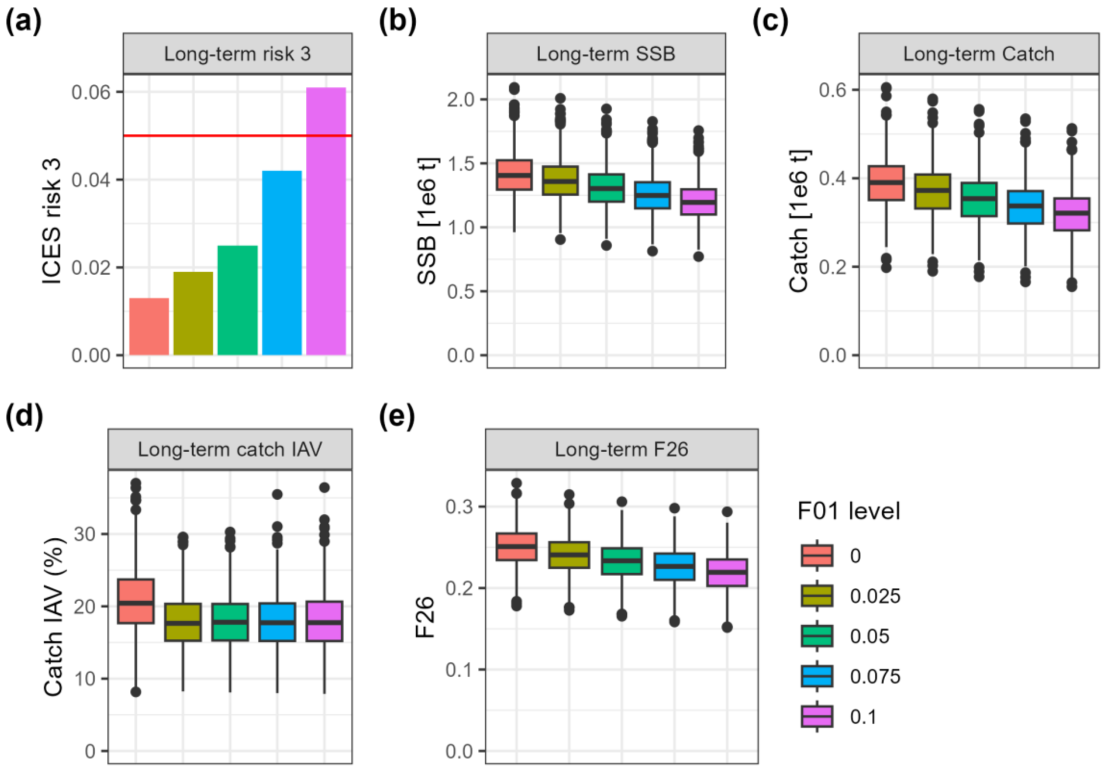


Figure 6.5: Long-term performance metrics over grid cell Btrigger=1.7e6-Ftarget=0.34 for the base OM with varying juvenile fishing pressure F01=0-0.1 in 0.025 steps. (a): Long-term ICES Risk 3. (b): Long term SSB. (c): Long-term catch. (d): Long-term IAV in catch. (e): Long term fishing pressure on adults. Note that the higher catch IAV for F01=0 is because the F01=0 "target" could not be achieved in that case due to the small amount of 0-1 catch that would continue in the A-fleet.

6.1.3 SEN4: shifting of selectivity at age

The SEN4 exploitation pattern scenario explores the impact of shifting the exploitation pattern on the A-fleet by one age up (SEN4.1) or down (SEN4.2), while setting exploitation on the remaining fleets (B-D) to zero. This is exemplified in Figure 6.6(a). The impact is generally negative on all performance statistics when moving the selection pattern down because of the increased fishing pressure on juveniles.

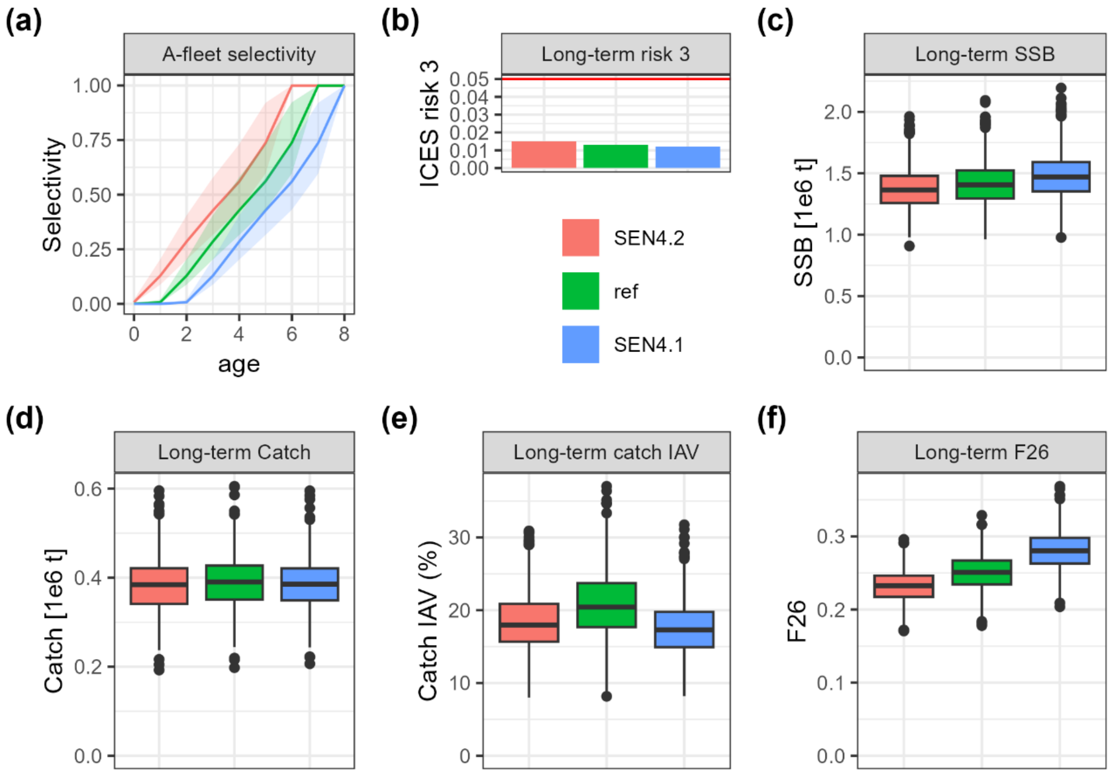


Figure 6.6: Effect of the shifting of the fishing selectivity of the A fleet. The reference case (green colour) considers a fishing selectivity resampled over the period 2013-2021. Case SEN4.1 (blue colour) corresponds to a fishing selectivity shifted by 1 age towards older ages. Case SEN4.2 (red colour) corresponds to a fishing selectivity shifted by 1 age towards younger ages. For all cases, the effort of fleets B-D are set to 0. (a) A-fleet fishing selectivity for the three cases considered. (b) Long-term ICES risk 3. (c) Long-term mean SSB. (d) Long-term mean catch. (e) Long-term mean catch Inter-Annual Variability (IAV). (f) Long-term mean fishing pressure on adults (over ages 2-6). For this comparison, Btrigger=1.7 million tonnes, and Ftarget=0.34.

6.2 Sensitivity to TAC constraints and banking and borrowing

All results so far include TAC constraints and exclude banking and borrowing, as described in Section 5.3. The sensitivity tests in this section consider the impact of excluding TAC constraints and including a banking and borrowing scheme.

6.2.1 SEN5: TAC constraint

TAC constraints are applied only if the SSB (measured at spawning time during the year that the TAC applies) is above Btrigger. If this condition is met, the TAC constraints ensure that a TAC cannot deviate by more than 20% below or 25% above the TAC of the preceding year. Figure 6.7 illustrates the impact of removing TAC constraints for a range of Btrigger-Ftarget combinations. The impact appears to be negligible on risk and small on interannual catch variation, with a slight increase for lower Btrigger values (due to TAC constraints being suspended less frequently in the case where they are included).

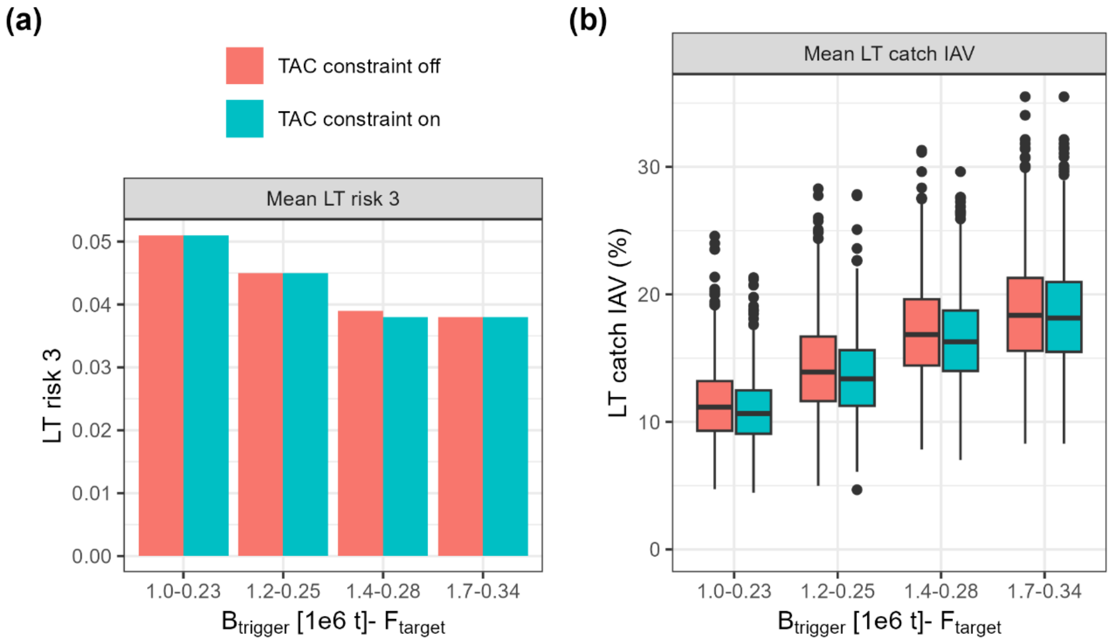


Figure 6.7: Effect of TAC constraints on (a) long-term ICES risk 3 and (b) mean catch Inter-Annual Variability (IAV). A set of four different combination of Btrigger-Ftarget are shown (MS=3, 5, 7, 10 from Table 5.1). The comparisons are for the base OM only.

6.2.2 SEN6: banking and borrowing

The implementation of the banking and borrowing scheme (BB) follows what was implemented during the WKNSMSE workshop (ICES (2019)). It is not possible to model exactly the behaviour of the banking and borrowing scheme, because one does not know to what extent it will be used each year, or what underlies the decision to bank/borrow each year. The approach taken here is to model an extreme version of banking and borrowing which consists in alternately banking and borrowing. Moreover, banking and borrowing is applied after the application of TAC constraints, and is modelled as implementation error (i.e. banking and borrowing does not affect the TAC from year to year, but rather the catch that is associated with the TAC). In addition, banking and borrowing only applies if the SSB (measured at spawning time during the year that the TAC applies) is above Btrigger.

The level of catch that includes banking and borrowing is calculated as follows:

$$C_y = \mathrm{TAC}_y \times (1 + q_y) - q_{y-1}\mathrm{TAC}_{y-1}$$

with the TAC resulting from the application of the HCR (Figure 3.16) and C_y the realised catch in year y that includes the application of banking and borrowing. The variable q_y alternates as: $q_y = -0.1$ for odd years (i.e. $y = 1,3,5,\ldots$) and $q_y = 0.1$ for even years (i.e. $y = 2,4,6,\ldots$). The first year of the scheme ($y = 1$) is 2023 which is the first year of projection (for this year $q_0 = 0$ in the above equation). This alternating scheme is further exemplified in Table 6.1.

The impact of the banking and borrowing scheme is shown in Figure 6.8 for a range of Btrigger-Ftarget combinations. It can be observed that the impact of this extreme version of the scheme on risk 3 is small. The inter annual variability in catch is strongly increasing at low Btrigger-values. Conversely, for higher Btrigger-Ftarget combinations the BB scheme is applied to a much lower degree, since the stock is predominantly below Btrigger (Figure 5.17b), where BB does not apply. However, it should be noted that the extreme banking and borrowing scheme tested is unlikely to be realistic as it deliberately induces fluctuations in catch, which is probably the opposite of how the scheme actually operates (with users of the scheme likely wanting to achieve year-to-year stability in their operations). This outcome is consistent with analyses of banking and borrowing tested elsewhere (De Oliveira (2013), ICES (2019)).

Table 6.1: Realisations of the banking and borrowing scheme tested. In the examples shown, TAC_y represents the TAC from the decision model in year y (and following implementation of any TAC constraints that are applicable for that year). The BB scheme presented here is the same as the one applied for Cod, haddock, whiting, and saithe at WKNSMSE (ICES 2019).

year	y1	y2	y3	y4	y5	y6
Realised catch	$0.9 \times \mathrm{TAC}_{y1}$	$1.1 \times \mathrm{TAC}_{y2} + 0.1{\times}\mathrm{TAC}_{y1}$	$0.9 \times \mathrm{TAC}_{y3} - 0.1{\times}\mathrm{TAC}_{y2}$	$1.1 \times \mathrm{TAC}_{y4} + 0.1{\times}\mathrm{TAC}_{y3}$	$0.9 \times \mathrm{TAC}_{y5} - 0.1{\times}\mathrm{TAC}_{y4}$	etc.

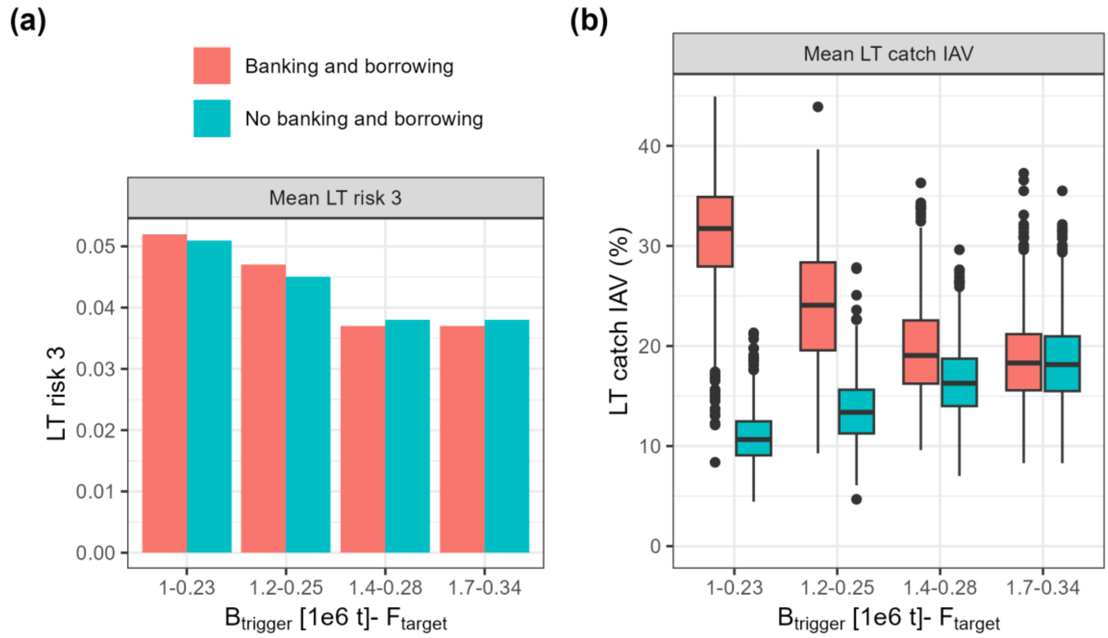


Figure 6.8: Effect of banking and borrowing on (a) long-term ICES risk 3 and (b) mean catch Inter-Annual Variability (IAV). A set of four different combination of Btrigger-Ftarget are shown (MS=3, 5, 7, 10 from Table 5.1). The comparisons are for the base OM only.

6.3 ROB.M: robustness tests on natural mortality

As described in Section 4.5.2, lower and higher additive scaling in natural mortality are considered as robustness tests (M4 and M5 in Figure 4.19). These tests are carried out using Btrigger=1.7e6-Ftarget=0.34 as control points. The annual performance metrics for different natural mortality assumptions are presented in Figure 6.9. As previously observed, higher natural mortality (M2 and M5) is linked with higher risk and higher SSB and conversely for lower natural mortality (M3 and M4). This is a result of SSB levels coupled with Blim estimates (Figure 4.26 which is the reference point risk is dependent on.

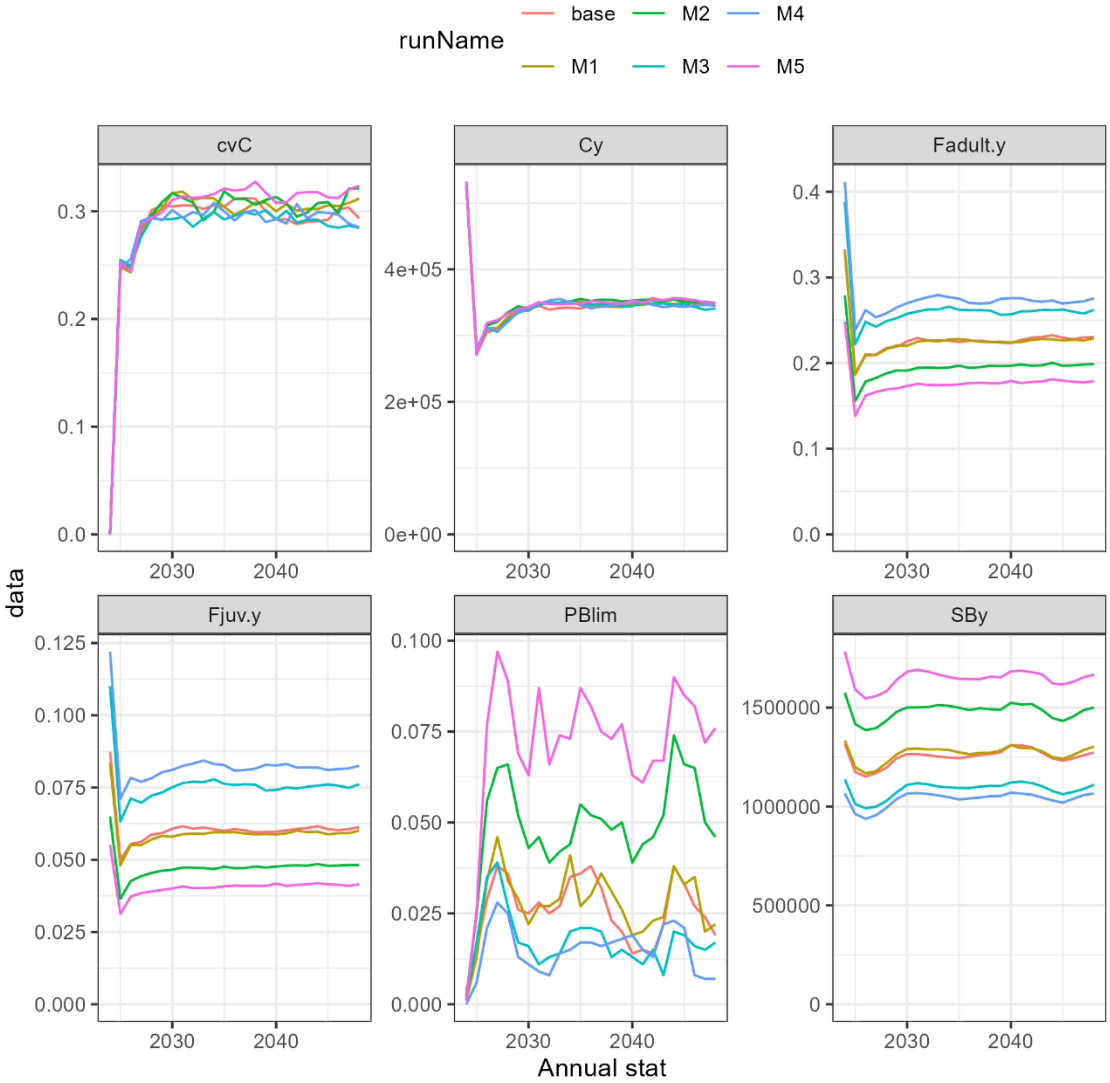


Figure 6.9: Annual stats over grid cell Btrigger=1.7e6-Ftarget=0.34 for the following OMs: base, M2, M3, M4 and M5. The yearly statistics are: CV in catch (cvC), catch (Cy), fishing pressure on adults (over age 2-6; Fadult.y), fishing pressure on juveniles (over age 0-1; Fjuv.y), probability of falling below Blim (PBlim) and SSB (SBy).

6.4 ROB.SR: robustness tests on depensation in recruitment

Depensation in recruitment can be impactful for the recovery of the stock at low stock size. The fitting of stock recruitment revealed that the fitting of segmented regression functions was poor and only a fitting of Beverton-Holt (BH) with depensation could be considered. For this reason, it was decided to consider depensation as a robustness test, conducted against replicates with BH as a stock-recruitment function. The fits between the different recruitment functions are shown in Figure 4.20 with SR3 being the fit of BH with depensation. It can be observed that the resulting depensation is small but results in higher steepness compared to other BH recruitment functions. The comparison of annual performance metrics between the SR2 and SR3 OMs is shown in Figure 6.10. A higher SSB (and in turn lower risk) is associated with the SR3 OM, which is due to the higher steepness estimated when including depensation as a parameter in the stock recruitment function.

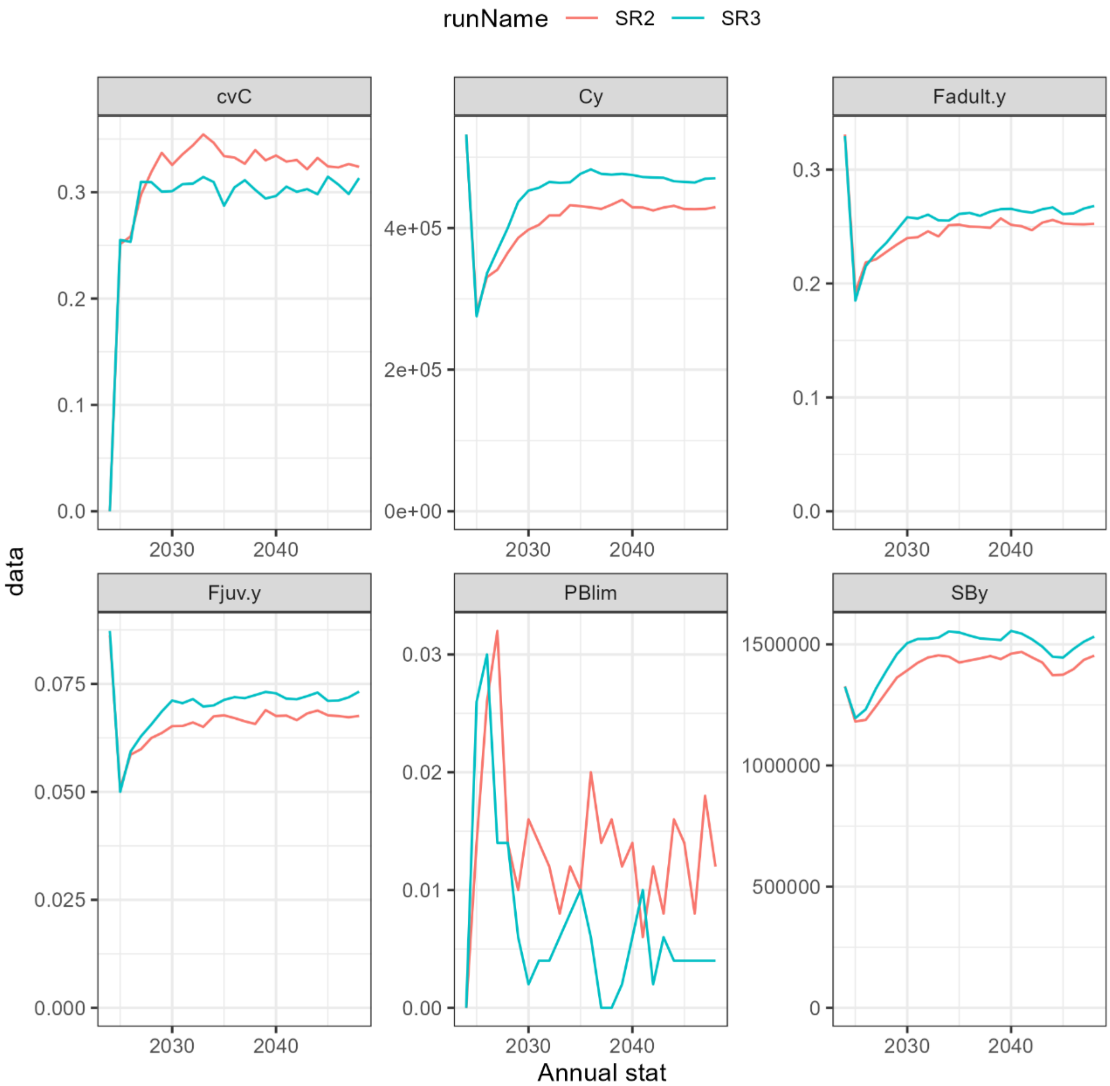


Figure 6.10: Annual stats over grid cell Btrigger=1.7e6-Ftarget=0.34 for the SR2 and SR3 OMs. Only the replicates with BH as stock recruitment function are used for the comparison. The yearly statistics are: CV in catch (cvC), catch (Cy), fishing pressure on adults (over age 2-6; Fadult.y), fishing pressure on juveniles (over age 0-1; Fjuv.y), probability of falling below Blim (PBlim) and SSB (SBy).

7 Testing of empirical MPs

This section describes additional testing of empirical rules that do not form part of the joint request, but is nevertheless presented as complementary analysis and an approach that may be useful in future.

7.1 Introduction

Default practices in ICES for MSEs are to use the working group stock assessment model and short-term forecasting procedure within the evaluated Management Procedures and only to consider alternative Harvest Control Rules (HCR). Although it is generally accepted that annual stock assessments and forecasts result in the 'best available science' to derive short-term management advice, the perception on usability and effectiveness of this approach changes when one tries to mimic this procedure within an MSE. A working document on 'SAM as Estimator' (see Annex 12 of the Scoping report: Annex 9) highlights some of the concerns with this approach, pointing specifically to underestimating uncertainty (i.e. hiding risk to stock collapse) and assuming that the stock assessment behaves fine for the next 30 years. It is not uncommon that ICES expert groups encounter issues with the stock assessment data input on a regular basis that affect estimated reference points and overall perception of the stock.

As an alternative, we here present an empirical-based method to sustainably manage the NSAS stock. In these empirical Management Procedures we use survey index values as direct input to an HCR. This bypasses the need to run stock assessments to set management advice and allows stakeholders without an in-depth knowledge of stock assessments and forecast methodology to calculate TAC advice. Advice based on empirical rules is, by design, not sensitive to changes in stock assessment assumptions on e.g. natural mortality, being some of the changes that have triggered several benchmarks for NSAS and overall changes in perception of the stock. This section describes two alternative methods to estimate the NSAS stock size matched with two HCR designs, tested within the NSAS MSE using the base OM.

7.2 Methods

For these analyses we use the base.OM, the base.OEM and within the MP we use the HERAS acoustic index as input to the HCR. The acoustic index is considered an accurate representation of the stock and estimated SSB from the SAM model is similar to the biomass estimate of the acoustic index (Figure 7.1). The acoustic index as used in the assessment holds data on numbers-at-age, but during the survey maturity at age and weight at age information is collected as well and used directly in the ICES working group assessment as input. This allows us to calculate SSB directly from the HERAS index data. The acoustic index is fitted with a catchability ≠ 1, and therefore, HERAS numbers-at-age are scaled by their catchability before being used in the HCR. Note however that this transformation is not needed *per se*, as it functions only as a scaling factor to allow direct comparison to e.g. biomass limit and target reference points. The HERAS data is used in two separate ways as an estimator of stock size. In the first approach, the SSB in the data year is taken as input to the HCR. In the second approach, the trend over the past 3 years as well as the average over the past 3 years is taken as input to the HCR. Two different HCRs are designed to process this information and set catch advice accordingly.

Figure 7.1: Time-trends of the SSB estimate of the 2024 SAM assessment (in red) and the bias corrected acoustic SSB from the HERAS survey (in blue).

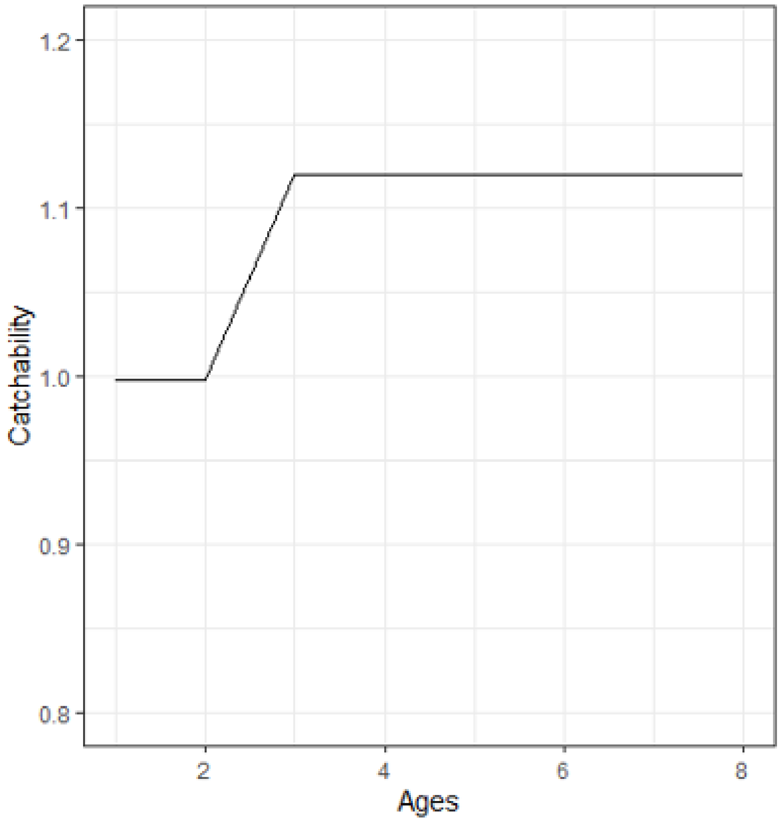


Figure 7.2: Estimated catchability at age for the HERAS survey in the 2024 HAWG assessment.

HCRs considered

Two different HCRs were designed to use the empirical HERAS data to set TAC advice. The first HCR uses the SSB estimate from the HERAS survey in the data year and proposes a catch target multiplier. This multiplier is derived from a relationship between SSB and the target multiplier as presented in Figure 7.3. It has a quadratic decline in target multiplier below a limit reference point, it has a multiplier of 1 in between an upper and lower buffer value and it has a linear increase in multiplier with increasing SSB above the upper buffer value. Tuning takes place over three parameter values, being the slope parameter, the upper buffer value and a target catch. For the upper buffer value we take either MSYBtrigger (at 1 130 747t; Table 3.1) or at $Bpa \times e^{1.65\sigma}$ where σ is the estimated σ from SSB in the terminal year of the assessment, resulting in a value of 1 306 773t. The mathematical implementation of the target multiplier of this HCR is given below:

$$\text{mult} = (SSB_{\text{HERAS}}/\text{lim})^2/2,\ \ SSB_{\text{HERAS}} \leq \text{lim}$$

$$\text{mult} = 0.5 \times \left(1 + (SSB_{\text{HERAS}} - \text{lim})/(\text{bufflow} - \text{lim})\right),\ \ SSB_{\text{HERAS}} > \text{lim}\ \&\ SSB_{\text{HERAS}} \leq \text{bufflow}$$

$$\text{mult} = 1,\ \ SSB_{\text{HERAS}} > \text{bufflow}\ \&\ SSB_{\text{HERAS}} \leq \text{buffup}$$

$$\text{mult} = 1 + \text{slope} \times 1/(2 \times (\text{bufflow} - \text{lim})) \times (SSB_{\text{HERAS}} - \text{buffup}), \quad SSB_{\text{HERAS}} > \text{buffup}$$

The target catch is then multiplied with 'mult' to set the TAC advice.

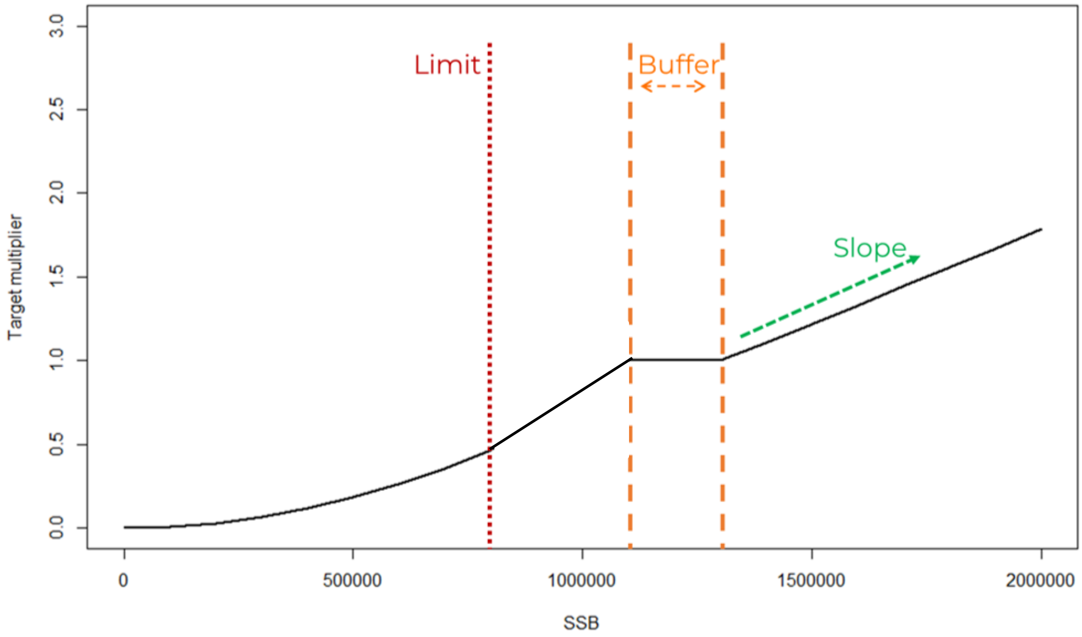


Figure 7.3: Design of HCR 1 (buffer hcr) showing a potential relationship (in black) between estimated SSB from the HERAS survey and the catch target multiplier. The changes in the shape of the relationship are defined by the limit biomass point (here taken as Blim from the HAWG 2024 assessment), the lower buffer value (here taken as Bpa) and the upper buffer value (taken as Bpa× exp[1.65×sigma], where sigma is the estimated sigma from SSB in the terminal year of the assessment) and a slope above the upper buffer value.

The second HCR uses the trend and average of the acoustic index over the past 3 years. It then scales both these values with a parameter being k1 (for the trend) and k3 (for the deviance of the average with the target). It is possible to expand the HCR to have different scaling parameters for increasing vs decreasing trends (which would add parameter k2) or having different parameters for the average being above or below the target (which would add parameter k4). Here, we only test situations where k1=k2 and k3=k4. The trend is calculated by fitting a linear model to the log(SSBHERAS) values over the most recent 3 datapoints.

$$TAC_{y+1} = TAC_y \times \left(1 + k_1 \times \text{trend} + k_3 \times (\text{bio}\bar{\text{m}}\text{as} - \text{tar})/\text{tar}\right)$$

where bio$\bar{\text{m}}$ass is the average of the biomass over specific number of years (e.g. 3 years in Figure 7.4). The parameters to tune in this case are k1, k3 and the target biomass. This HCR is referred to as the slope HCR.

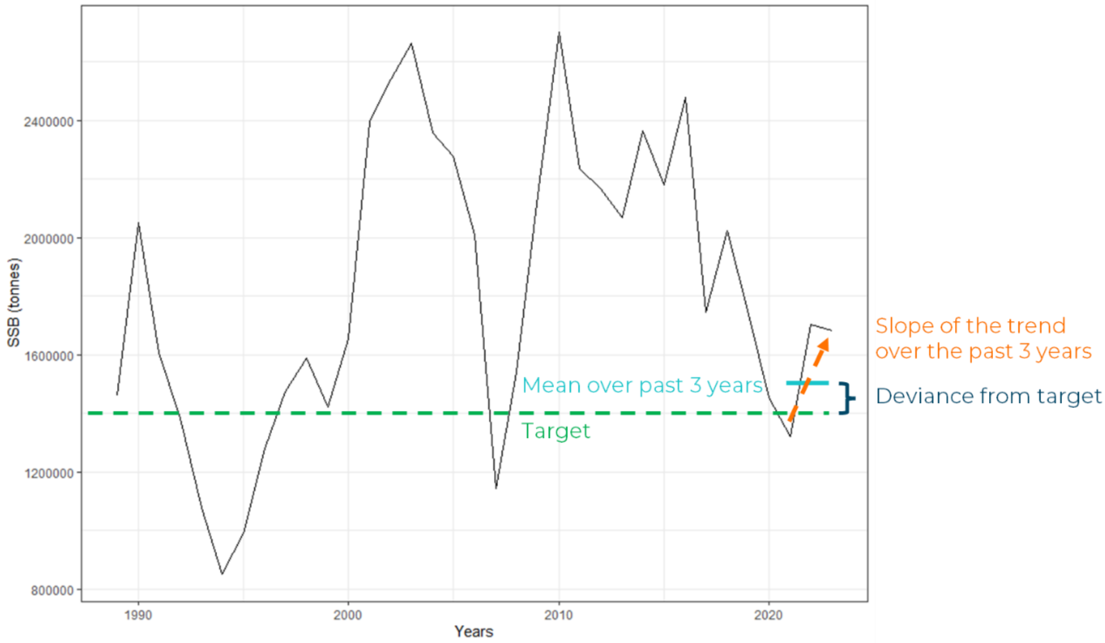


Figure 7.4: Schematic of the slope rule.

7.3 Results

The results of the MP with SSB taken from the HERAS survey and used directly in the buffer.hcr is presented in Figure 7.5. The results show that target catches at around 300,000 t would be feasible in the long run with risks ≤ 5%. When interpolating risk, optimal performance in terms of long-term risk (ICES risk 3) and catch would be achieved at a target catch of 300,000 t in combination with a slope (i.e. k1) of the HCR at 0.2. Realised long-term catches in this situation equates to 333877t annually.

Explorations changing the upper buffer value did not yield better performance and are hence not shown.

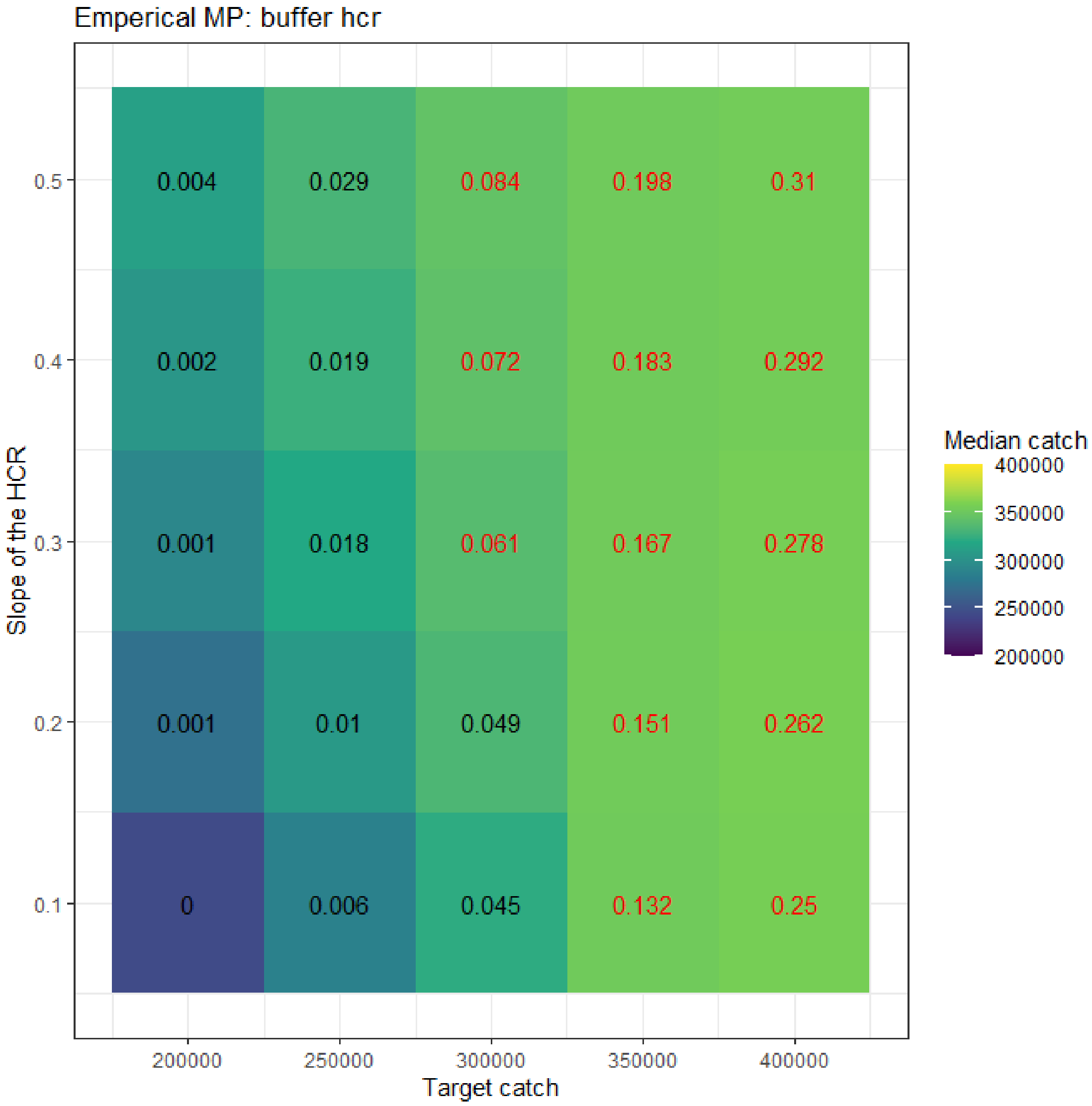


Figure 7.5: Tuning results of the empirical MP including the buffer HCR. The colour of the panels refer to long-term (2034-2048) average catch (t), the numbers in each of these grid cells refer to the long term risk (fraction) which is colour-coded for value > 5% risk (ICES risk 3 definition).

The results for the slope HCR, tuning 3 parameters, need to be visualized in a multi-panel plot. Initial performance was evaluated for k1 parameters between a value of 0.74– 1.6 in steps of 0.1, for k3 between 0.05 – 0.20 in steps of 0.025 and for target biomass at 1.2Mt to 1.4Mt in steps of 0.1Mt. Results are given in Figure 7.6 for a target biomass of 1.4Mt, and comparisons with the model-based management strategy of previous sections shown in Figure 7.7. Long-term average catches at a risk of 5% are at 335kt. It can be observed that the management procedure described in Section 5 exemplifies the best performances (highest catch with limited risk). However, the empirical rules have similar performances, especially considering tuning could be further improved.

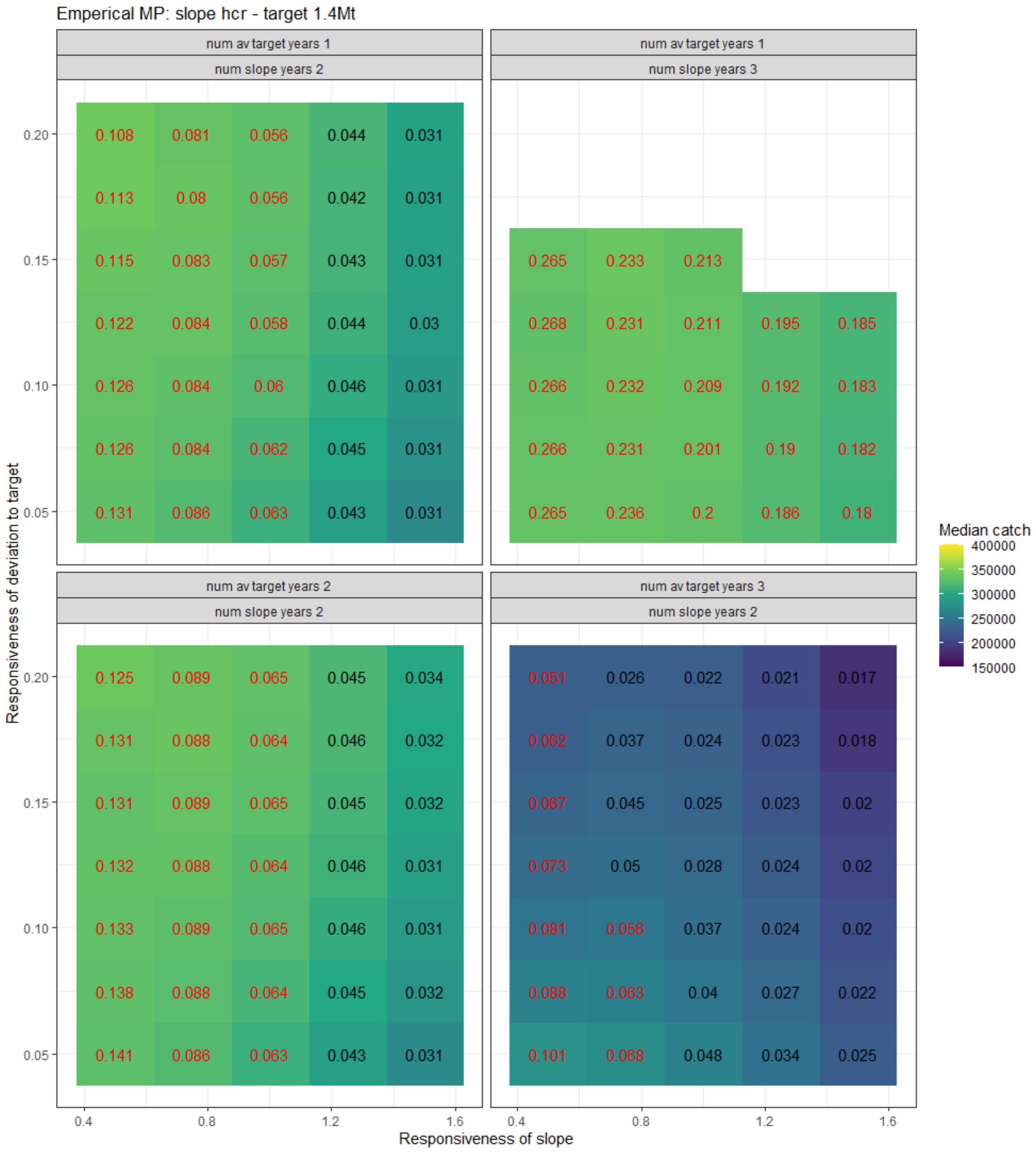


Figure 7.6: Tuning results of the empirical MP including the slope HCR. The colour of the panels refer to long-term (2034-2048) average catch (t), the numbers in each of these grid cells refer to the long-term risk (fraction) which is colour-coded for value > 5% risk (ICES risk 3 definition). The panels refer to how the slope and deviance of the target are averaged over most recent years.

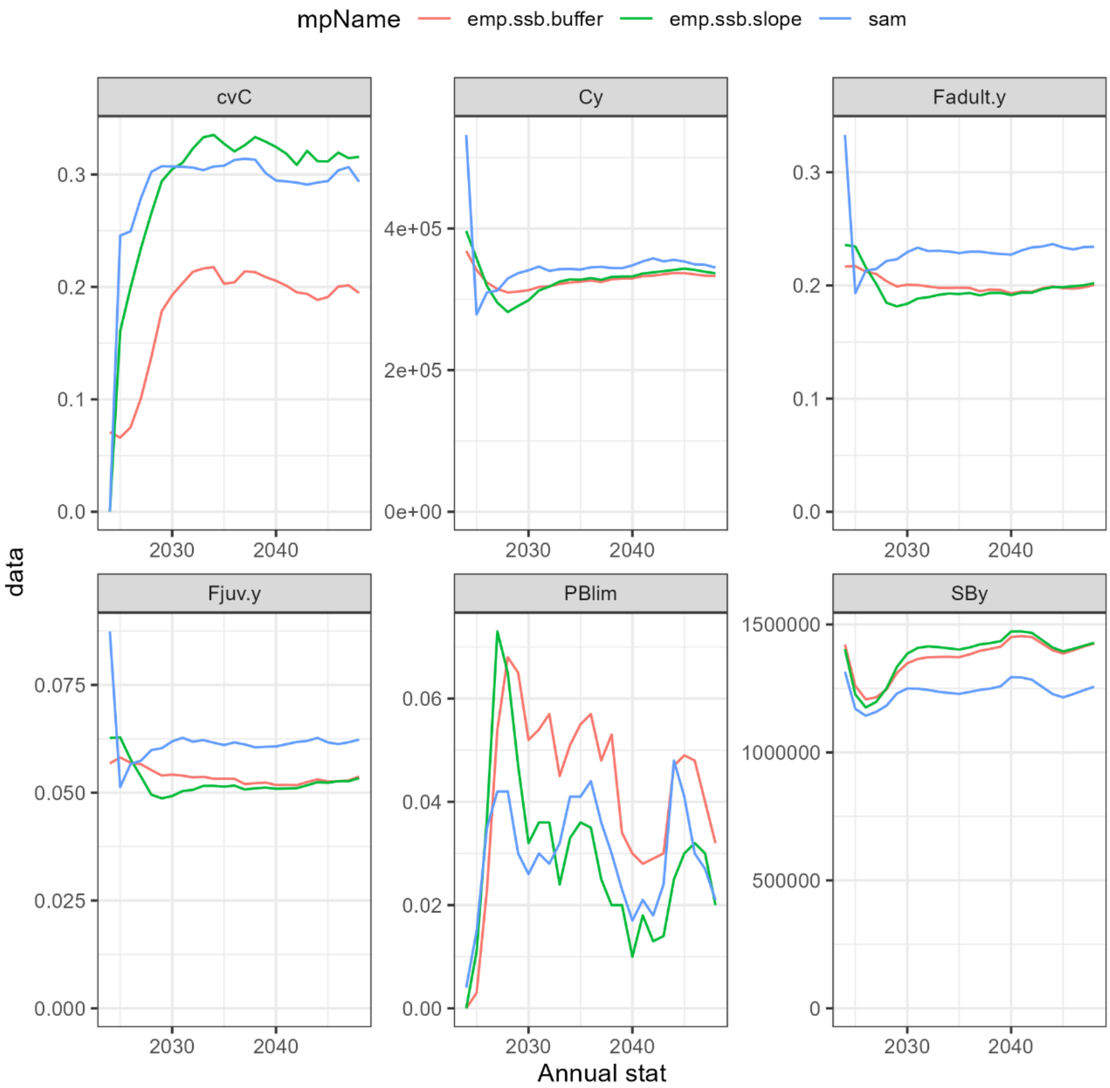


Figure 7.7. Annual stats for the base OM with different types of management procedures. Blue line: stock assessment model combined with HCR as shown in Figure 3.16 (over grid cell Btrigger=1.7e6-Ftarget=0.35). Red line: empirical buffer rule (slope of 0.2, target biomass of 300,000 t). Green lines: empirical slope rule (k1 = 1, k3 = 0.05, target biomass of 1.1 mt and averaging in years of 2, green lines). The computations are done on the base OM only. For each type of management procedure, tuning is done separately. The yearly statistics are: CV in catch (cvC), catch (Cy), fishing pressure on adults (over age 2-6; Fadult.y), fishing pressure on juveniles (over age 0-1; Fjuv.y), probability of falling below Blim (PBlim) and SSB (SBy).

7.4 Conclusion

Evaluation of two MPs that used the HERAS acoustic biomass as input to an HCR showed that sustainably managing the NSAS is possible and results in maintaining the stock at biomass levels well-above MSYBtrigger and fishing mortality at or below FMSY. The results furthermore indicate that both HCRs evaluated perform similarly resulting in the tuned situation in long-term catches at around 335,000 t.

Additional benefits of these empirical MPs are the ability to directly derive TAC from the HERAS index, reducing the dependency on extensive data collection and complex stock assessment procedures which are prone to errors and negatively impact ICES credibility when mistakes have to be corrected past advice publication date.

8 References

Bekkevold, Dorte, Florian Berg, Patrick Polte, Valerio Bartolino, Henn Ojaveer, Henrik Mosegaard, Edward D. Farrell, et al. 2023. “Mixed-stock analysis of Atlantic herring (Clupea harengus): a tool for identifying management units and complex migration dynamics.” *ICES Journal of Marine Science* 80 (1): 173–84. https://doi.org/10.1093/ICESJMS/FSAC223.

De Oliveira, J. 2013. Evaluation of a Banking and Borrowing Scheme for Saithe. ICES CM 2013/ACOM:65. 71 pp.

Fisher, Simon. 2024. “Operating Models for Plaice.” https://github.com/shfischer/WKBPLAICE2024_ple.27.7e_MSE/blob/WKBPLAICE2024/WKBPLAICE2024_ple.27.7e_OM.pdf.

ICES. 2018. “Report of the Benchmark Workshop on Pelagic Stocks (WKPELA 2018).” In *ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:32*, 313.

———. 2019. “Workshop on North Sea stocks management strategy evaluation (WKNSMSE).” *ICES Scientific Reports* 1:12: 378. https://doi.org/doi.org/10.17895/ices.pub.5090.

———. 2021. “Inter-Benchmark Protocol of North Sea Herring (IBPNSHerring).” *ICES Scientific Reports. 3:98*.

———. 2023. “Advice on fishing opportunities.” *Report of the ICES Advisory Committee, 2023 section 1.1.1.* https://doi.org/10.17895/ices.advice.22240624

———. 2024a. Herring Assessment Working Group (HAWG). ICES Scientific Reports. 06:24. 969 pp. https://doi.org/10.17895/ices.pub.25305532

———. 2024b. “Report of the Working Group on Multispecies Assessment Methods (WGSAM).”———.

———. 2024c. “Msy r Package.” https://github.com/ices-tools-prod/msy.

Kell, L. T., I. Mosqueira, P. Grosjean, J-M. Fromentin, D. Garcia, R. Hillary, E. Jardim, et al. 2007. “FLR: an open-source framework for the evaluation and development of management strategies.” *ICES Journal of Marine Science* 64 (4): 640–46. https://doi.org/10.1093/icesjms/fsm012.

Nielsen, Anders, and Casper W. Berg. 2014. “Estimation of time-varying selectivity in stock assessments using state-space models.” *Fisheries Research* 158 (October): 96–101. https://doi.org/10.1016/j.fishres.2014.01.014.

Nielsen, Anders, Niels T Hintzen, Henrik Mosegaard, Vanessa Trijoulet, and Casper W Berg. 2021. “Multi-fleet state-space assessment model strengthens confidence in single-fleet SAM and provides fleet-specific forecast options.” *ICES Journal of Marine Science*, July. https://doi.org/10.1093/ICESJMS/FSAB078.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A. and M. Haddon. 2016. Management strategy evaluation: best practices. Fish and Fisheries 17: 303-334.

Roustant, Olivier, David Ginsbourger, and Yves Deville. 2012. “DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization.” *Journal of Statistical Software* 51 (1): 1–55. https://www.jstatsoft.org/v51/i01/.

Scott, Finlay, and Iago Mosqueira. 2023a. “FLasher FLasher_reference Vignette.” https://github.com/flr/FLasher/blob/master/vignettes/FLasher_reference.Rmd.

———. 2023b. “FLasher r Package.” https://github.com/flr/FLasher.

Annex 1: Long Term Management Strategy (LTMS) request

The European Union, Norway, and the United Kingdom jointly request ICES to advise on the long-term management strategies on North Sea autumn spawners herring (Clupea harengus) in Subarea 4 and divisions 3.a and 7.d, (North Sea, Skagerrak and Kattegat, eastern English Channel). A request is provided below.

ICES is requested to identify appropriate precautionary combinations in the format of Tables given in its response to the EU, Norway and the United Kingdom request to ICES to evaluate a multi-annual management strategy for herring (Clupea harengus) in Subarea 4 and divisions 3.a and 7.d, autumn spawners (North Sea, Skagerrak and Kattegat, eastern English Channel) (her.27.3a47d), using:

- *A harvest control rule with a fishing mortality equal to the target F when SSB is at or above* $B_{trigger}$. *In the case that the SSB is forecast to be less than* $B_{trigger}$ *at spawning time in the year for which the TAC is to be set, the TAC shall be fixed consistently with a fishing mortality that is given by:* $F = F_{target} \times SSB/B_{trigger}$
- *A range of* $B_{trigger}$ *from 800 000 to 1 700 000 tonnes with a range of target Fs up to* F_{lim}
- *For the combinations above explore the following exploitation pattern scenarios:*
  1. *Recent exploitation pattern (averaged over 2012-2021).*
  2. *A historic exploitation pattern (averaged over 1998-2007).*
  3. *Ranges of assumptions for values of* F_{0-1} *that vary between 0-0.1 independent from recent exploitation patterns for older fish* (F_{2+}).
  4. *The recent exploitation pattern with* $F_{0-1}=0$ *from above contrasted with exploitation patterns moved to one year older and one year younger fish (three scenarios).*

Long term goals:

- *Maximise yield*
- *Minimising the risk of falling below* B_{lim}
- *Achieve stability of catches*

All alternatives should be evaluated with and without a constraint on the inter-annual variation of TAC. When the rules would lead to a TAC, which deviates by more than 20% below or 25% above the TAC of the preceding year, the Parties shall fix a TAC that is respectively no more than 20% less or 25% more than the TAC of the preceding year. The TAC constraint shall not apply if the SSB at spawning time in the year for which the TAC is to be set is less or equal to $B_{trigger}$.

The constraint mechanism shall be tested separately from and in combination with 10% banking and borrowing mechanism. Banking and borrowing should be suspended when SSB is below $B_{trigger}$.

Evaluation and performance criteria:
Each alternative shall be assessed in relation to how it performs in the short term (5 years), medium term (next 10 years) and long term (next 25 years) in relation to:

- *Average SSB*
- *Average yield*
- *Indicator for year to year variability in SSB and yield*
- *Risk of SSB falling below* B_{lim}

Clarification on request:
Further amendments and clarifications were agreed with requesters, as follows:

- The recent exploitation pattern (point 1 of exploitation pattern scenarios) should be over the period 2013-2021
- The historic exploitation pattern (point 2 of exploitation pattern scenarios) should be over the period 1998-2003
- The interpretation of short-, medium-, and long-term (under evaluation performance criteria) should be short = years 1-5, medium = years 6-10, and long = years 11-25 of the projection period.

Annex 2: List of Participants

In-person and online meetings

Name	Institute	Country (of institute)	Email
Aaron Brazier			Aaron.brazier@cefas.gov.uk
Ann-Kristin Westberg			Ann-kristin.westberg@nfd.dep.no
Ashley Wilson			awilson@pewtrusts.org
Benoit Berges			benoit.berges@wur.nl
Bernhard Kuehn			bernhard.kuehn@thuenen.de
Campbell Pert			Campbell.Pert@gov.scot
Carryn de Moor [Reviewer]			carryn.demoor@uct.ac.za
Cecilie Kvamme			cecilie.kvamme@hi.no
Claus Reedtz Sparrevohn			crs@pelagisk.dk
Daniel Howell			daniel.howell@hi.no
Esben Sverdrup			es@pelagisk.dk
Florian Berg			florian.berg@hi.no
Hannah Sterling			hannah.sterling@defra.gov.uk
Helen Bradbrook			helen.bradbrook@defra.gov.uk
Henrik Mosegaard			hm@aqua.dtu.dk
Imogen Cessford			imogen.cessford@defra.gov.uk
Jaylene Mbararia [Secretariat]			Jaylene.Mbararia@ices.dk
John Tyler Trochta			john.tyler.trochta@hi.no
José De Oliveira [Chair]			jose.deoliveira@cefas.gov.uk
Jurgen Batsleer			Jurgen.Batsleer@ec.europa.eu
Lise Laustsen			ll@pelagisk.dk
Niels Hintzen			nhintzen@pelagicfish.eu
Nis Sand Jacobsen			nsja@aqua.dtu.dk
Norbert Rohlf			norbert.rohlf@thuenen.de

Name	Institute	Country (of institute)	Email
Paul Marchal			paul.marchal@ifremer.fr
Richard Nash			Richard.nash@cefas.gov.uk
Sarah Louise Millar [Secretariat]			sarah-louise.millar@ices.dk
Snorri Palmason			Snorri.Palmason@fiskeridir.no
Tamara Talevska			tamarat@nsrac.org
Tanja Miethe			tanja.miethe@gov.scot
Thomas Carruthers [Reviewer]			tom@bluematterscience.com
Valerio Bartolino			valerio.bartolino@slu.se
Vanessa Trijoulet			vtri@aqua.dtu.dk

Stakeholder engagement meeting (participants not listed above)

Name	Institute	Country (of institute)	Email
Anne-Marie Kats			a.j.kats@minlnv.nl
Anthony Banny			tonyban83@gmail.com
Arthur Yon			a.yon@fromnord.fr
Gjert Dingsør			gjert@fiskebat.no
Jean-Christophe Vandevelde			jvandevelde@pewtrusts.org
Johanne Frost Klepp			jfk@naturvernforbundet.no
Jules Danto			jules.danto@eapo.com
Justyna Zajchowska			jzajchowska@endoverfishing.eu
Katie Brigden			katie.brigden@uhi.ac.uk
Marta Ballesteros [Chair stakeholder engagement meeting]			marta.ballesteros@ieo.csic.es
Paul Thomas			p.thomas@pelagic-ac.org
Per-Erik Schulze			pes@naturvernforbundet.no
Tamara Talevska			tamarat@nsrac.org

Annex 3: OM trajectories

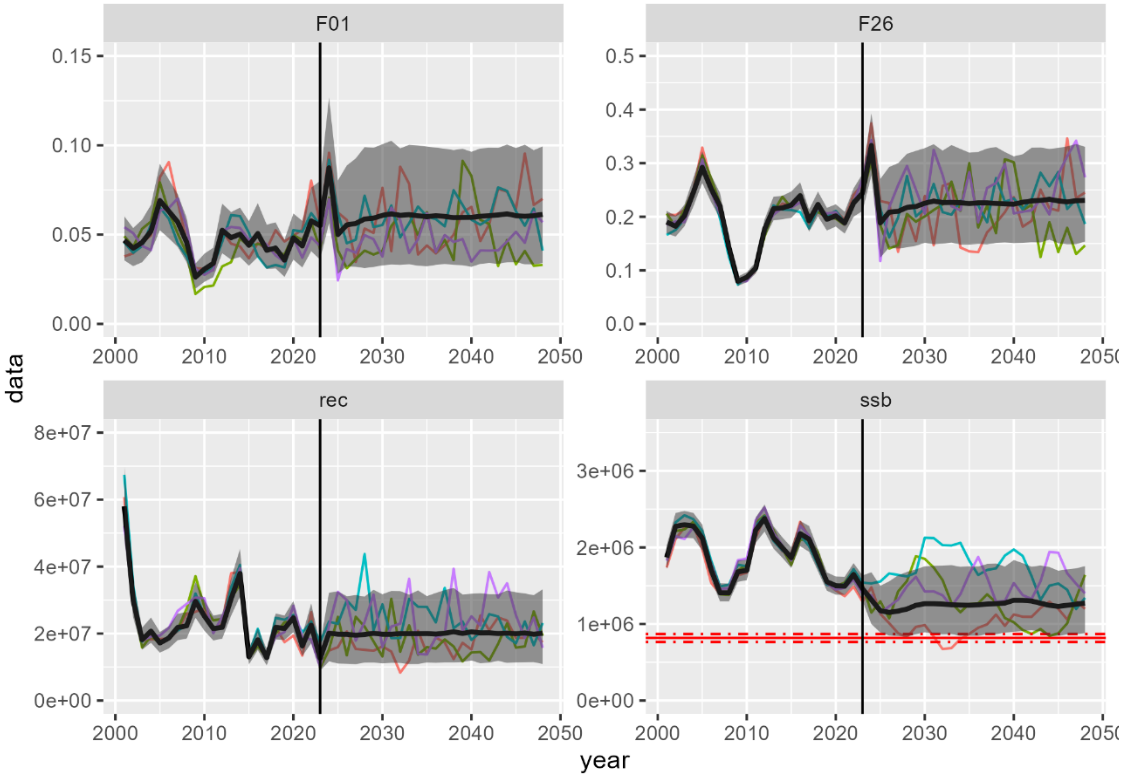


Figure A3.1: Base OM stock trajectory for Btrigger=1.7e6-Ftarget=0.34. The horizontal red lines in the ssb facet show the distribution in biological reference point Blim across replicates (solid line as the median, dashed lines as the 95th/5th percentiles). The grey shaded ribbons are the 95th/5th quantiles and the solid black line the median for each quantity.

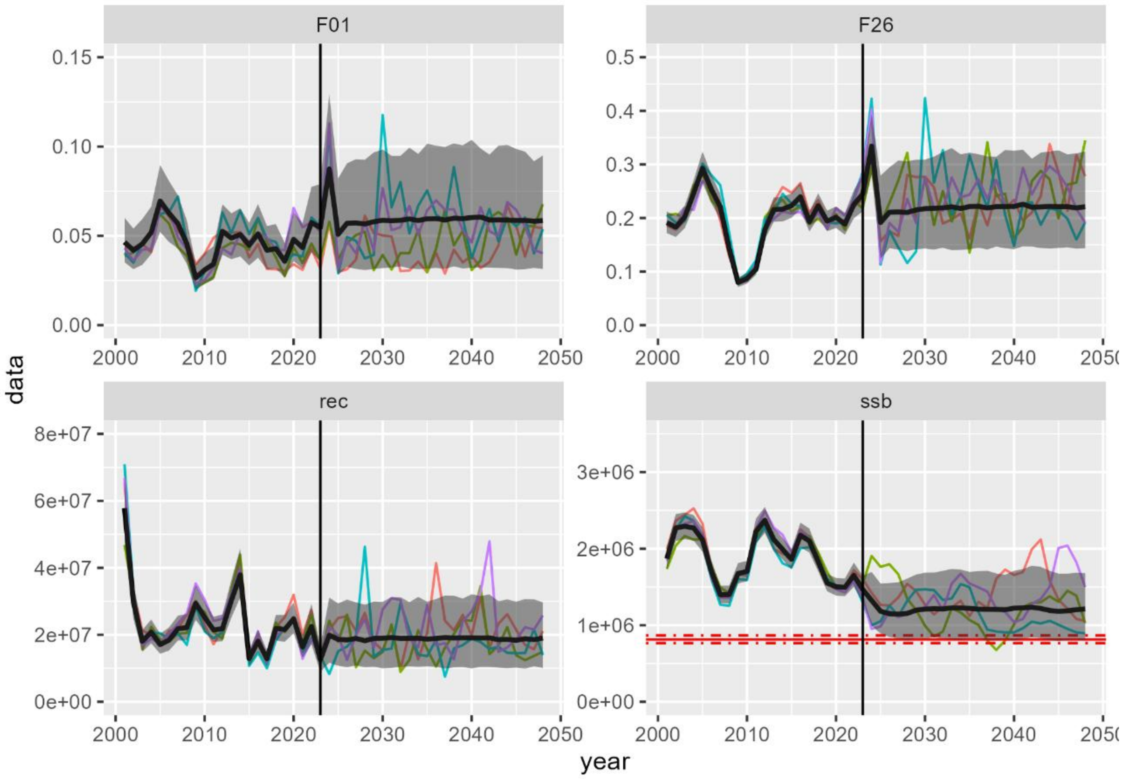


Figure A3.2: SR1 OM stock trajectory for Btrigger=1.7e6-Ftarget=0.34. The horizontal red lines in the ssb facet show the distribution in biological reference point Blim across replicates (solid line as the median, dashed lines as the 95th/5th percentiles). The grey shaded ribbons are the 95th/5th quantiles and the solid black line the median for each quantity.

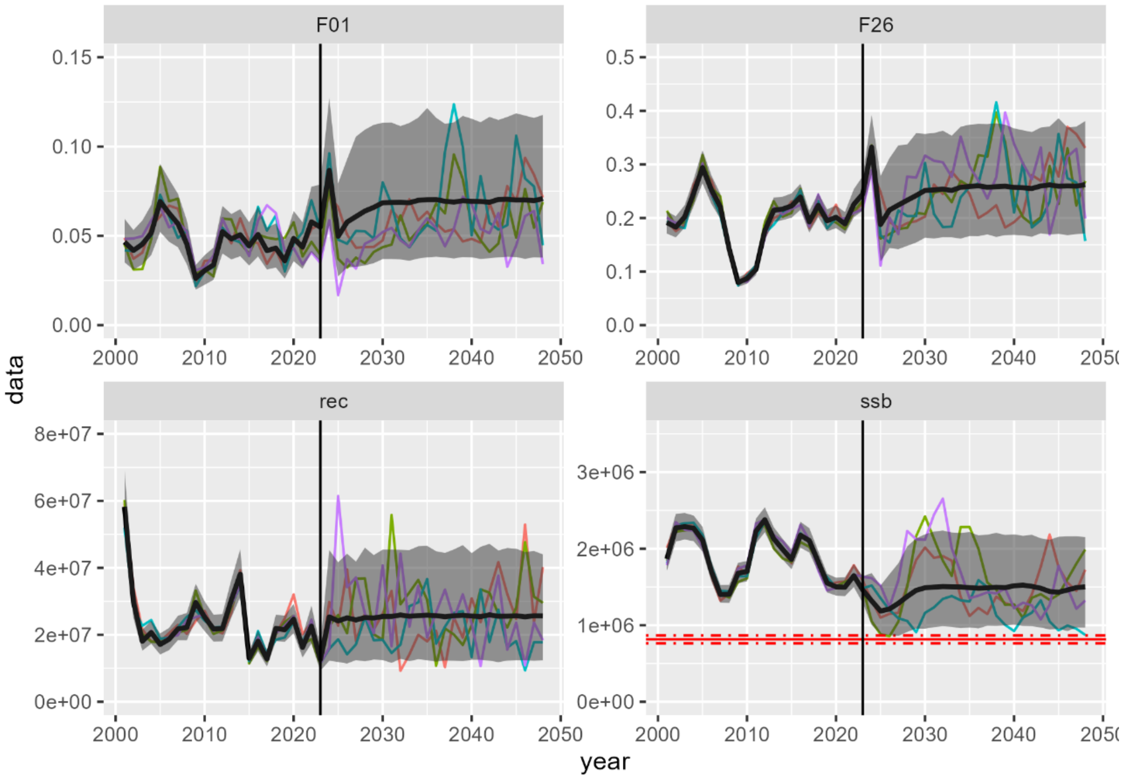


Figure A3.3: SR2 OM stock trajectory for Btrigger=1.7e6-Ftarget=0.34. The horizontal red lines in the ssb facet show the distribution in biological reference point Blim across replicates (solid line as the median, dashed lines

as the 95th/5th percentiles). The grey shaded ribbons are the 95th/5th quantiles and the solid black line the median for each quantity.

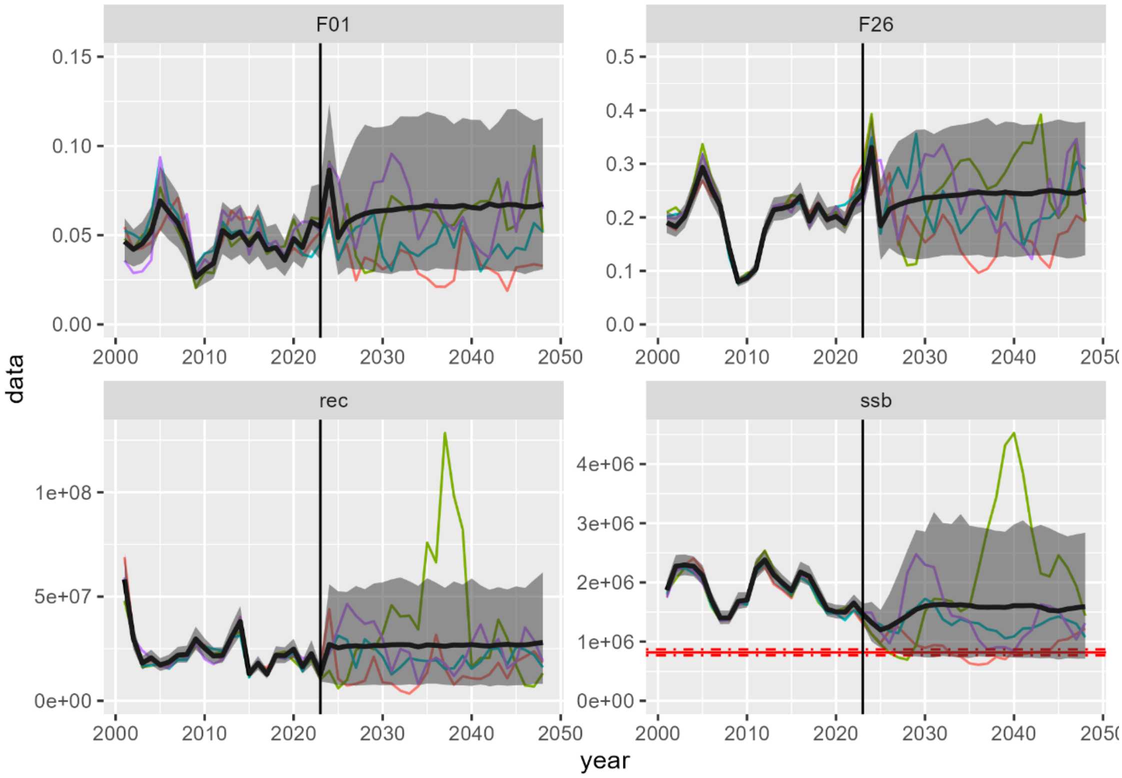


Figure A3.4: SR4 OM stock trajectory for Btrigger=1.7e6-Ftarget=0.34. The horizontal red lines in the ssb facet show the distribution in biological reference point Blim across replicates (solid line as the median, dashed lines as the 95th/5th percentiles). The grey shaded ribbons are the 95th/5th quantiles and the solid black line the median for each quantity.

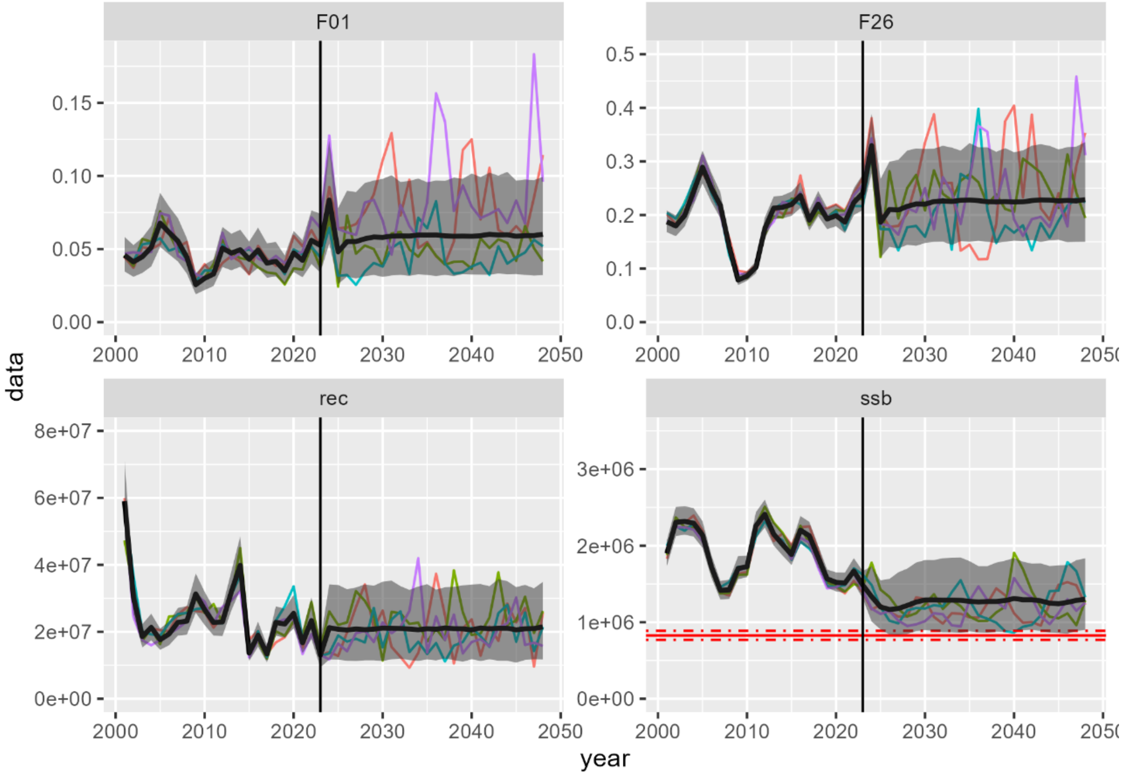


Figure A3.5: M1 OM stock trajectory for Btrigger=1.7e6-Ftarget=0.34. The horizontal red lines in the ssb facet show the distribution in biological reference point Blim across replicates (solid line as the median, dashed lines as the 95th/5th percentiles). The grey shaded ribbons are the 95th/5th quantiles and the solid black line the median for each quantity.

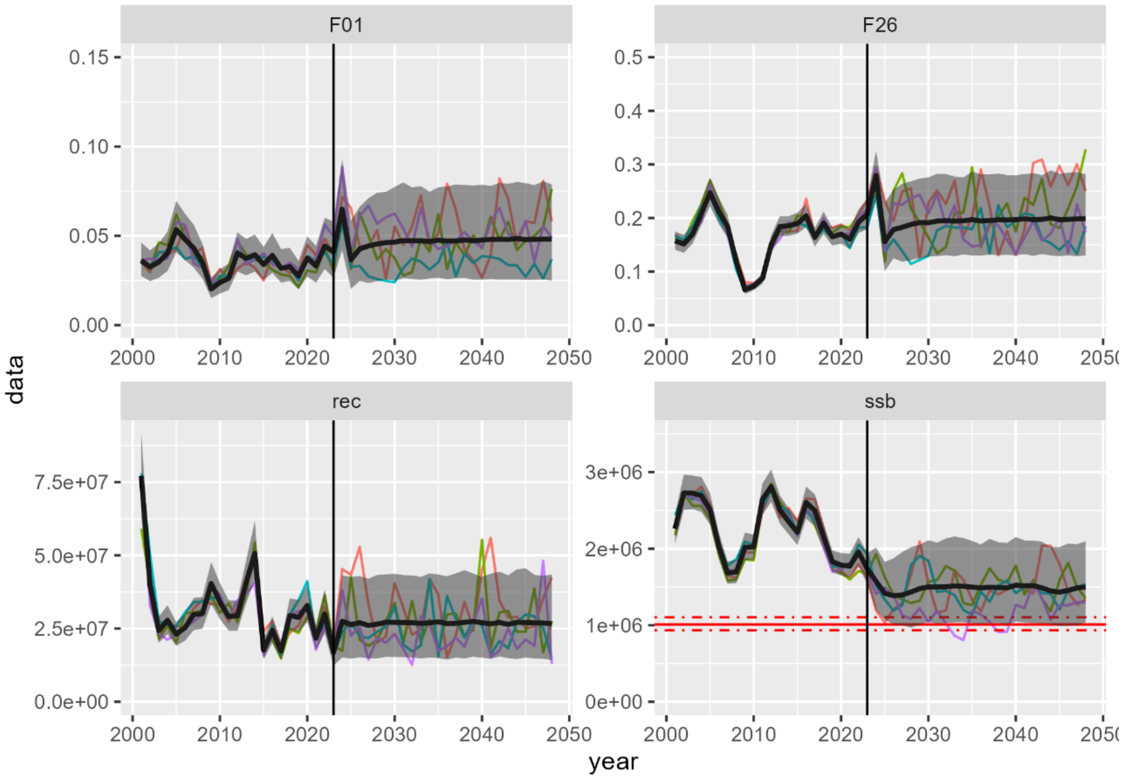


Figure A3.6: M2 OM stock trajectory for Btrigger=1.7e6-Ftarget=0.34. The horizontal red lines in the ssb facet show the distribution in biological reference point Blim across replicates (solid line as the median, dashed lines

as the 95th/5th percentiles). The grey shaded ribbons are the 95th/5th quantiles and the solid black line the median for each quantity.

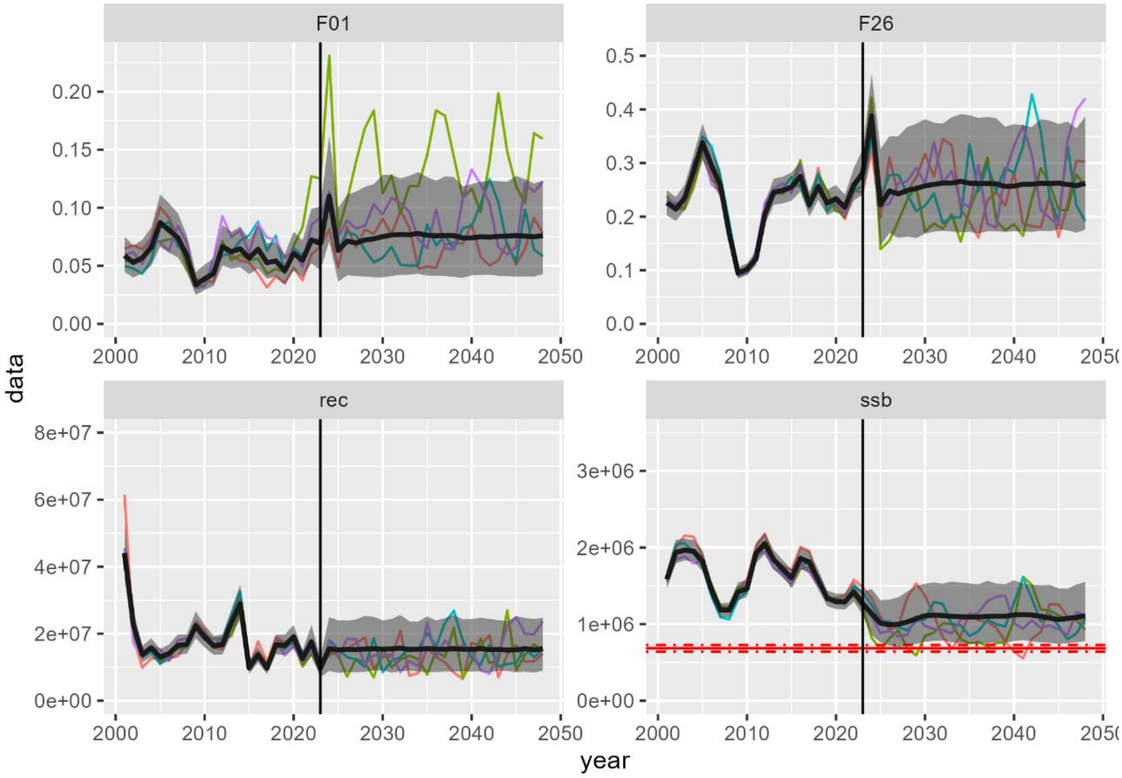


Figure A3.7: M3 OM stock trajectory for Btrigger=1.7e6-Ftarget=0.34. The horizontal red lines in the ssb facet show the distribution in biological reference point Blim across replicates (solid line as the median, dashed lines as the 95th/5th percentiles). The grey shaded ribbons are the 95th/5th quantiles and the solid black line the median for each quantity.

Annex 4: Performance metrics grids

Table A4.1: long-term ICES risk 3 over grid cells covered with computation. The results are the OM combining individuals OMs in the reference set.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	0.4449									
0.38					0.2314					
0.37									0.0818	
0.36			0.2902							
0.35							0.1093			0.0511
0.34	0.3217								0.0590	0.0440
0.33					0.1439				0.0510	0.0369
0.32								0.0597	0.0439	
0.31			0.1856					0.0512		
0.3							0.0593	0.0433		
0.29	0.1947						0.0509			
0.28					0.0723	0.0573	0.0413			
0.27						0.0491			0.0203	
0.26			0.0863		0.0551					
0.25					0.0459		0.0233			
0.24	0.0812			0.0501						
0.23			0.0492	0.0406	0.0286					
0.22	0.0529	0.0466	0.0411							
0.21	0.0407		0.0297							
0.2										
0.19	0.0222								0.0043	
0.18										

Table A4.2: long-term ICES risk 3 over all combination of Btrigger-Ftarget using interpolation. The results are the OM combining individuals OMs in the reference set. Cells that coincide with the non-empty cells in the preceding table are actual values and are not interpolated.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	0.4449	0.3903	0.3453	0.2993	0.2366	0.1925	0.1464	0.1172	0.0916	0.0698
0.38	0.4352	0.3821	0.3385	0.2934	0.2314	0.1880	0.1424	0.1138	0.0886	0.0673
0.37	0.4067	0.3565	0.3152	0.2724	0.2137	0.1738	0.1316	0.1051	0.0818	0.0620
0.36	0.3764	0.3291	0.2902	0.2500	0.1948	0.1585	0.1198	0.0956	0.0742	0.0562
0.35	0.3497	0.3051	0.2683	0.2302	0.1782	0.1450	0.1093	0.0872	0.0676	0.0511
0.34	0.3217	0.2797	0.2452	0.2094	0.1608	0.1308	0.0983	0.0773	0.0590	0.0440
0.33	0.2934	0.2543	0.2222	0.1888	0.1439	0.1169	0.0875	0.0679	0.0510	0.0369
0.32	0.2718	0.2351	0.2048	0.1733	0.1312	0.1058	0.0785	0.0597	0.0439	0.0323
0.31	0.2480	0.2138	0.1856	0.1563	0.1172	0.0938	0.0689	0.0512	0.0380	0.0284
0.3	0.2200	0.1886	0.1628	0.1369	0.1022	0.0814	0.0593	0.0433	0.0325	0.0247
0.29	0.1947	0.1660	0.1424	0.1195	0.0888	0.0704	0.0509	0.0375	0.0284	0.0219
0.28	0.1581	0.1348	0.1156	0.0974	0.0723	0.0573	0.0413	0.0309	0.0237	0.0185
0.27	0.1359	0.1163	0.1001	0.0849	0.0633	0.0491	0.0348	0.0263	0.0203	0.0162
0.26	0.1165	0.1000	0.0863	0.0737	0.0551	0.0419	0.0291	0.0221	0.0173	0.0139
0.25	0.0996	0.0857	0.0741	0.0625	0.0459	0.0342	0.0233	0.0180	0.0141	0.0116
0.24	0.0812	0.0699	0.0604	0.0501	0.0360	0.0272	0.0188	0.0146	0.0116	0.0097
0.23	0.0648	0.0564	0.0492	0.0406	0.0286	0.0218	0.0153	0.0120	0.0096	0.0081
0.22	0.0529	0.0466	0.0411	0.0341	0.0243	0.0186	0.0132	0.0104	0.0084	0.0072
0.21	0.0407	0.0347	0.0297	0.0250	0.0181	0.0141	0.0101	0.0081	0.0066	0.0058
0.2	0.0301	0.0260	0.0225	0.0192	0.0141	0.0111	0.0081	0.0065	0.0053	0.0048
0.19	0.0222	0.0195	0.0171	0.0147	0.0109	0.0087	0.0064	0.0052	0.0043	0.0039
0.18	0.0226	0.0199	0.0175	0.0151	0.0113	0.0091	0.0068	0.0055	0.0046	0.0042

Table A4.3: long-term mean catch over grid cells covered with computation. The results are the OM combining individuals OMs in the reference set.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	370098									
0.38					375579					
0.37									374679	
0.36			374063							
0.35							374979			372608
0.34	371727								372985	371888
0.33					374567				372277	371080
0.32								372531	371450	
0.31			373147					371659		
0.3							371761	370733		
0.29	370796						370817			
0.28					371155	370605	369702			
0.27						369373			365678	
0.26			368927		368661					
0.25					367098		365377			
0.24	365079			365618						
0.23			363512	363505	363110					
0.22	360611	360838	361020							
0.21	357762		358115							
0.2										
0.19	350581								346206	
0.18										

Table A4.4: long-term mean catch over all combination of Btrigger-Ftarget using interpolation. The results are the OM combining individuals OMs in the reference set. Cells that coincide with the non-empty cells in the preceding table are actual values and are not interpolated.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	370098	371301	372518	373473	374295	374300	374115	373761	373219	372238
0.38	371334	372554	373791	374754	375579	375548	375318	374912	374311	373250
0.37	371847	373041	374252	375175	375952	375935	375709	375297	374679	373579
0.36	371780	372914	374063	374908	375597	375551	375282	374815	374130	373042
0.35	371838	372921	374018	374781	375377	375296	374979	374451	373692	372608
0.34	371727	372750	373785	374455	374947	374816	374436	373834	372985	371888
0.33	371708	372662	373625	374191	374567	374376	373919	373229	372277	371080
0.32	371674	372557	373449	373909	374169	373892	373335	372531	371450	370330
0.31	371527	372334	373147	373495	373630	373258	372589	371659	370492	369455
0.3	371201	371922	372647	372913	372953	372514	371761	370733	369481	368528
0.29	370796	371427	372059	372238	372176	371663	370817	369700	368369	367504
0.28	369985	370568	371148	371280	371155	370605	369702	368505	367103	366332
0.27	369043	369572	370092	370167	369969	369373	368412	367144	365678	365011
0.26	368004	368473	368927	368939	368661	368009	367008	365686	364173	363610
0.25	366653	367044	367414	367410	367098	366405	365377	364012	362461	362014
0.24	365079	365379	365653	365618	365255	364511	363443	362051	360478	360160
0.23	363002	363274	363512	363505	363110	362331	361239	359834	358250	358072
0.22	360611	360838	361020	360976	360556	359765	358670	357269	355693	355672
0.21	357762	357964	358115	358057	357634	356850	355770	354392	352837	352987
0.2	354162	354394	354567	354530	354137	353392	352356	351027	349520	349862
0.19	350581	350843	351037	351021	350657	349949	348954	347670	346206	346739
0.18	351104	351354	351539	351524	351177	350500	349549	348322	346923	347432

Table A4.5: long-term mean catch IAV over grid cells covered with computation. The results are the OM combining individuals OMs in the reference set.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	15.52									
0.38					18.43					
0.37									19.05	
0.36			16.37							
0.35							18.34			18.68
0.34	13.68								18.44	18.49
0.33					16.80				18.22	18.29
0.32								17.81	17.99	
0.31			14.38					17.53		
0.3							16.85	17.23		
0.29	11.96						16.52			
0.28					14.81	15.55	16.15			
0.27						15.12			16.61	
0.26			12.26		13.92					
0.25					13.46		14.93			
0.24	10.52			12.18						
0.23			11.09	11.74	12.50					
0.22	10.06	10.32	10.73							
0.21	9.85		10.40							
0.2										
0.19	9.48								13.39	
0.18										

Table A4.6: long-term mean catch IAV over all combination of Btrigger-Ftarget using interpolation. The results are the OM combining individuals OMs in the reference set. Cells that coincide with the non-empty cells in the preceding table are actual values and are not interpolated.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	15.52	16.23	17.15	17.83	18.59	18.98	19.23	19.34	19.34	19.29
0.38	15.25	15.99	16.94	17.64	18.43	18.84	19.11	19.23	19.25	19.21
0.37	14.91	15.65	16.62	17.33	18.14	18.57	18.86	19.02	19.05	19.03
0.36	14.62	15.39	16.37	17.11	17.94	18.41	18.73	18.91	18.97	18.96
0.35	14.12	14.88	15.86	16.61	17.47	17.98	18.34	18.56	18.66	18.68
0.34	13.68	14.44	15.41	16.20	17.08	17.63	18.04	18.31	18.44	18.49
0.33	13.41	14.15	15.11	15.90	16.80	17.37	17.80	18.08	18.22	18.29
0.32	13.09	13.81	14.75	15.54	16.44	17.05	17.51	17.81	17.99	18.05
0.31	12.77	13.47	14.38	15.18	16.08	16.71	17.20	17.53	17.74	17.81
0.3	12.36	13.03	13.90	14.71	15.63	16.31	16.85	17.23	17.48	17.56
0.29	11.96	12.60	13.44	14.27	15.20	15.93	16.52	16.93	17.22	17.30
0.28	11.70	12.29	13.08	13.89	14.81	15.55	16.15	16.59	16.91	17.01
0.27	11.41	11.95	12.67	13.47	14.37	15.12	15.78	16.25	16.61	16.71
0.26	11.11	11.60	12.26	13.04	13.92	14.69	15.36	15.83	16.19	16.30
0.25	10.82	11.26	11.87	12.62	13.46	14.23	14.93	15.39	15.75	15.88
0.24	10.52	10.91	11.47	12.18	12.98	13.74	14.42	14.88	15.24	15.39
0.23	10.29	10.61	11.09	11.74	12.50	13.24	13.90	14.36	14.73	14.89
0.22	10.06	10.32	10.73	11.36	12.10	12.82	13.47	13.93	14.30	14.48
0.21	9.85	10.05	10.40	11.01	11.73	12.44	13.07	13.52	13.89	14.08
0.2	9.67	9.87	10.21	10.82	11.52	12.20	12.83	13.27	13.64	13.83
0.19	9.48	9.69	10.03	10.62	11.31	11.98	12.59	13.02	13.39	13.59
0.18	9.57	9.78	10.12	10.70	11.37	12.03	12.62	13.05	13.40	13.60

Table A4.7: long-term mean SSB over grid cells covered with computation. The results are the OM combining individuals OMs in the reference set.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	961754									
0.38					1102266					
0.37									1245990	
0.36			1072050							
0.35							1215761			1309441
0.34	1058797								1296265	1326810
0.33					1189063				1314210	1344934
0.32								1302085	1333049	
0.31			1174818					1322332		
0.3							1312484	1343138		
0.29	1187731						1334686			
0.28					1301165	1328812	1358071			
0.27						1354463			1441931	
0.26			1312671		1355804					
0.25					1385804		1436822			
0.24	1361444			1397925						
0.23			1419420	1434036	1452357					
0.22	1446222	1451542	1460138							
0.21	1492399		1503662							
0.2										
0.19	1592881								1692795	
0.18										

Table A4.8: long-term mean SSB over all combination of Btrigger-Ftarget using interpolation. The results are the OM combining individuals OMs in the reference set. Cells that coincide with the non-empty cells in the preceding table are actual values and are not interpolated.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	961754	988485	1016328	1047940	1081486	1112683	1144804	1176510	1207881	1237810
0.38	983263	1009633	1037190	1068708	1102266	1133387	1165495	1197193	1228566	1258462
0.37	1002204	1027956	1054952	1086079	1119324	1150483	1182692	1214497	1245990	1275972
0.36	1020637	1045694	1072050	1102708	1135563	1166687	1198921	1230751	1262275	1292400
0.35	1039684	1064062	1089803	1119971	1152419	1183505	1215761	1247616	1279171	1309441
0.34	1058797	1082465	1107564	1137221	1169247	1200288	1232563	1264558	1296265	1326810
0.33	1082871	1105332	1129294	1157953	1189063	1219555	1251328	1282920	1314210	1344934
0.32	1107592	1128813	1151615	1179267	1209461	1239469	1270813	1302085	1333049	1362413
0.31	1133220	1153183	1174818	1201473	1230776	1260345	1291316	1322332	1353011	1380962
0.3	1159988	1178686	1199162	1224688	1252975	1281996	1312484	1343138	1373526	1400028
0.29	1187731	1205139	1224443	1248840	1276123	1304634	1334686	1364979	1395058	1420055
0.28	1218883	1234606	1252357	1275252	1301165	1328812	1358071	1387975	1417720	1441148
0.27	1251673	1265685	1281875	1303281	1327858	1354463	1383087	1412559	1441931	1463706
0.26	1285833	1298079	1312671	1332567	1355804	1381386	1409013	1437689	1466313	1486403
0.25	1322554	1333047	1346081	1364178	1385804	1410272	1436822	1464634	1492447	1510752
0.24	1361444	1370156	1381627	1397925	1417968	1441380	1466927	1493788	1520711	1537109
0.23	1402381	1409399	1419420	1434036	1452357	1474613	1499066	1524896	1550854	1565227
0.22	1446222	1451542	1460138	1473253	1490175	1511119	1534339	1559009	1583886	1596060
0.21	1492399	1496315	1503662	1515117	1530498	1550006	1571880	1595290	1618993	1628823
0.2	1541752	1543566	1548982	1558659	1572395	1590375	1610822	1632902	1655371	1662745
0.19	1592881	1592455	1595816	1603608	1615605	1631974	1650925	1671613	1692795	1697607
0.18	1585910	1585497	1588760	1596325	1607972	1623860	1642248	1662316	1682857	1687523

Table A4.9: long-term mean fishing pressure on adults over grid cells covered with computation. The results are the OM combining individuals OMs in the reference set.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	0.337									
0.38					0.292					
0.37									0.252	
0.36			0.301							
0.35							0.260			0.236
0.34	0.306								0.240	0.233
0.33					0.267				0.236	0.228
0.32								0.239	0.231	
0.31			0.272					0.234		
0.3							0.237	0.229		
0.29	0.269						0.231			
0.28					0.240	0.233	0.226			
0.27						0.227			0.208	
0.26			0.237		0.227					
0.25					0.221		0.210			
0.24	0.226			0.218						
0.23			0.214	0.211	0.207					
0.22	0.208	0.207	0.206							
0.21	0.199		0.197							
0.2										
0.19	0.181								0.163	
0.18										

Table A4.10: long-term mean fishing pressure on adults over all combination of Btrigger-Ftarget using interpolation. The results are the OM combining individuals OMs in the reference set. Cells that coincide with the non-empty cells in the preceding table are actual values and are not interpolated.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	0.337	0.328	0.318	0.307	0.297	0.287	0.277	0.268	0.260	0.252
0.38	0.330	0.321	0.312	0.302	0.292	0.282	0.273	0.264	0.256	0.248
0.37	0.324	0.315	0.307	0.297	0.287	0.278	0.269	0.260	0.252	0.244
0.36	0.318	0.309	0.301	0.292	0.282	0.273	0.264	0.256	0.248	0.240
0.35	0.312	0.304	0.296	0.287	0.277	0.269	0.260	0.252	0.244	0.236
0.34	0.306	0.298	0.291	0.282	0.273	0.264	0.256	0.248	0.240	0.233
0.33	0.298	0.291	0.284	0.276	0.267	0.259	0.251	0.243	0.236	0.228
0.32	0.291	0.284	0.278	0.270	0.262	0.254	0.246	0.239	0.231	0.225
0.31	0.283	0.278	0.272	0.265	0.257	0.249	0.241	0.234	0.227	0.221
0.3	0.276	0.271	0.265	0.259	0.251	0.244	0.237	0.229	0.222	0.217
0.29	0.269	0.264	0.259	0.253	0.246	0.239	0.231	0.224	0.218	0.213
0.28	0.260	0.256	0.252	0.246	0.240	0.233	0.226	0.220	0.213	0.208
0.27	0.252	0.249	0.245	0.239	0.234	0.227	0.221	0.214	0.208	0.204
0.26	0.244	0.241	0.237	0.233	0.227	0.221	0.215	0.209	0.203	0.200
0.25	0.235	0.233	0.230	0.226	0.221	0.215	0.210	0.204	0.198	0.195
0.24	0.226	0.225	0.222	0.218	0.214	0.209	0.204	0.198	0.193	0.190
0.23	0.217	0.216	0.214	0.211	0.207	0.202	0.197	0.192	0.187	0.185
0.22	0.208	0.207	0.206	0.203	0.200	0.195	0.191	0.186	0.182	0.180
0.21	0.199	0.198	0.197	0.195	0.192	0.188	0.184	0.180	0.176	0.174
0.2	0.190	0.189	0.188	0.187	0.184	0.181	0.177	0.173	0.169	0.169
0.19	0.181	0.181	0.180	0.179	0.176	0.174	0.170	0.167	0.163	0.163
0.18	0.182	0.182	0.181	0.180	0.177	0.175	0.172	0.168	0.165	0.164

Table A4.11 long-term mean fishing pressure on juveniles over grid cells covered with computation. The results are the OM combining individuals OMs in the reference set.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	0.090									
0.38					0.078					
0.37									0.067	
0.36			0.080							
0.35							0.069			0.063
0.34	0.082								0.064	0.062
0.33					0.072				0.063	0.061
0.32								0.064	0.062	
0.31			0.073					0.063		
0.3							0.063	0.061		
0.29	0.072						0.062			
0.28					0.064	0.062	0.061			
0.27						0.061			0.056	
0.26			0.064		0.061					
0.25					0.059		0.056			
0.24	0.061			0.059						
0.23			0.057	0.057	0.055					
0.22	0.056	0.056	0.055							
0.21	0.054		0.053							
0.2										
0.19	0.049								0.044	
0.18										

Table A4.12: long-term mean fishing pressure on juveniles over all combination of Btrigger-Ftarget using interpolation. The results are the OM combining individuals OMs in the reference set. Cells that coincide with the non-empty cells in the preceding table are actual values and are not interpolated.

Ftarget-Btrigger	8e+05	9e+05	1e+06	1100000	1200000	1300000	1400000	1500000	1600000	1700000
0.39	0.090	0.088	0.085	0.082	0.079	0.077	0.074	0.072	0.069	0.067
0.38	0.088	0.086	0.083	0.081	0.078	0.075	0.073	0.071	0.068	0.066
0.37	0.087	0.084	0.082	0.079	0.077	0.074	0.072	0.069	0.067	0.065
0.36	0.085	0.083	0.080	0.078	0.075	0.073	0.071	0.068	0.066	0.064
0.35	0.083	0.081	0.079	0.077	0.074	0.072	0.069	0.067	0.065	0.063
0.34	0.082	0.080	0.078	0.075	0.073	0.071	0.068	0.066	0.064	0.062
0.33	0.080	0.078	0.076	0.074	0.072	0.069	0.067	0.065	0.063	0.061
0.32	0.078	0.076	0.074	0.072	0.070	0.068	0.066	0.064	0.062	0.060
0.31	0.076	0.074	0.073	0.071	0.069	0.067	0.065	0.063	0.061	0.059
0.3	0.074	0.073	0.071	0.069	0.067	0.065	0.063	0.061	0.059	0.058
0.29	0.072	0.071	0.069	0.068	0.066	0.064	0.062	0.060	0.058	0.057
0.28	0.070	0.069	0.067	0.066	0.064	0.062	0.061	0.059	0.057	0.056
0.27	0.068	0.067	0.066	0.064	0.063	0.061	0.059	0.057	0.056	0.055
0.26	0.065	0.065	0.064	0.062	0.061	0.059	0.058	0.056	0.054	0.053
0.25	0.063	0.062	0.062	0.060	0.059	0.058	0.056	0.055	0.053	0.052
0.24	0.061	0.060	0.060	0.059	0.057	0.056	0.055	0.053	0.052	0.051
0.23	0.058	0.058	0.057	0.057	0.055	0.054	0.053	0.052	0.050	0.050
0.22	0.056	0.056	0.055	0.054	0.053	0.052	0.051	0.050	0.049	0.048
0.21	0.054	0.053	0.053	0.052	0.051	0.050	0.049	0.048	0.047	0.047
0.2	0.051	0.051	0.051	0.050	0.049	0.048	0.047	0.046	0.045	0.045
0.19	0.049	0.049	0.048	0.048	0.047	0.047	0.046	0.045	0.044	0.044
0.18	0.049	0.049	0.049	0.048	0.048	0.047	0.046	0.045	0.044	0.044

Annex 5: Stock assessment model configuration

```
An object of class "FLSAM.control"
Slot "name":
[1] "North Sea Herring"

Slot "desc":
[1] "Imported from a VPA file. ( ./bootstrap/data/index.txt ).  Tue Apr  9 20:26:01 2024"

Slot "range":
     min       max plusgroup   minyear   maxyear   minfbar   maxfbar
       0         8         8      1947      2024         2         6

Slot "fleets":
catch unique       HERAS     IBTS-Q1       IBTS0     IBTS-Q3
           0           2           2           2           2

Slot "plus.group":
[1] 1 1 0 0 0

Slot "states":
             age
fleet          0  1  2  3  4  5  6  7  8
  catch unique  0  1  2  3  4  5  6  7  7
  HERAS        -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q1      -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS0        -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q3      -1 -1 -1 -1 -1 -1 -1 -1 -1

Slot "logN.vars":
0 1 2 3 4 5 6 7 8
0 1 1 1 1 1 1 1 1

Slot "logP.vars":
numeric(0)

Slot "catchabilities":
             age
fleet          0  1  2  3  4  5  6  7  8
  catch unique -1 -1 -1 -1 -1 -1 -1 -1 -1
  HERAS        -1  1  1  2  2  2  2  2  2
  IBTS-Q1      -1  3 -1 -1 -1 -1 -1 -1 -1
  IBTS0         0 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q3       4  5  6  7  8  9 -1 -1 -1

Slot "power.law.exps":
             age
fleet          0  1  2  3  4  5  6  7  8
  catch unique -1 -1 -1 -1 -1 -1 -1 -1 -1
```

```
  HERAS        -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q1      -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS0        -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q3      -1 -1 -1 -1 -1 -1 -1 -1 -1

Slot "f.vars":
              age
fleet          0  1  2  3  4  5  6  7  8
  catch unique  0  0  1  1  1  1  2  2  2
  HERAS        -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q1      -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS0        -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q3      -1 -1 -1 -1 -1 -1 -1 -1 -1

Slot "obs.vars":
              age
fleet          0  1  2  3  4  5  6  7  8
  catch unique  0  0  1  1  1  1  1  2  2
  HERAS        -1  3  4  5  6  6  6  7  7
  IBTS-Q1      -1  8 -1 -1 -1 -1 -1 -1 -1
  IBTS0         9 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q3      10 10 11 11 11 11 -1 -1 -1

Slot "srr":
[1] 0

Slot "scaleNoYears":
[1] 0

Slot "scaleYears":
[1] NA

Slot "scalePars":
     age
years  0 1 2 3 4 5 6 7 8

Slot "cor.F":
[1] 2

Slot "cor.obs":
              age
fleet          0-1 1-2 2-3 3-4 4-5 5-6 6-7 7-8
  catch unique  NA  NA  NA  NA  NA  NA  NA  NA
  HERAS         -1  NA  NA  NA  NA  NA  NA  NA
  IBTS-Q1       -1  -1  -1  -1  -1  -1  -1  -1
  IBTS0         -1  -1  -1  -1  -1  -1  -1  -1
  IBTS-Q3        0   0   0   0   0  -1  -1  -1

Slot "cor.obs.Flag":
[1] ID ID ID ID AR
Levels: ID AR US
```

```
Slot "biomassTreat":
[1] -1 -1 -1 -1 -1

Slot "timeout":
[1] 3600

Slot "likFlag":
[1] LN LN LN LN LN
Levels: LN ALN

Slot "fixVarToWeight":
[1] FALSE FALSE FALSE FALSE FALSE

Slot "fracMixF":
[1] 0

Slot "fracMixN":
[1] 0 0 0 0 0 0 0 0 0

Slot "fracMixObs":
catch unique       HERAS     IBTS-Q1       IBTS0     IBTS-Q3
           0           0           0           0           0

Slot "constRecBreaks":
numeric(0)

Slot "predVarObsLink":
              age
fleet          0  1  2  3  4  5  6  7  8
  catch unique -1 -1 -1 -1 -1 -1 -1 -1 -1
  HERAS        -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q1      -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS0        -1 -1 -1 -1 -1 -1 -1 -1 -1
  IBTS-Q3      -1 -1 -1 -1 -1 -1 -1 -1 -1

Slot "stockWeightModel":
[1] FALSE

Slot "stockWeightMean":
 0  1  2  3  4  5  6  7  8
NA NA NA NA NA NA NA NA NA

Slot "stockWeightObsVar":
 0  1  2  3  4  5  6  7  8
NA NA NA NA NA NA NA NA NA

Slot "catchWeightModel":
[1] FALSE

Slot "catchWeightMean":
```

```
 0  1  2  3  4  5  6  7  8
NA NA NA NA NA NA NA NA NA

Slot "catchWeightObsVar":
 0  1  2  3  4  5  6  7  8
NA NA NA NA NA NA NA NA NA

Slot "maturityModel":
[1] FALSE

Slot "maturityMean":
 0  1  2  3  4  5  6  7  8
NA NA NA NA NA NA NA NA NA

Slot "mortalityModel":
[1] FALSE

Slot "mortalityMean":
 0  1  2  3  4  5  6  7  8
NA NA NA NA NA NA NA NA NA

Slot "mortalityObsVar":
 0  1  2  3  4  5  6  7  8
NA NA NA NA NA NA NA NA NA

Slot "XtraSd":
     [,1] [,2] [,3] [,4]

Slot "logNMeanAssumption":
[1] 0 0

Slot "initState":
[1] 0

Slot "simulate":
[1] FALSE

Slot "residuals":
[1] TRUE

Slot "sumFleets":
logical(0)
```

Annex 6: Stock assessment model diagnostics

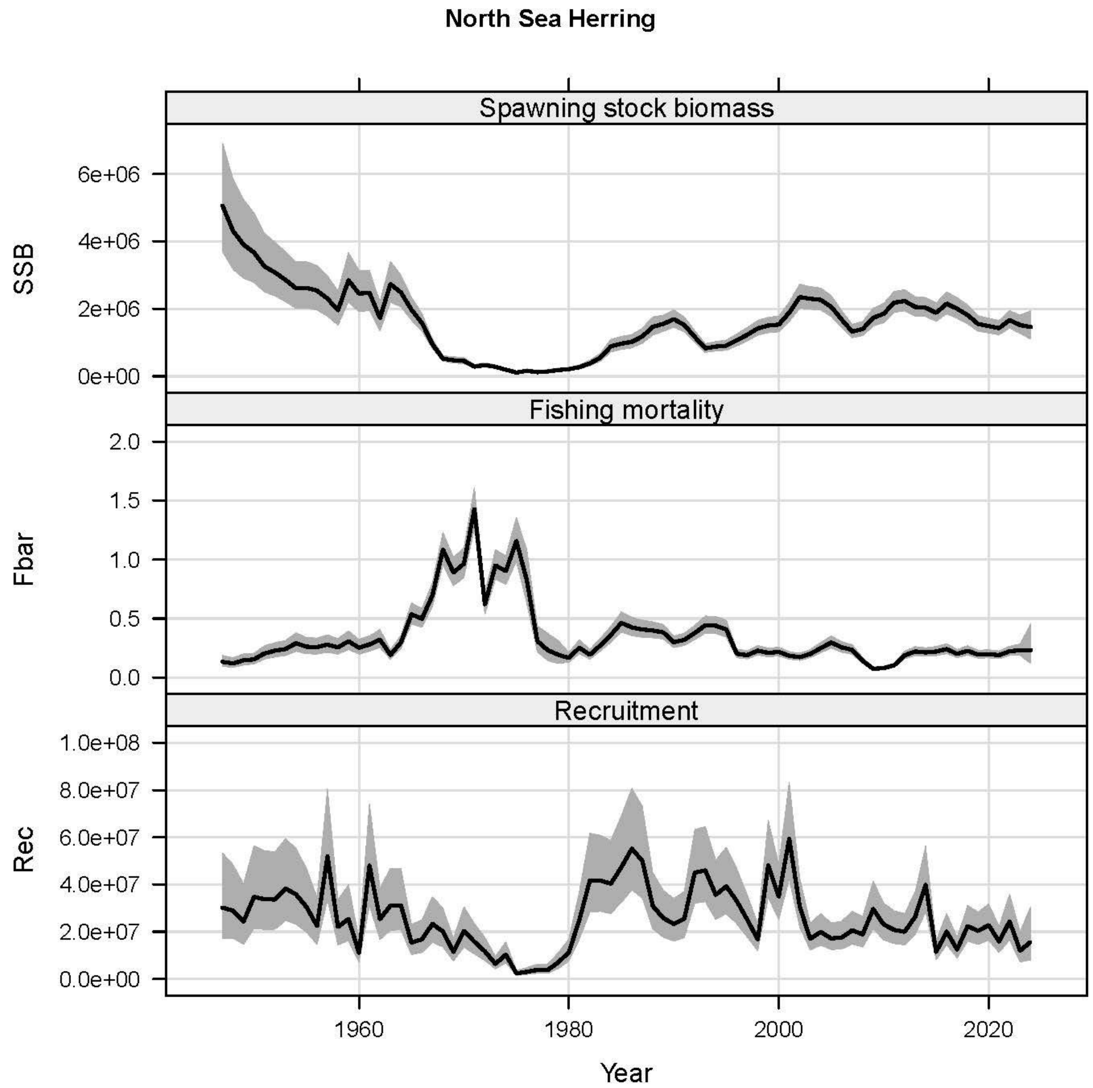


Figure A6.1. North Sea herring. Stock summary plot of North Sea herring with associated uncertainty for SSB (top panel), F ages 2–6 (middle panel) and recruitment (bottom panel).

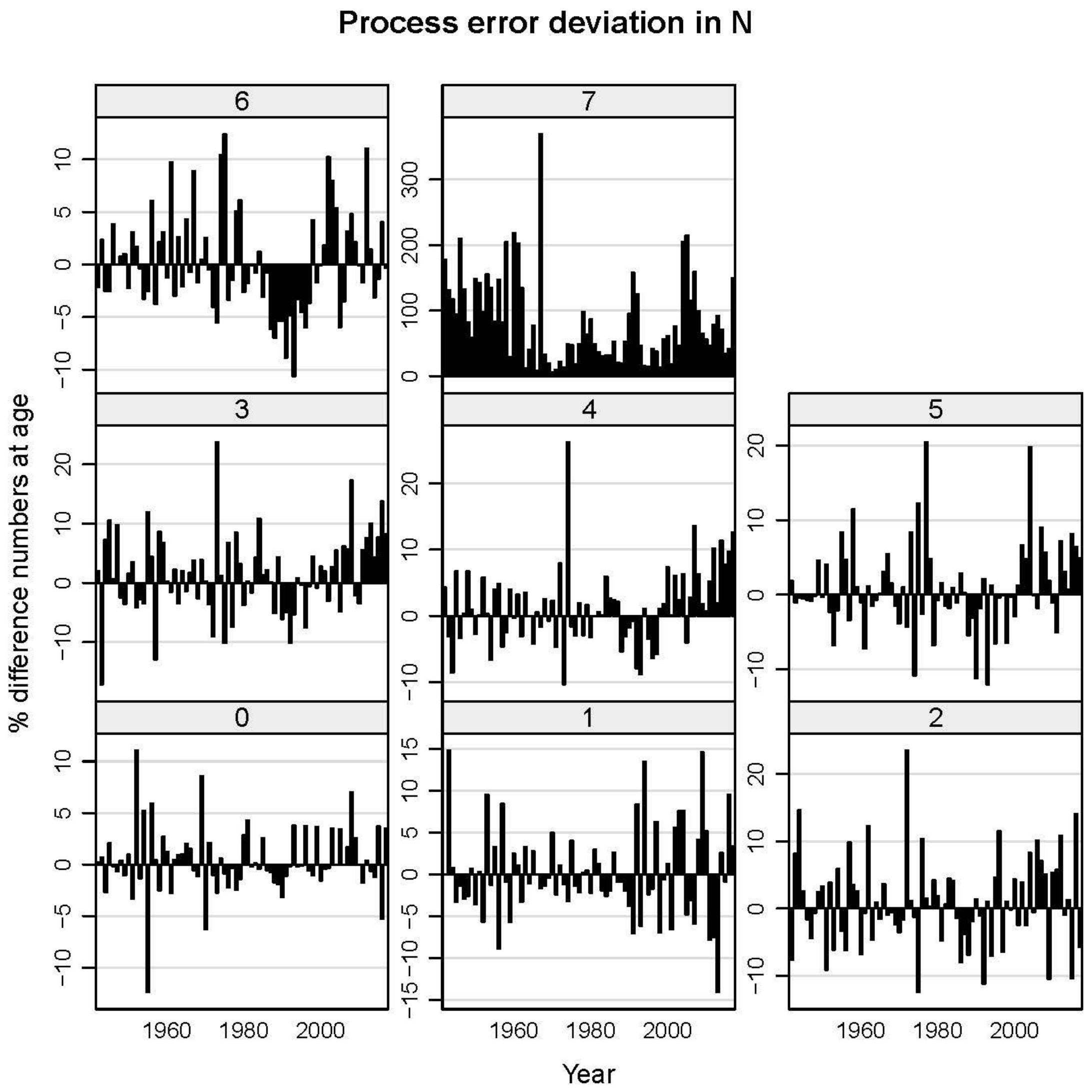


Figure A6.2. Yearly process error deviation in stock numbers at age.

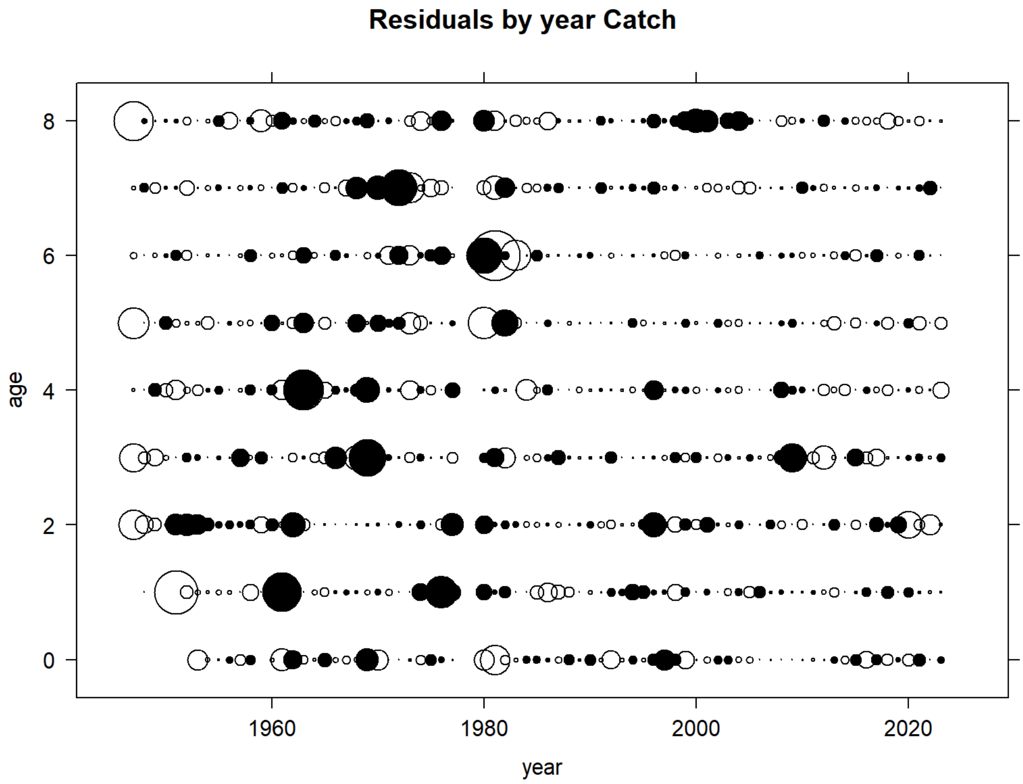


Figure A6.3. North Sea herring. Bubble plot of standardized catch residual at age.

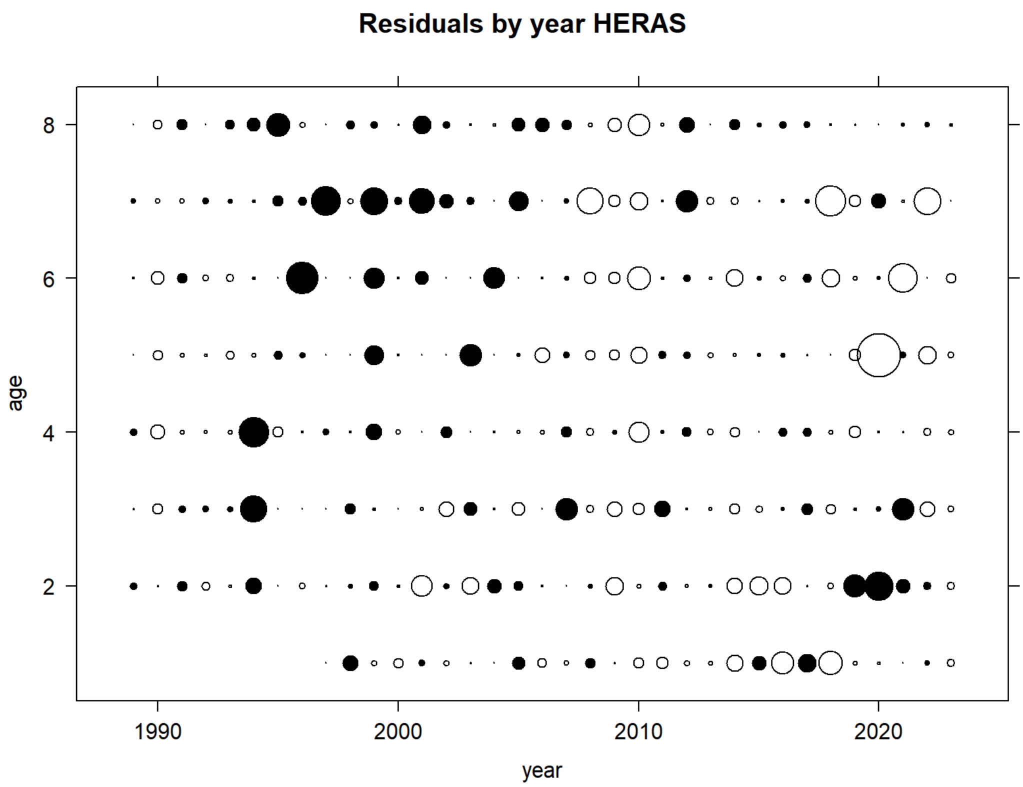


Figure A6.4. North Sea herring. Bubble plot of standardized acoustic survey residuals at age.

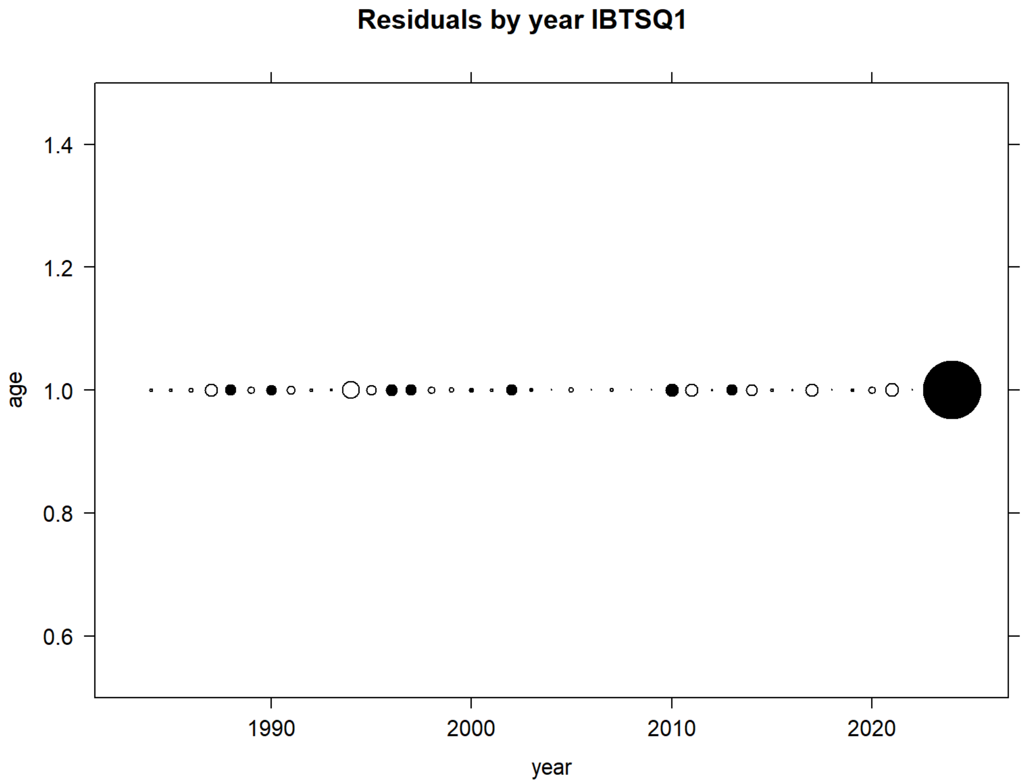


Figure A6.5. North Sea herring. Bubble plot of standardized IBTS-Q1 residuals at age.

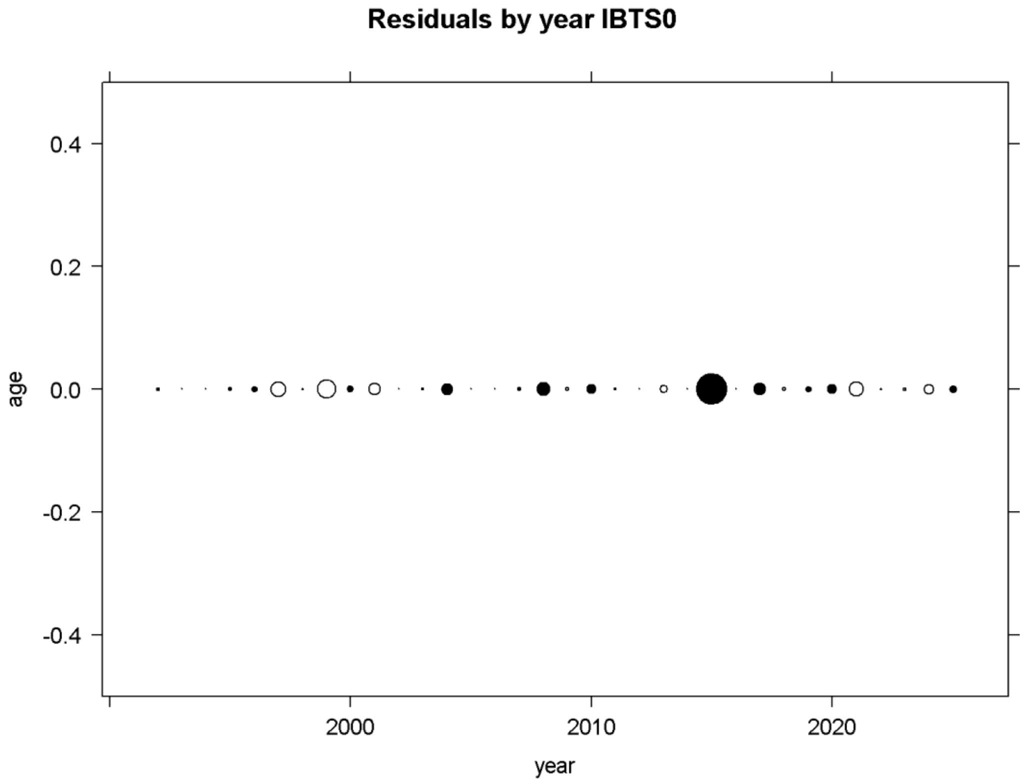


Figure A6.6. North Sea herring. Bubble plot of standardized IBTS0 residuals at age.

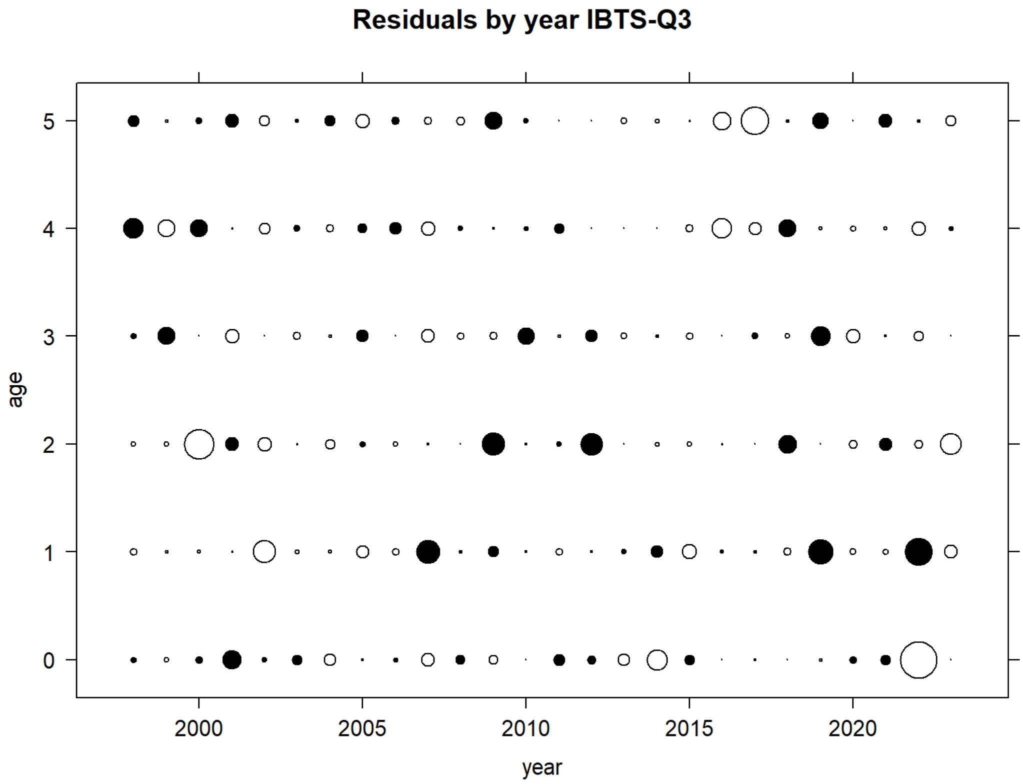


Figure A6.7. North Sea herring. Bubble plot of standardized IBTS-Q3 residuals at age.

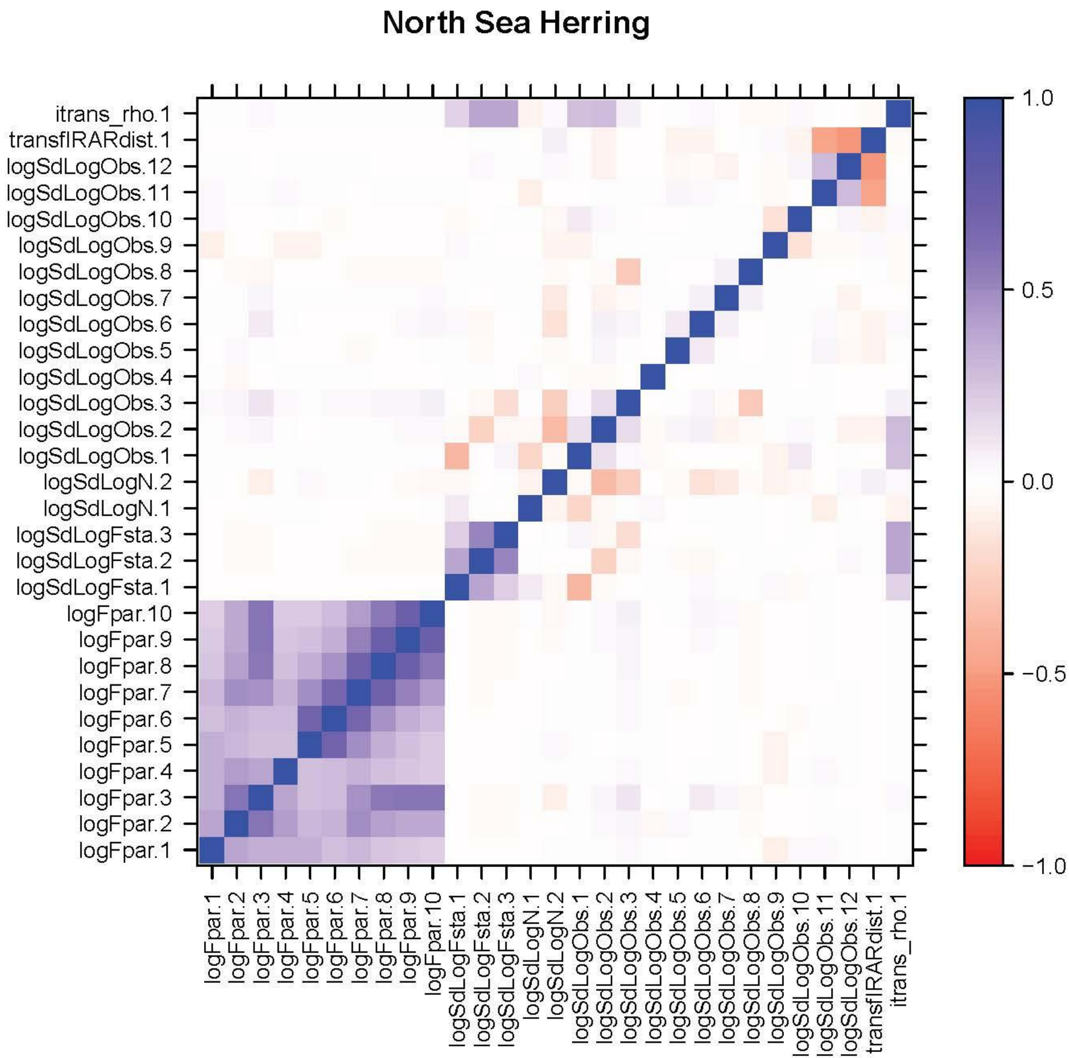


Figure A6.8. North Sea herring. Correlation plot of the FLSAM assessment model with the final set of parameters estimated in the model. The diagonal represents the correlation with the data source itself.

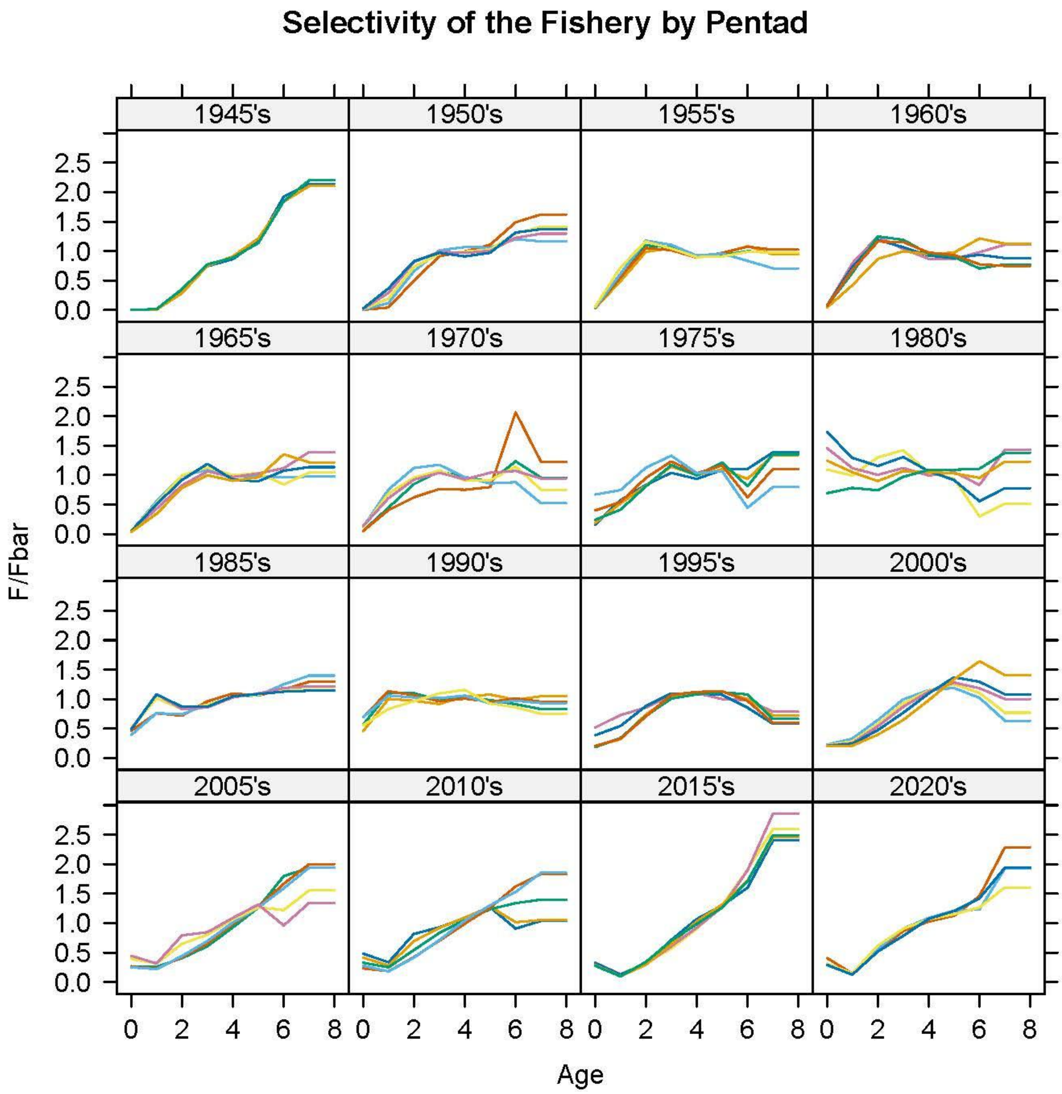


Figure A6.9. North Sea herring. Fishing selectivity by pentad.

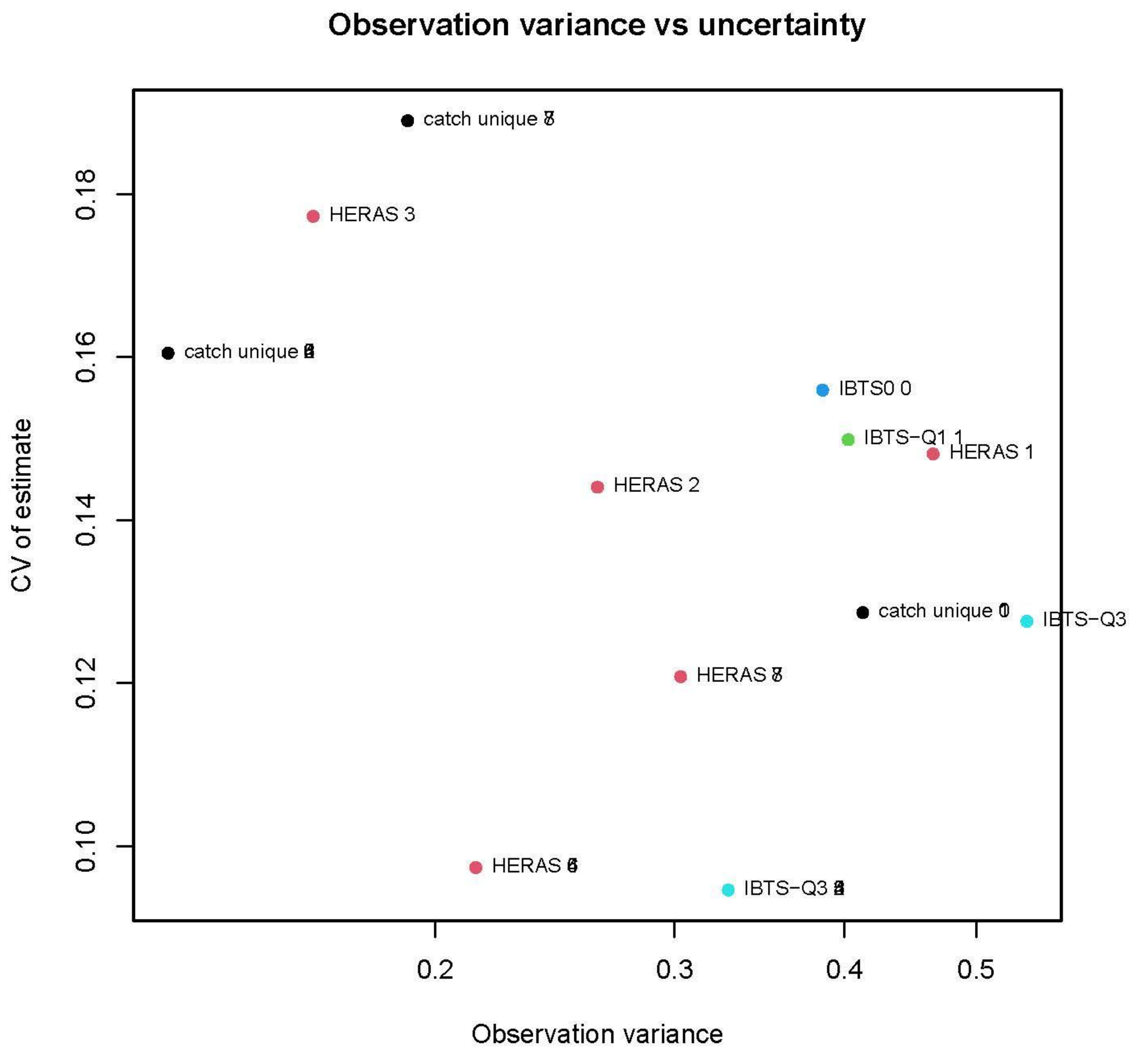


Figure A6.10. North Sea herring. Observation variance by data source as estimated by the assessment model plotted against the CV estimate of the observation variance parameter.

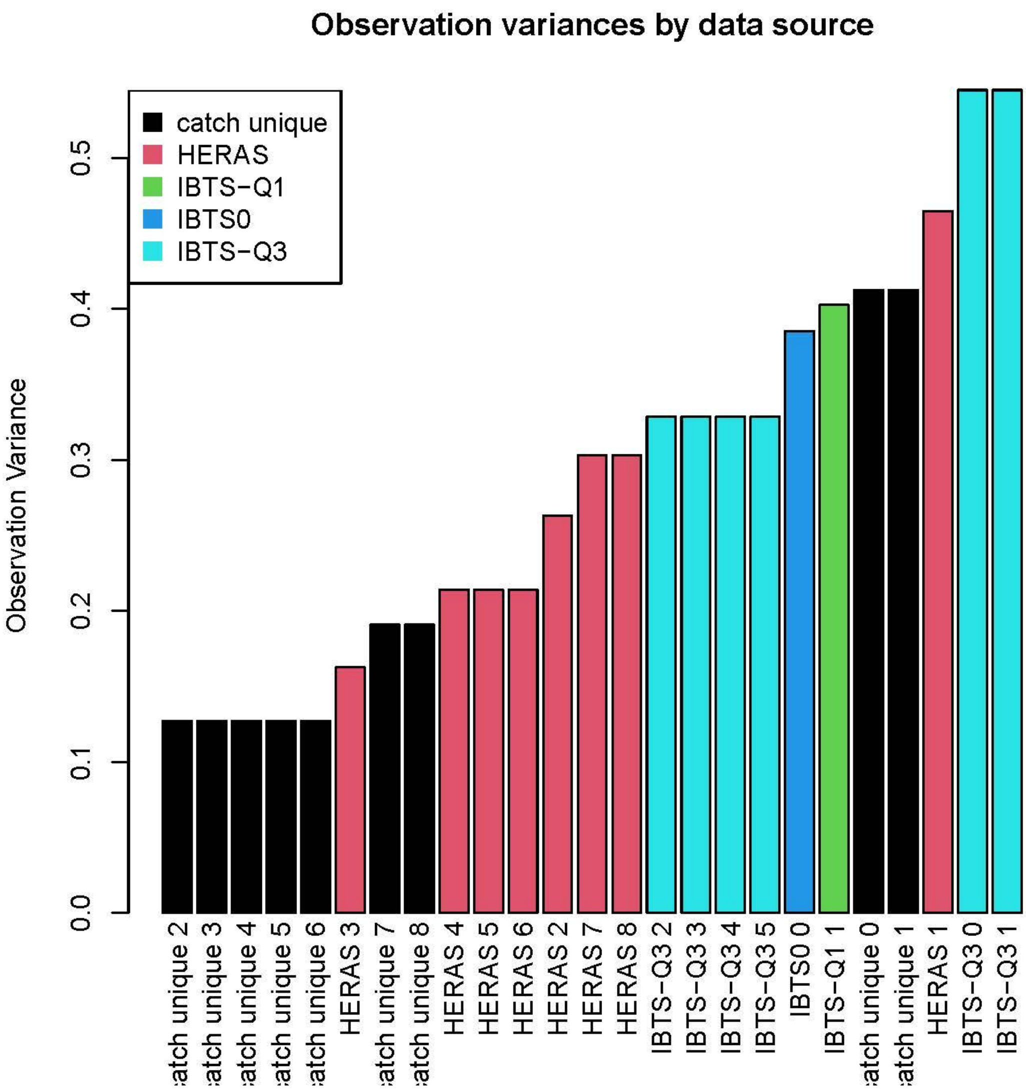


Figure A6.11. North Sea herring. Observation variance by data source as estimated by the assessment model. Observation variance is ordered from least (left) to most (right). Colours indicate the different data sources. Observation variance is not individually estimated for each data source thereby reducing the parameters needed to be estimated in the assessment model. In these cases of parameter bindings, observation variances have equal values.

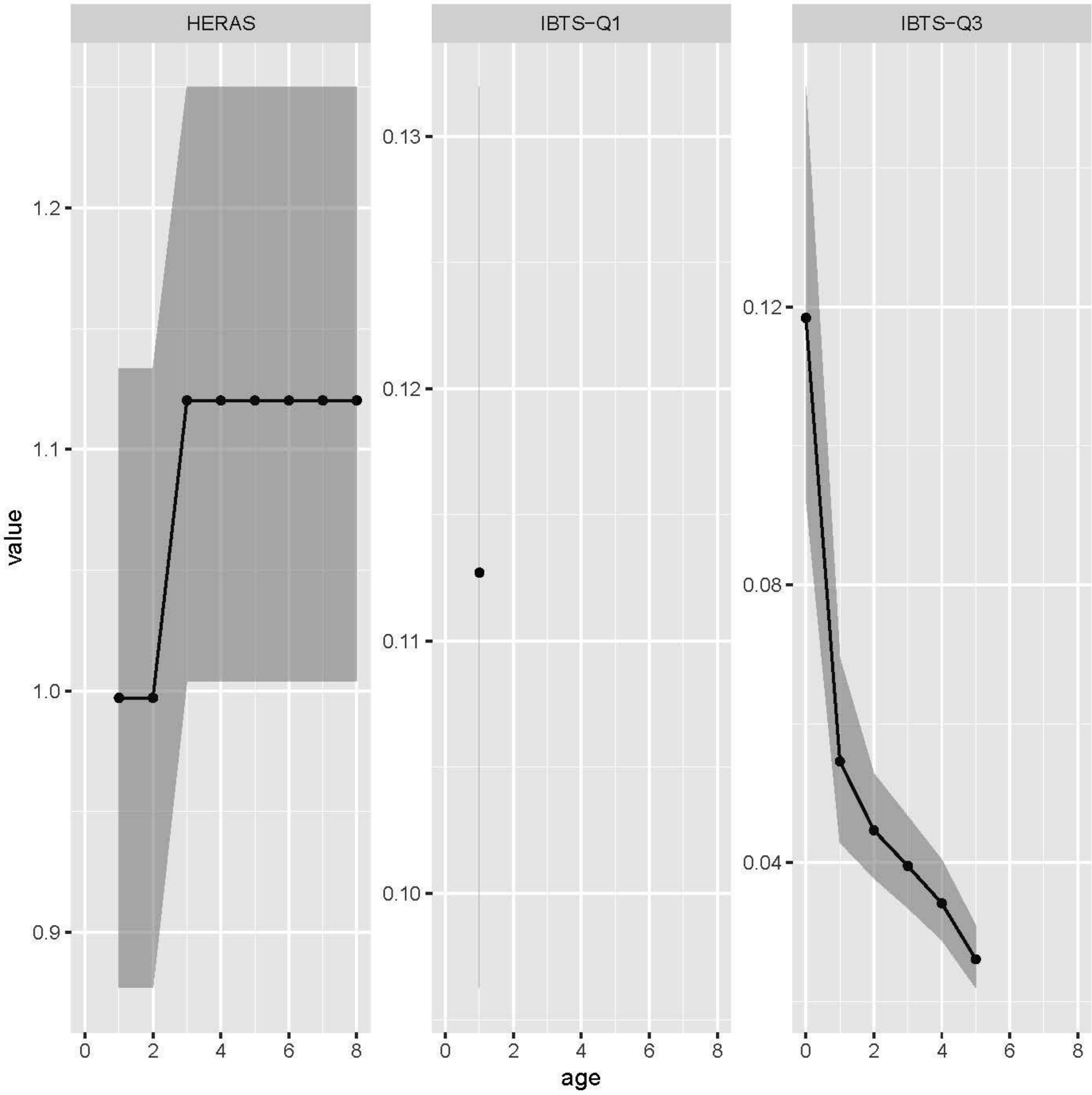


Figure A6.12. North Sea herring. Catchability at age for the HERAS, IBTSQ1 and IBTSQ3 surveys.

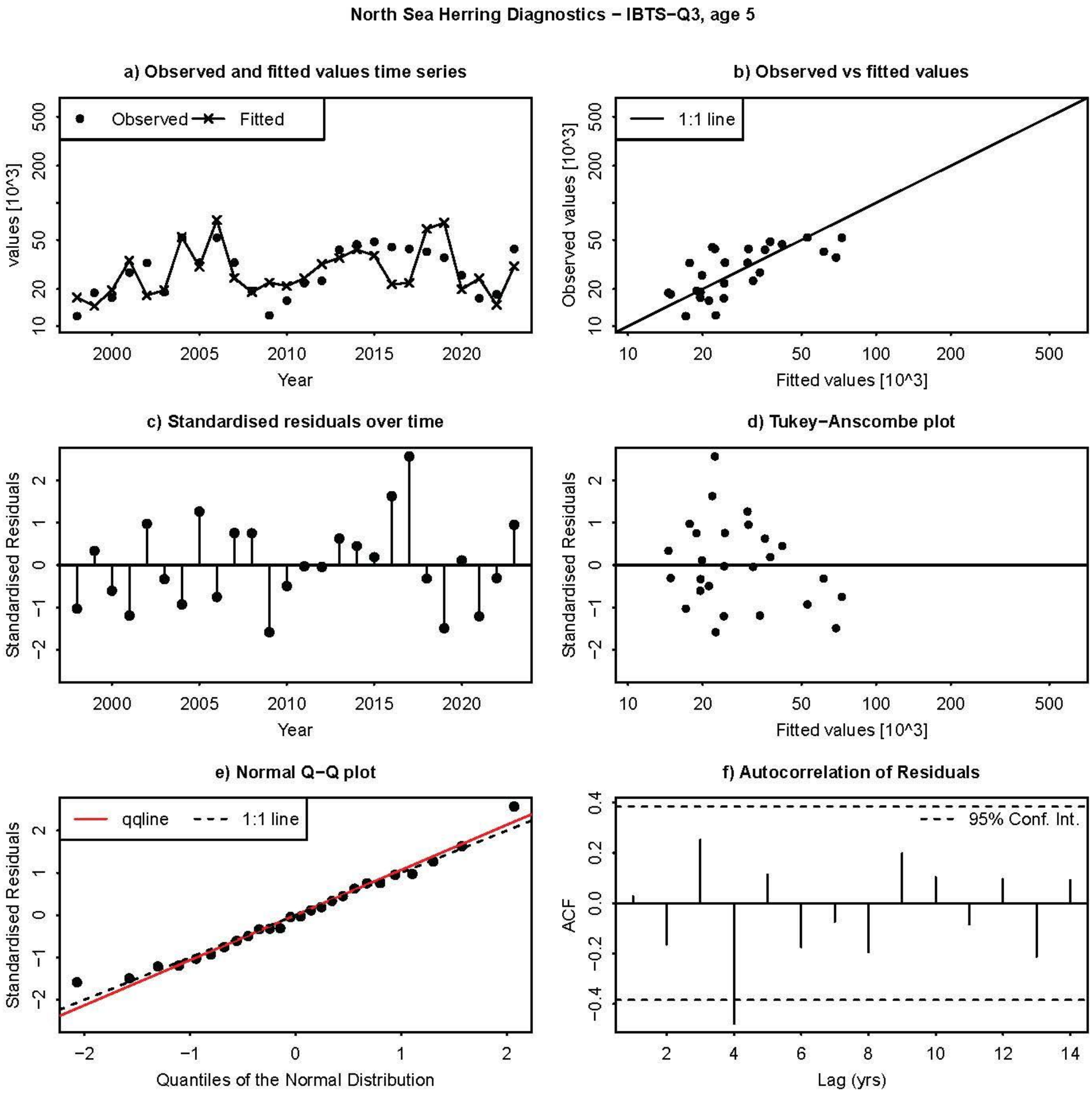


Figure A6.13. Assessment fit IBTS-Q3 age 5.

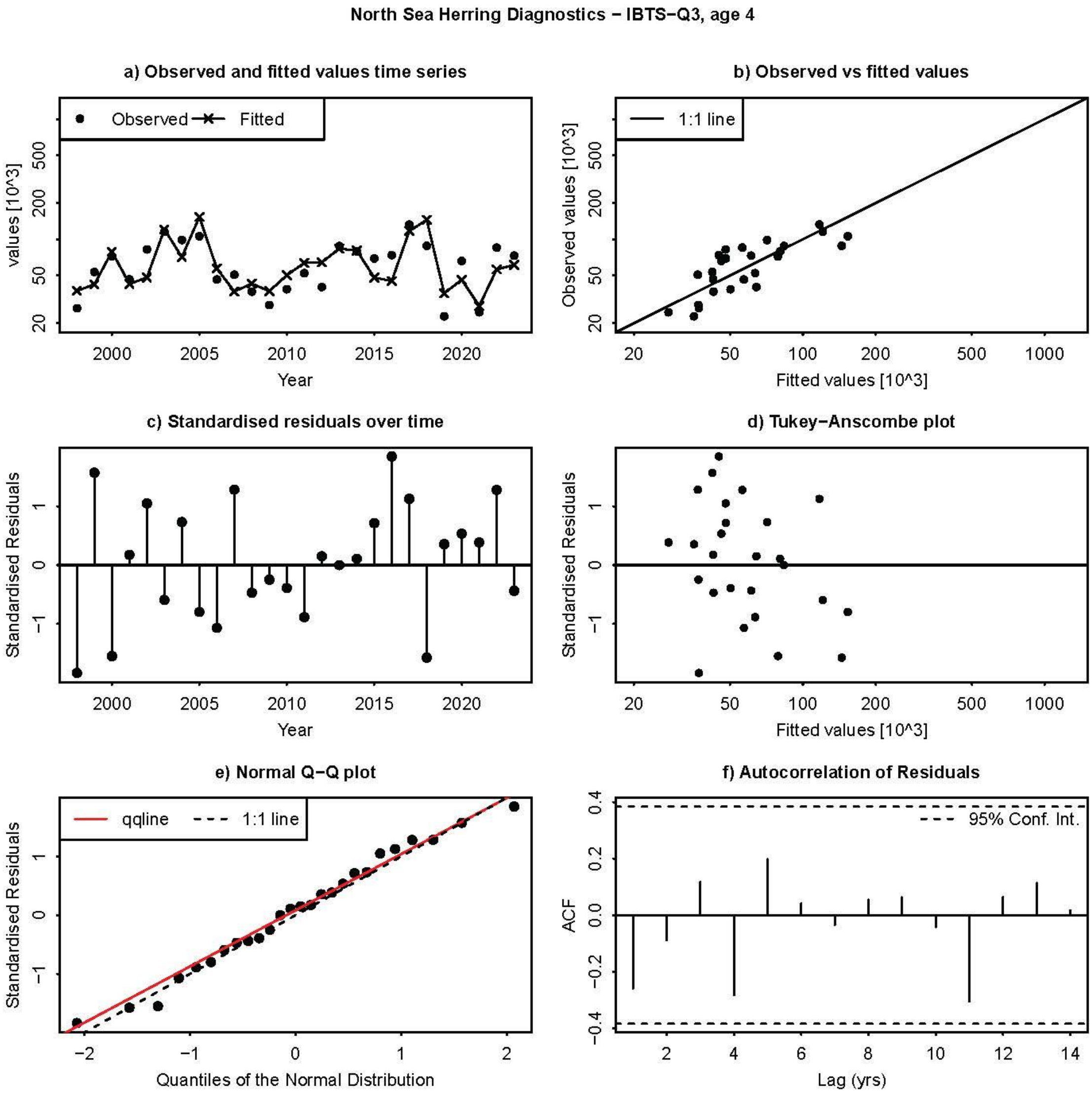


Figure A6.14. Assessment fit IBTS-Q3 age 4.

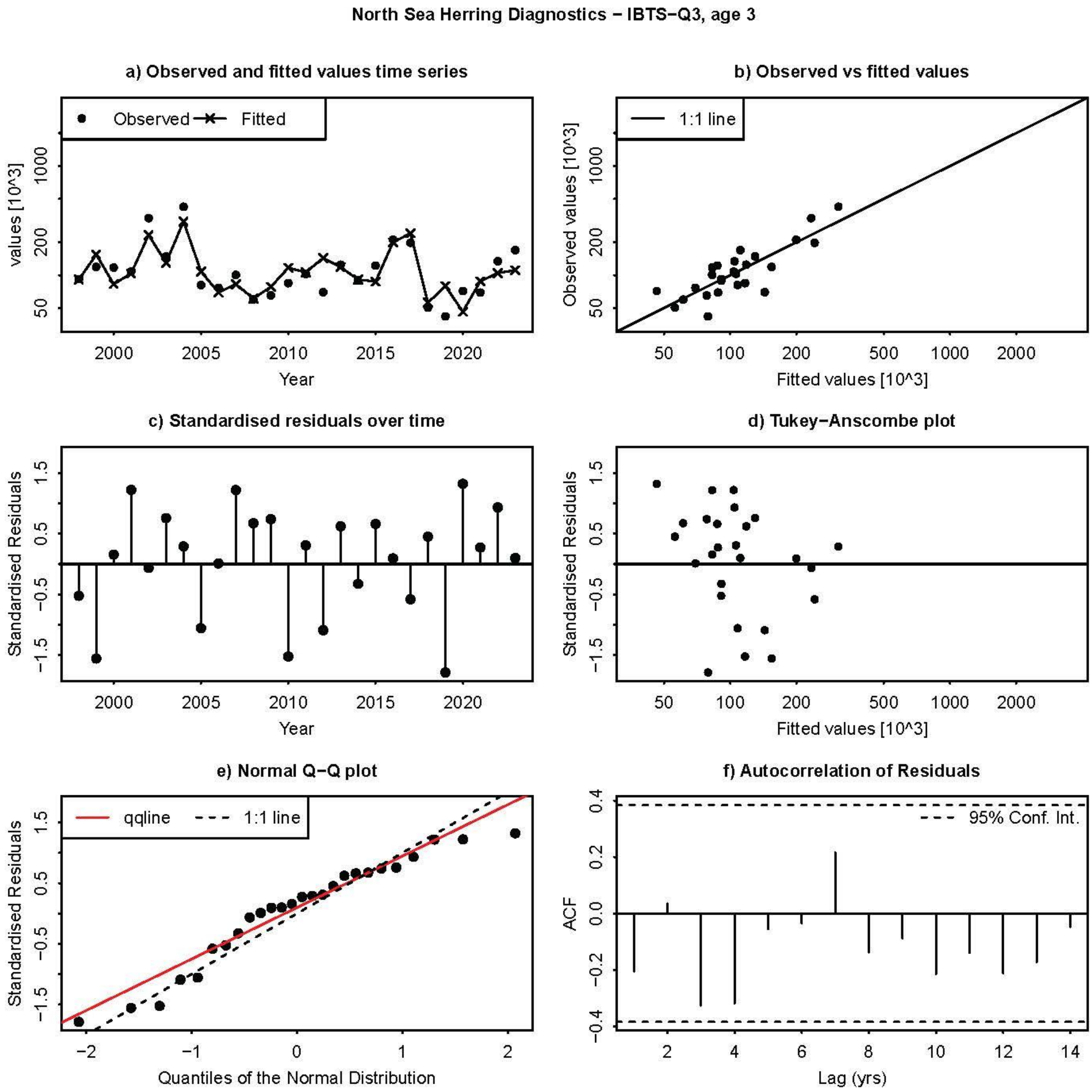


Figure A6.15. Assessment fit IBTS-Q3 age 3.

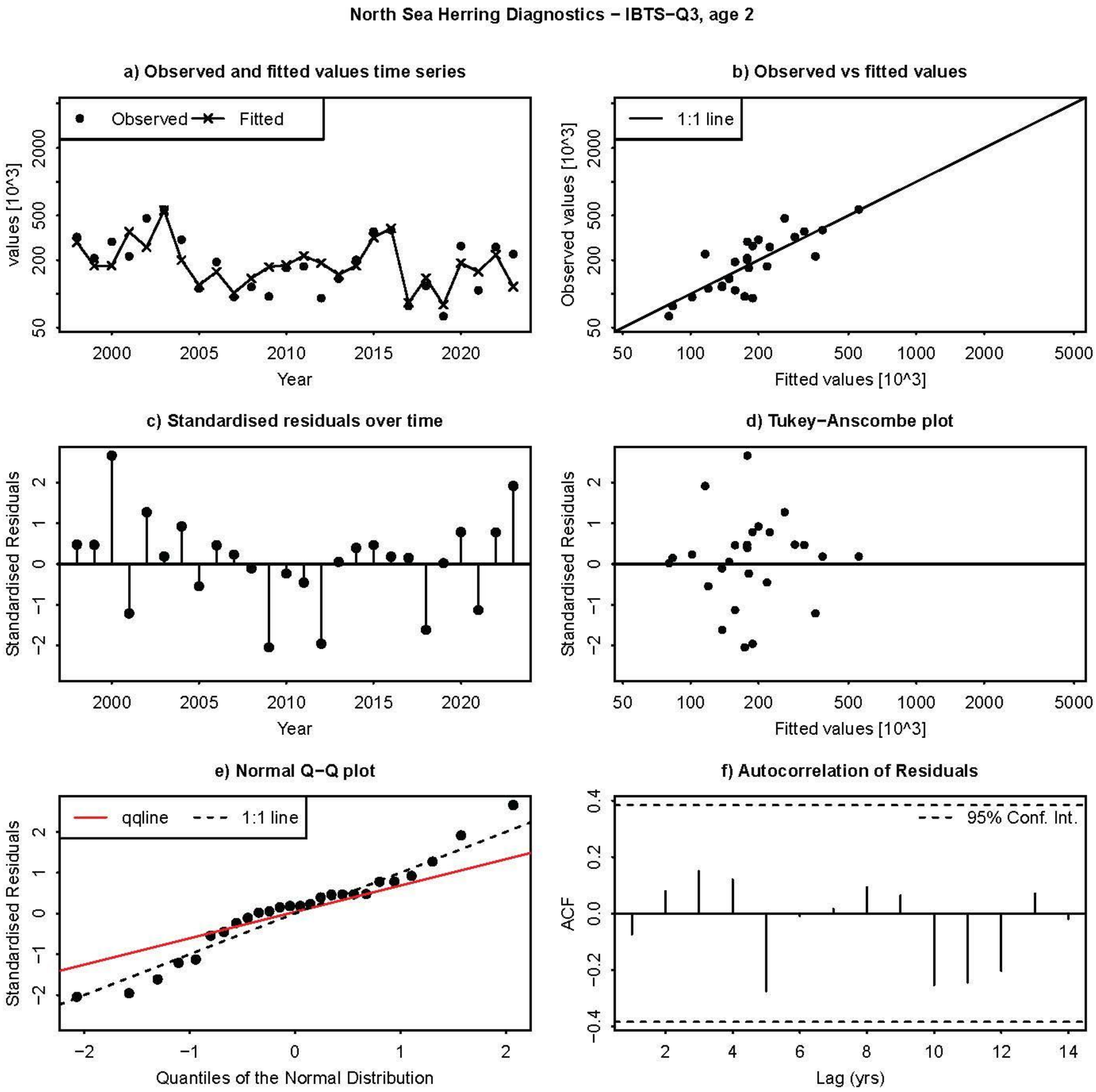


Figure A6.16. Assessment fit IBTS-Q3 age 2.

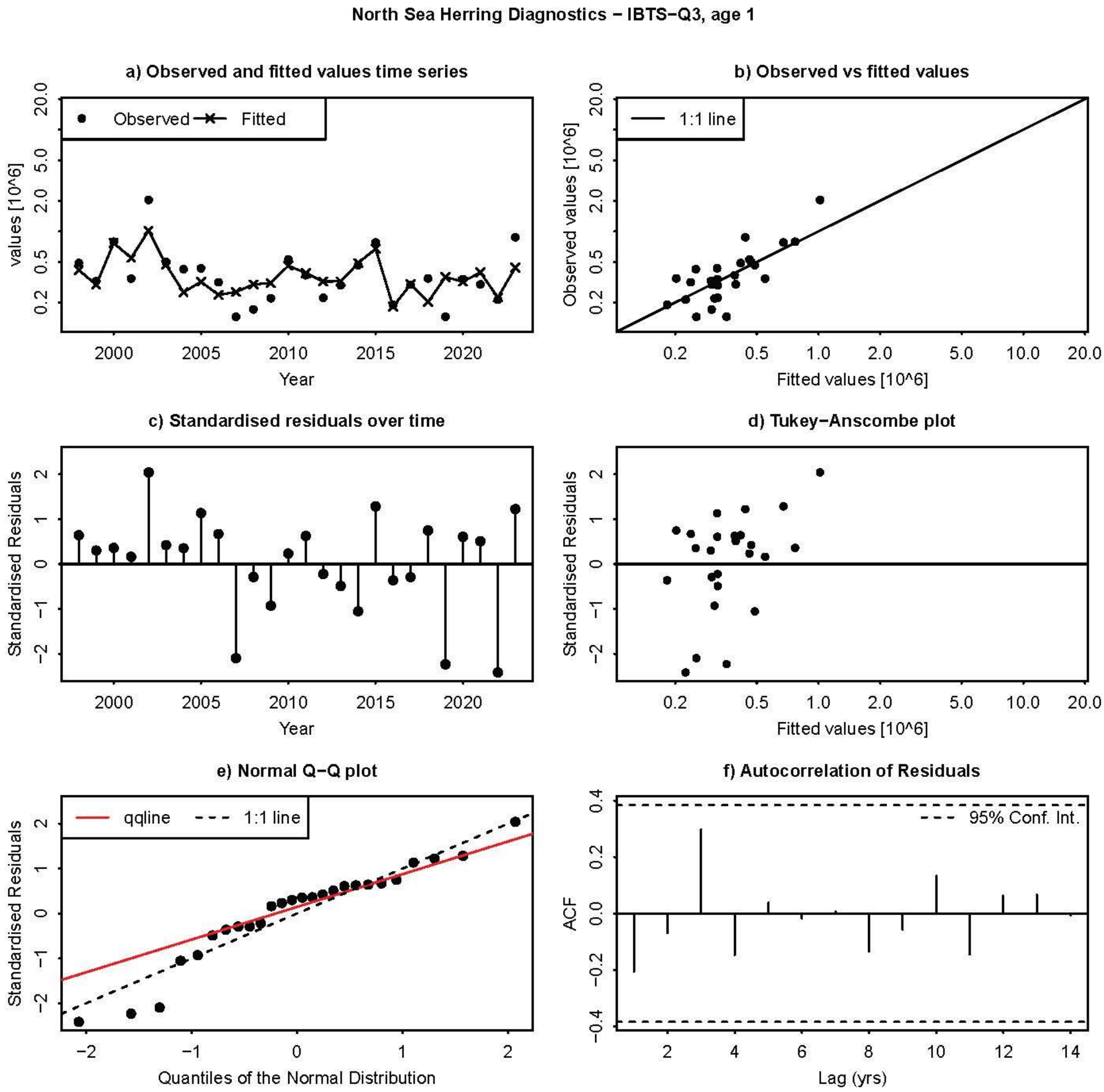


Figure A6.17. Assessment fit IBTS-Q3 age 1.

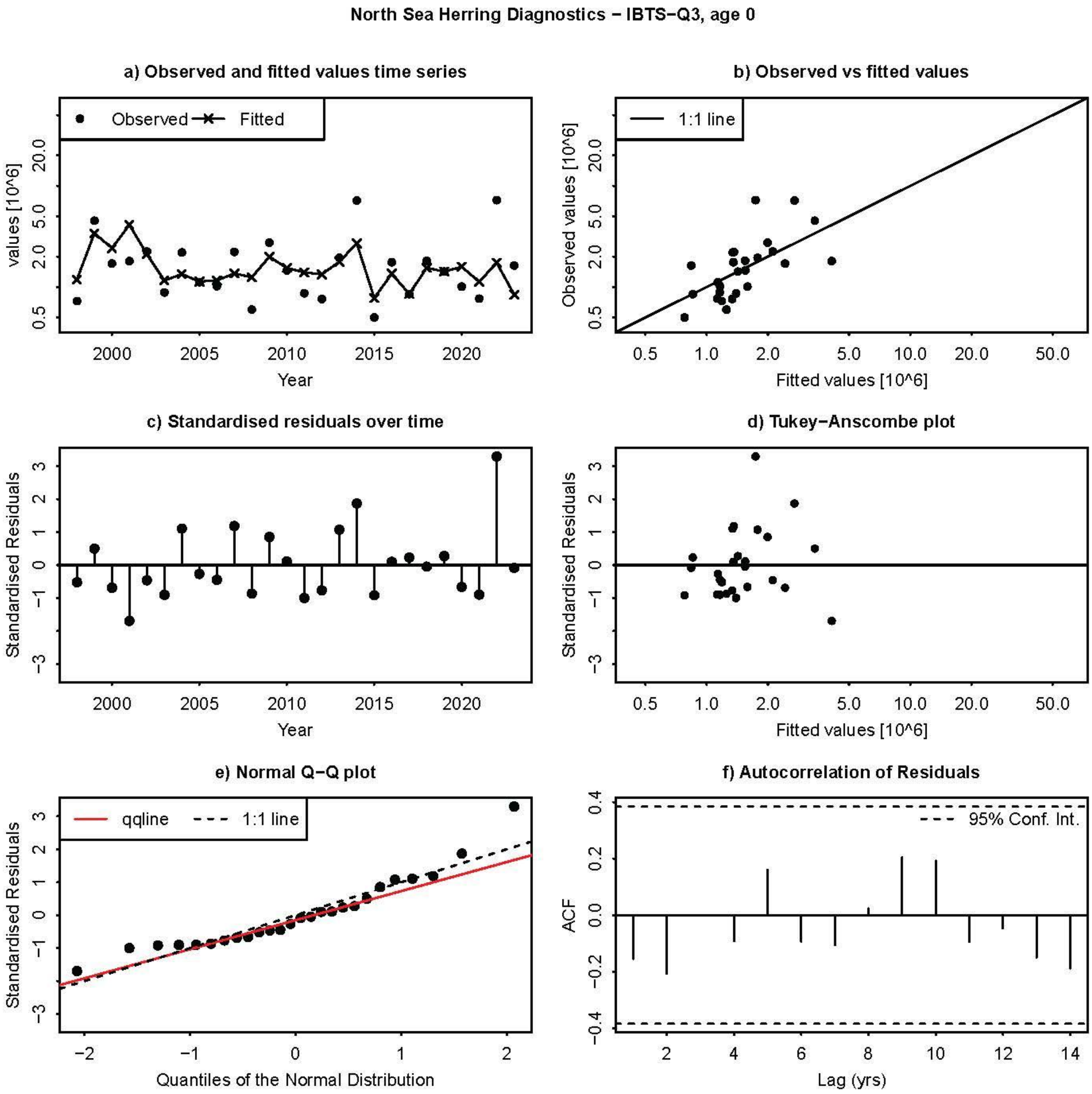


Figure A6.18. Assessment fit IBTS-Q3 age 0.

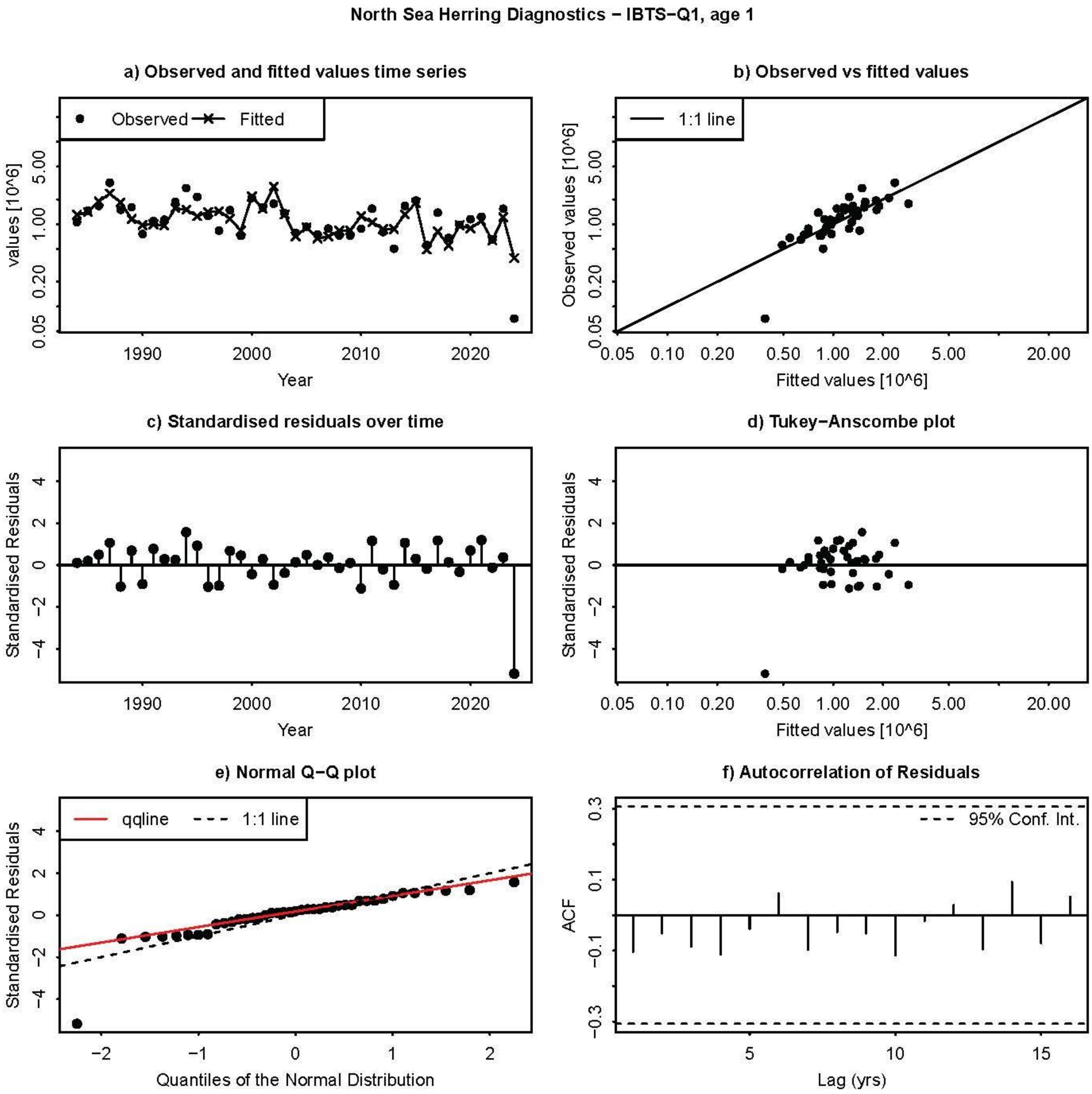


Figure A6.19. Assessment fit IBTS-Q1 age 1.

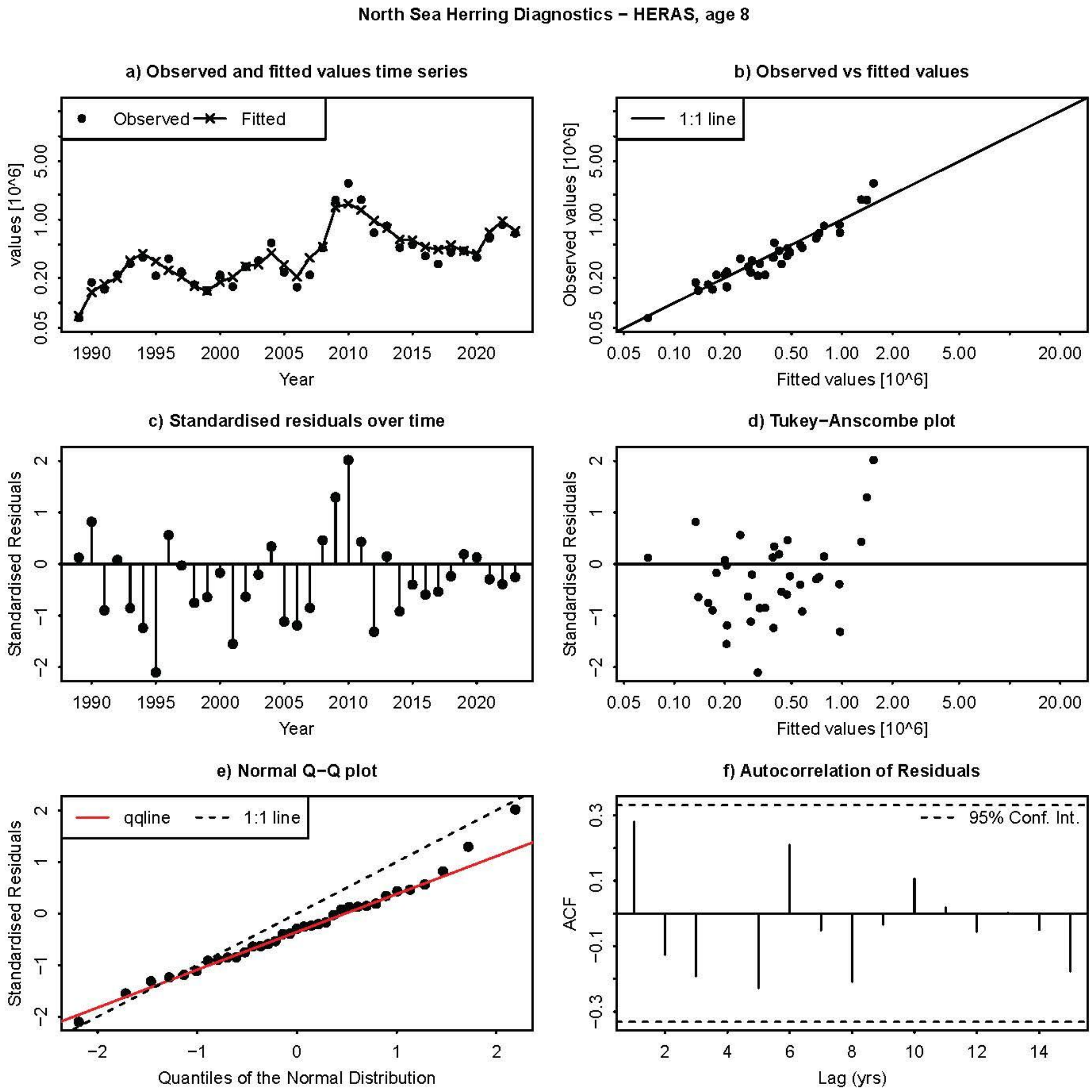


Figure A6.20. Assessment fit HERAS age 8.

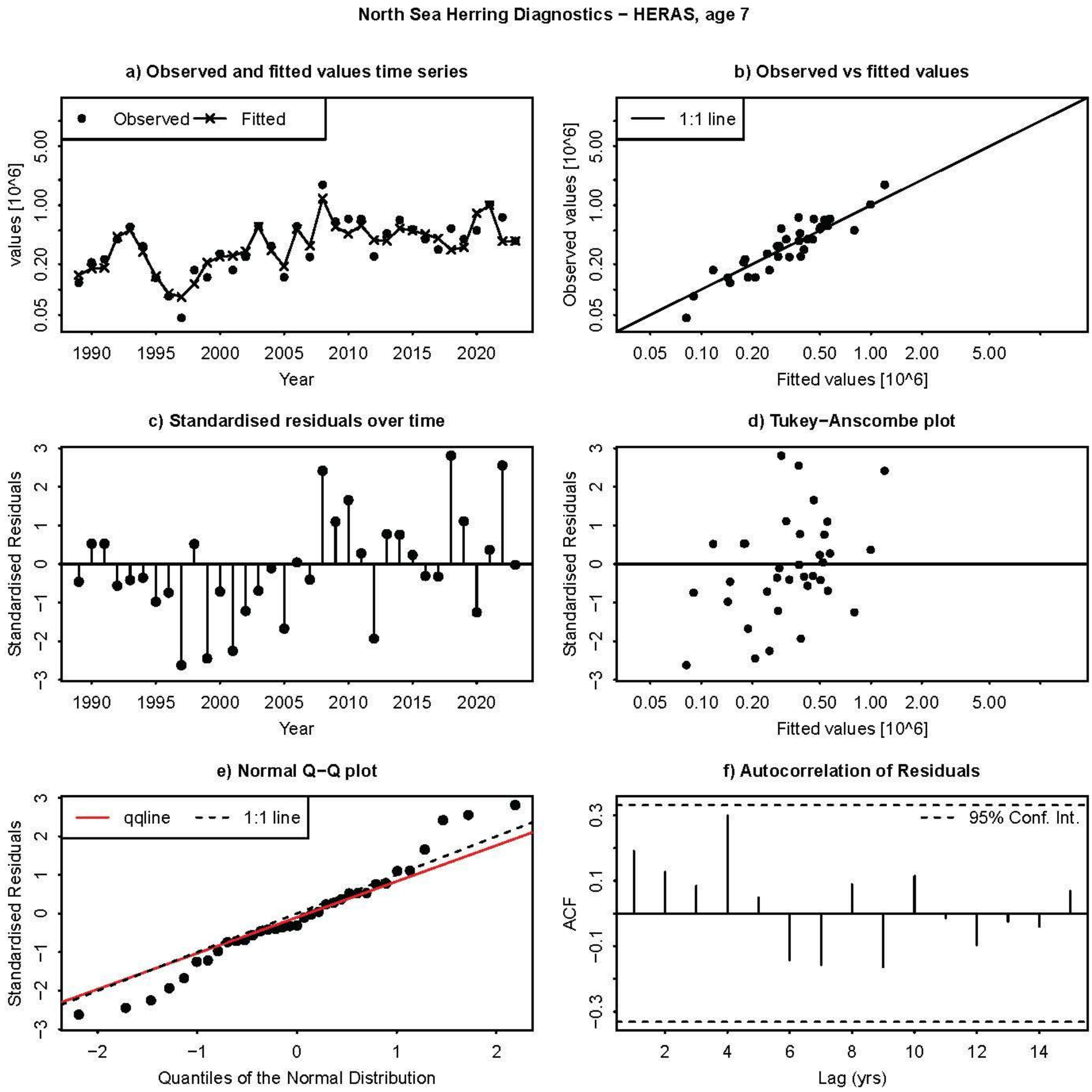


Figure A6.21. Assessment fit HERAS age 7.

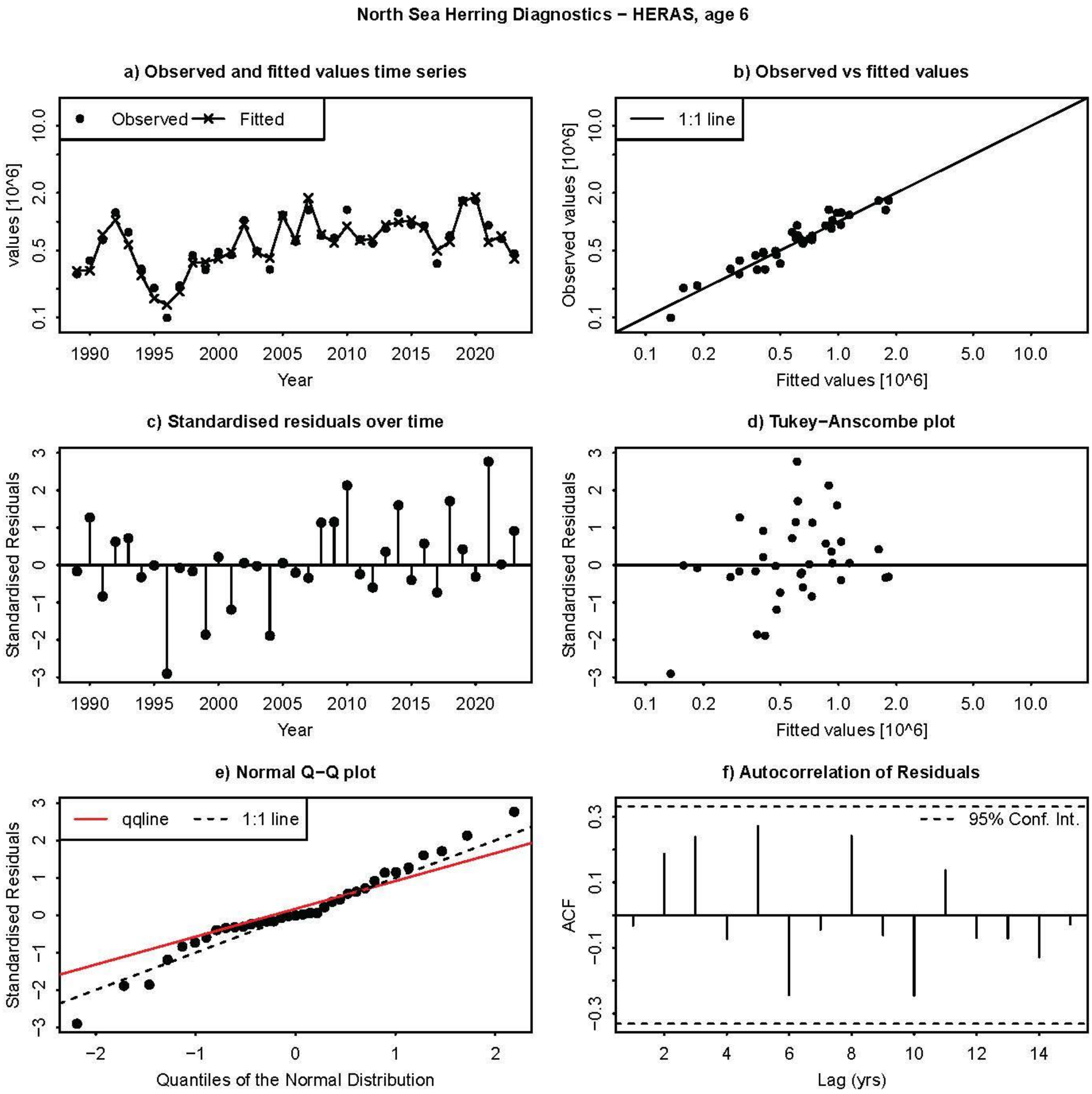


Figure A6.22. Assessment fit HERAS age 6.

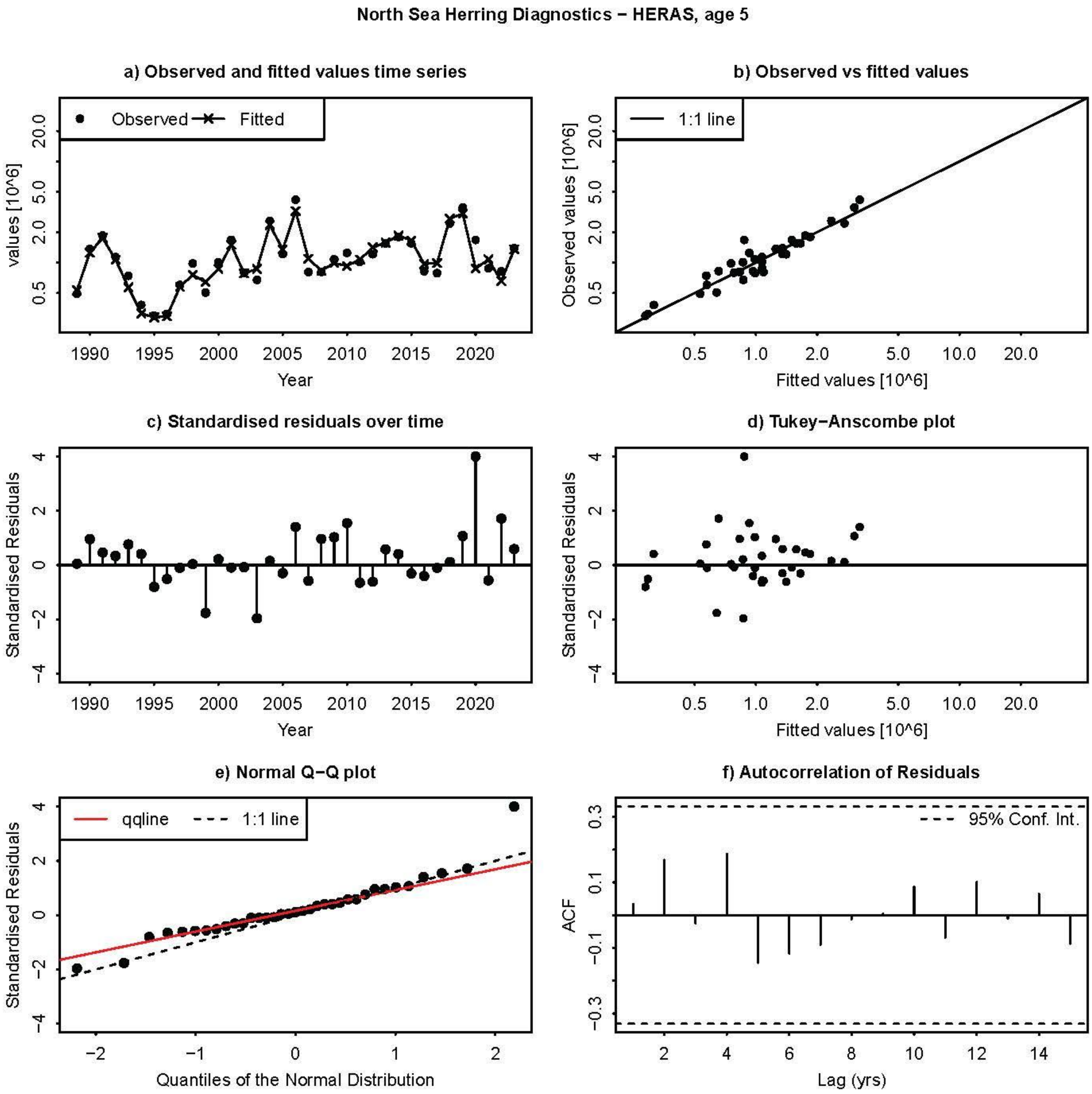


Figure A6.23. Assessment fit HERAS age 5.

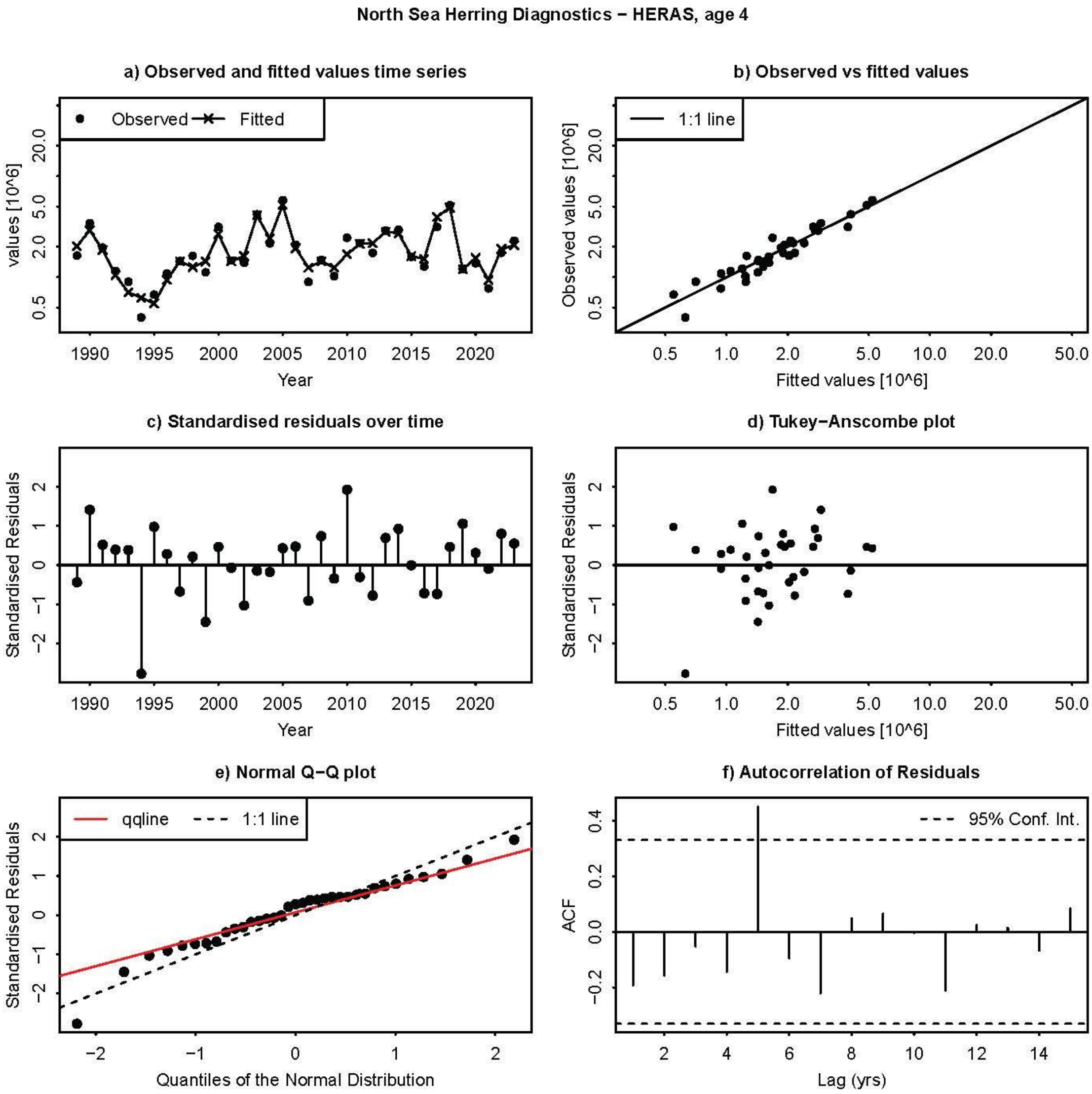


Figure A6.24. Assessment fit HERAS age 4.

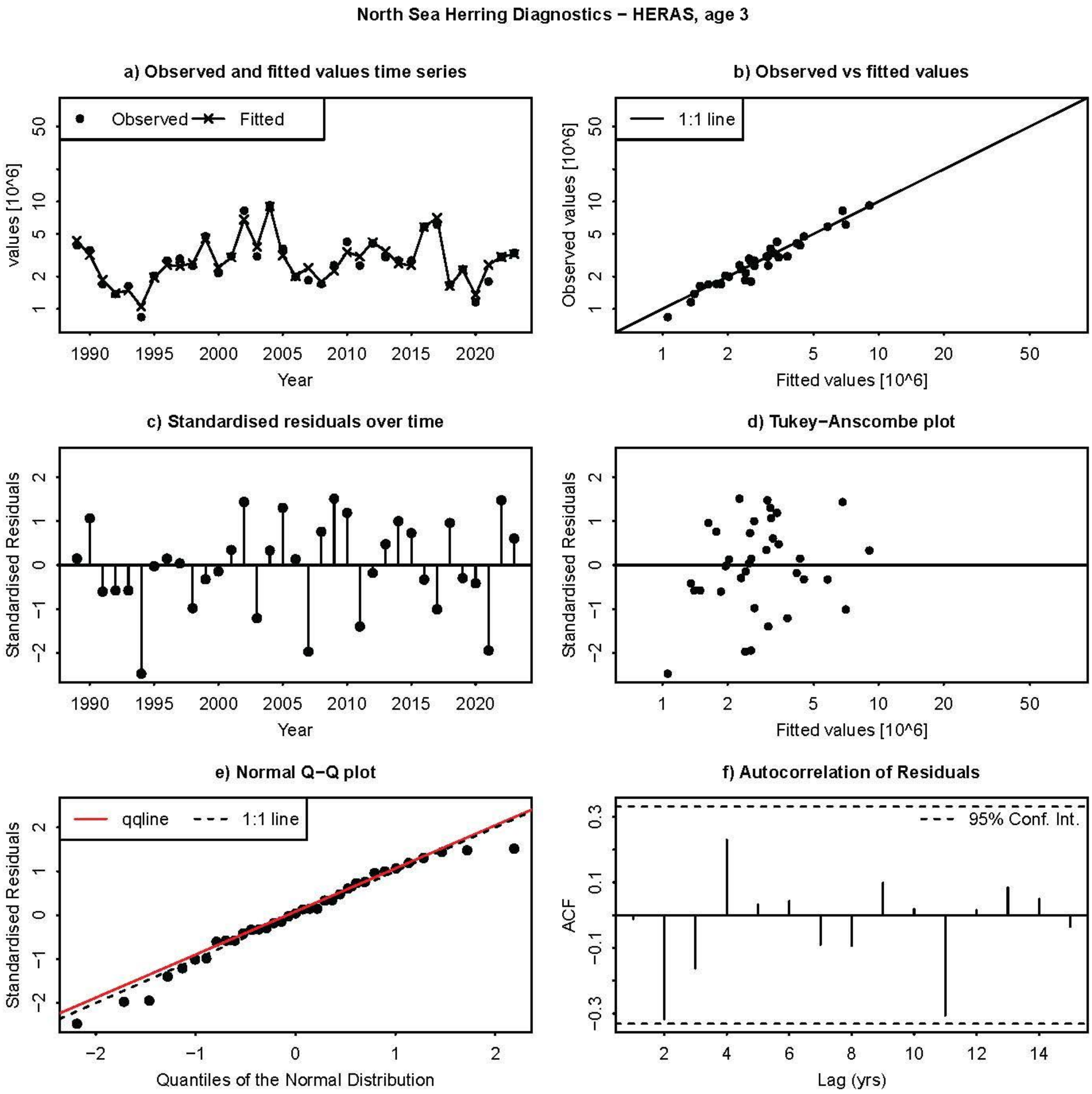


Figure A6.25. Assessment fit HERAS age 3.

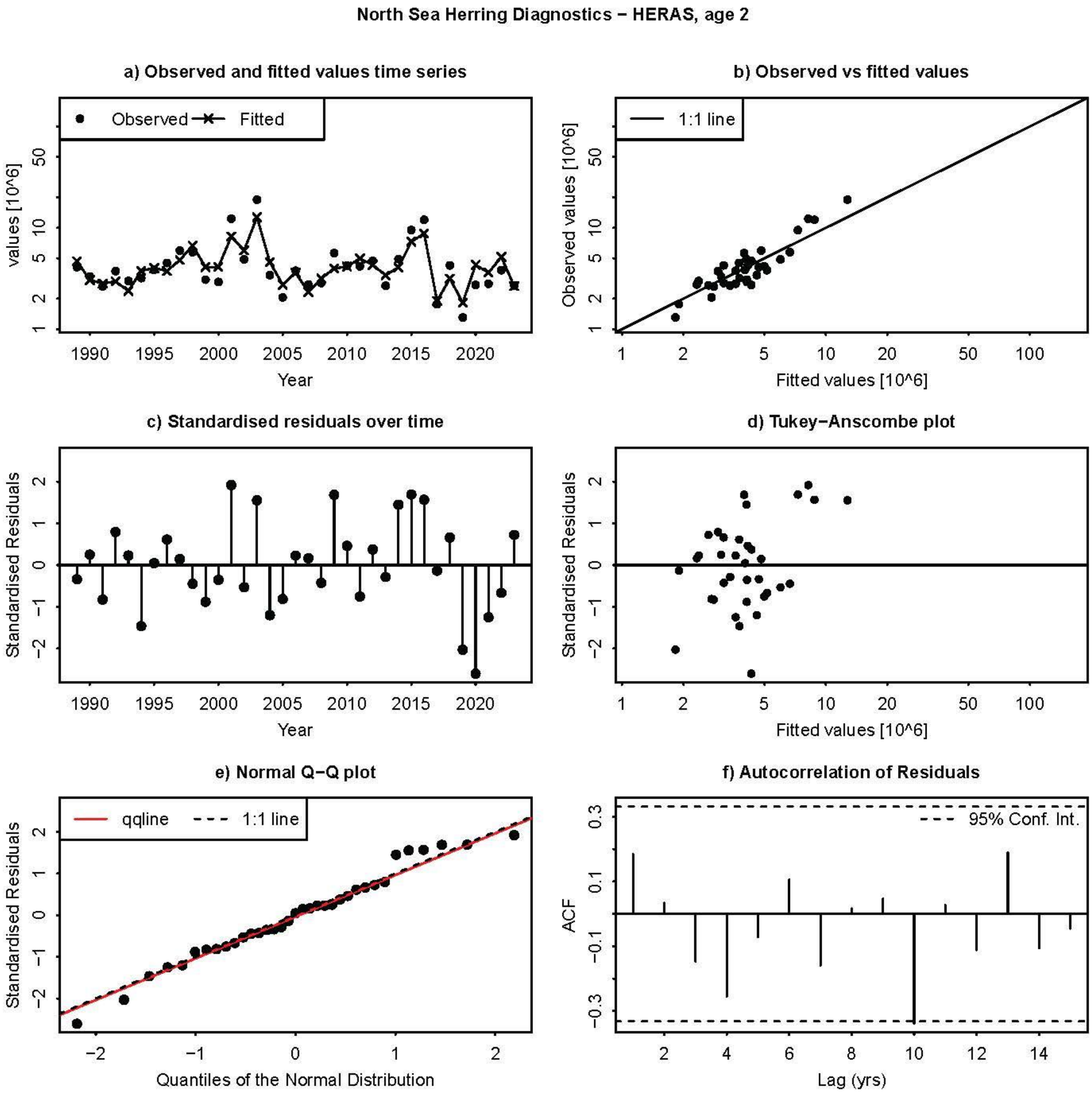


Figure A6.26. Assessment fit HERAS age 2.

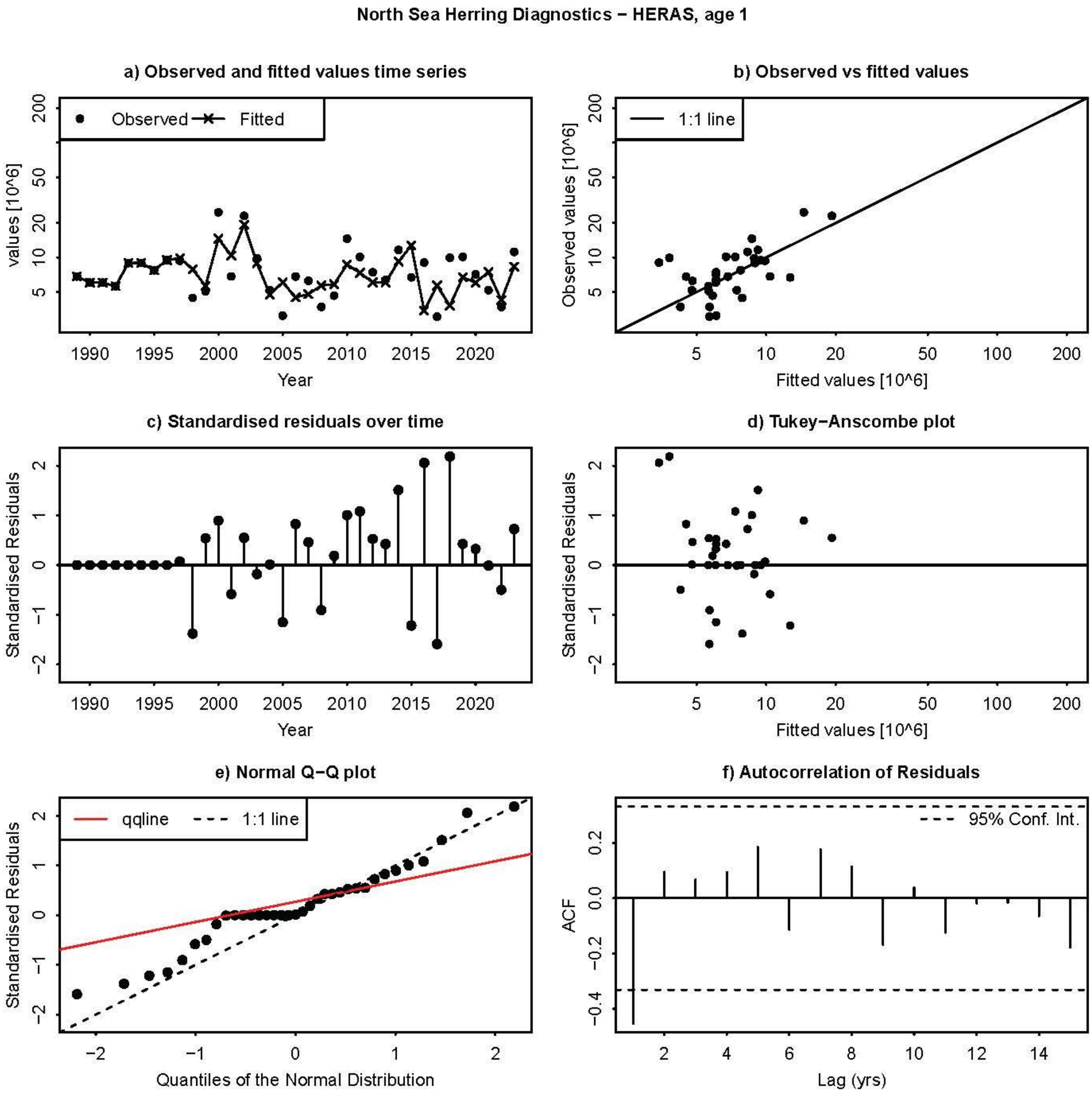


Figure A6.27. Assessment fit HERAS age 1.

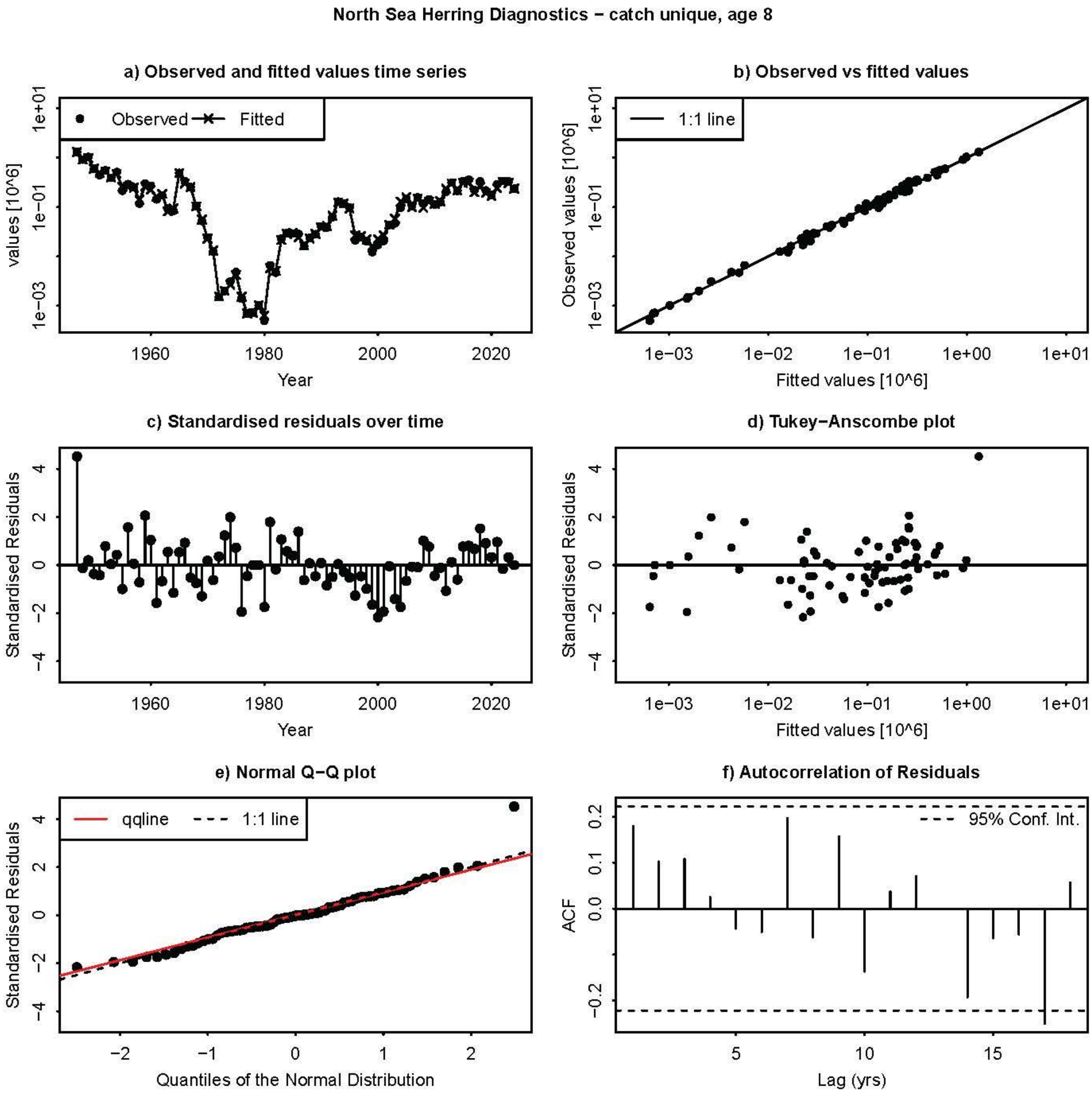


Figure A6.28. Assessment fit catch age 8.

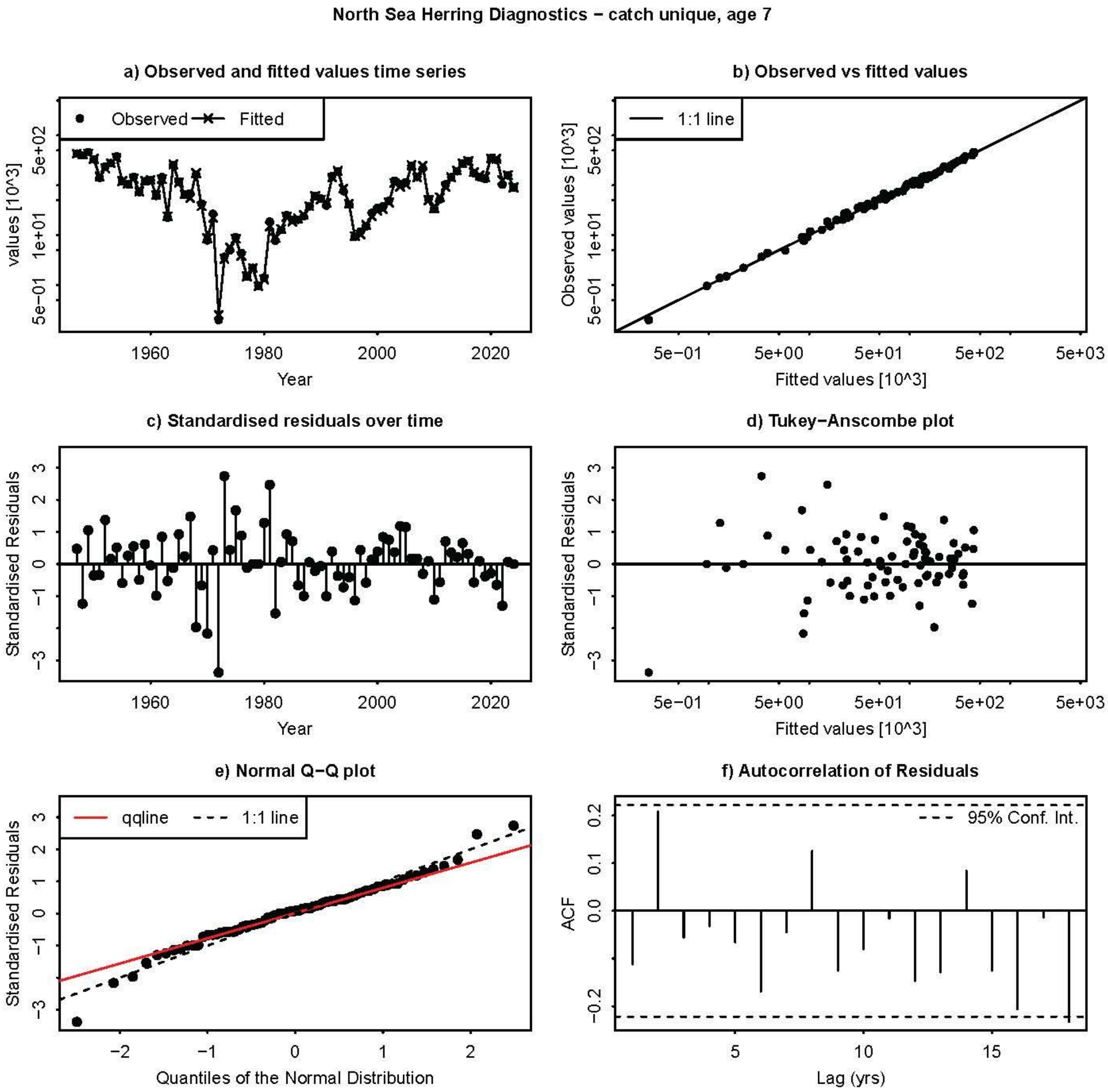


Figure A6.29. Assessment fit catch age 7.

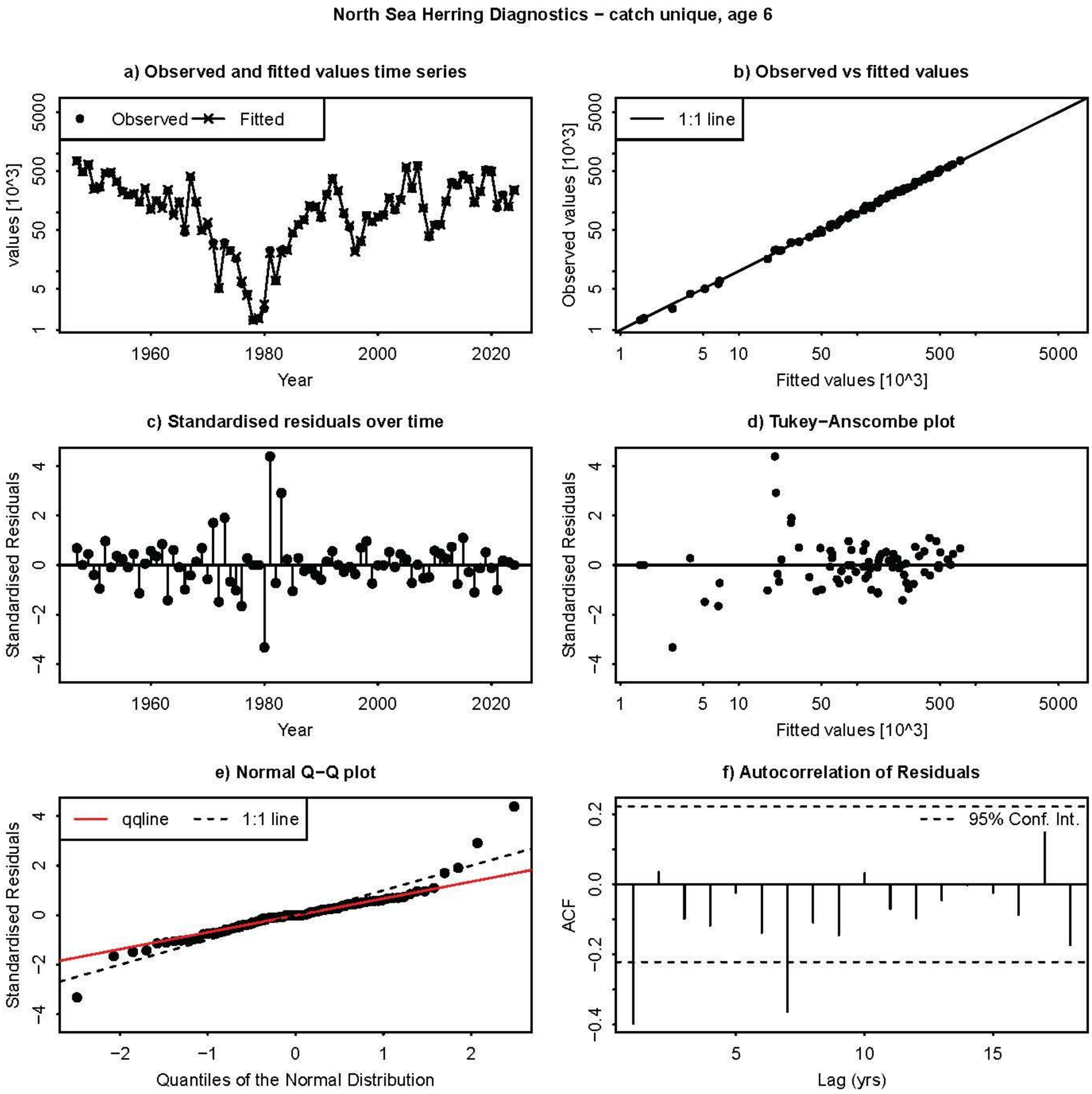


Figure A6.30. Assessment fit catch age 6.

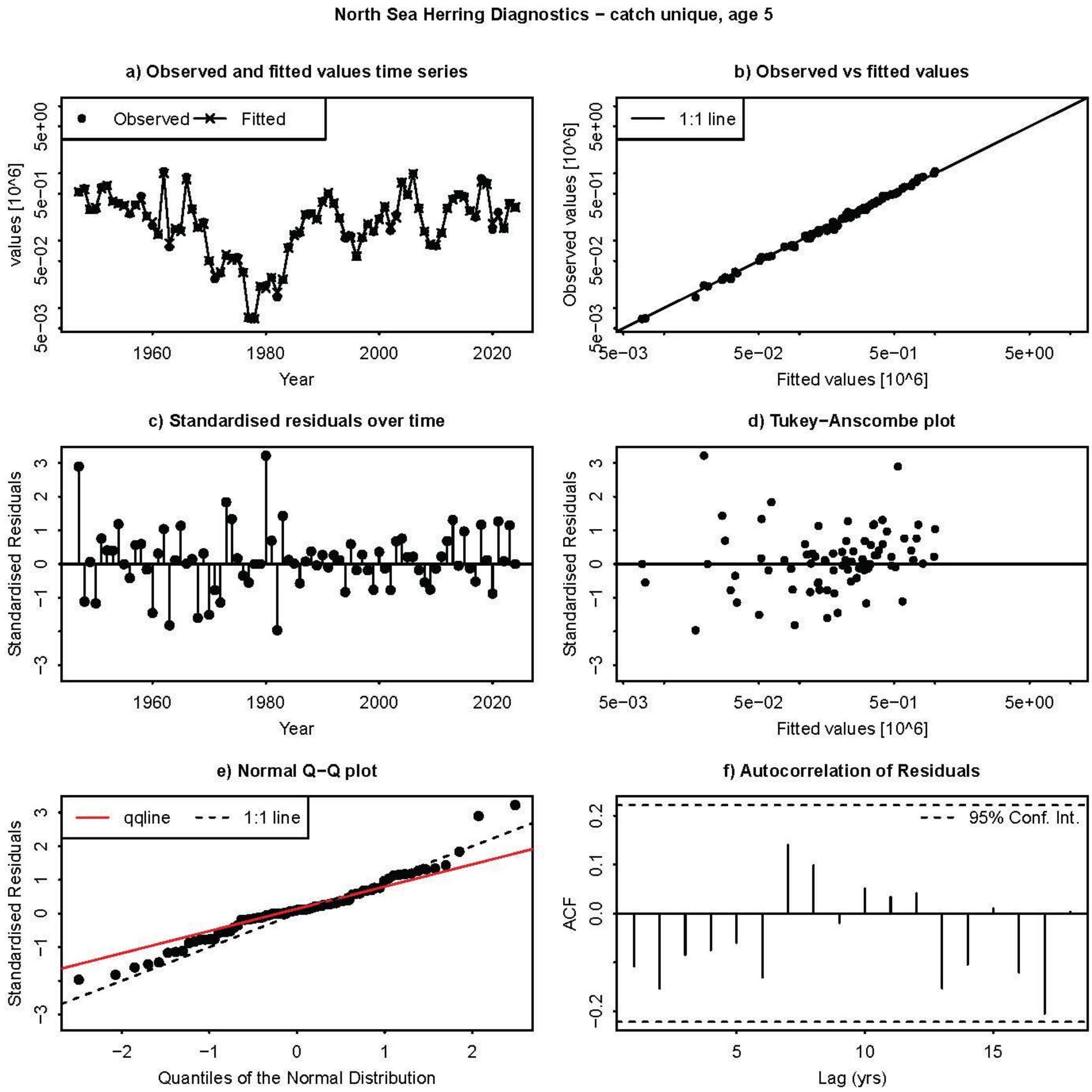


Figure A6.31. Assessment fit catch age 5.

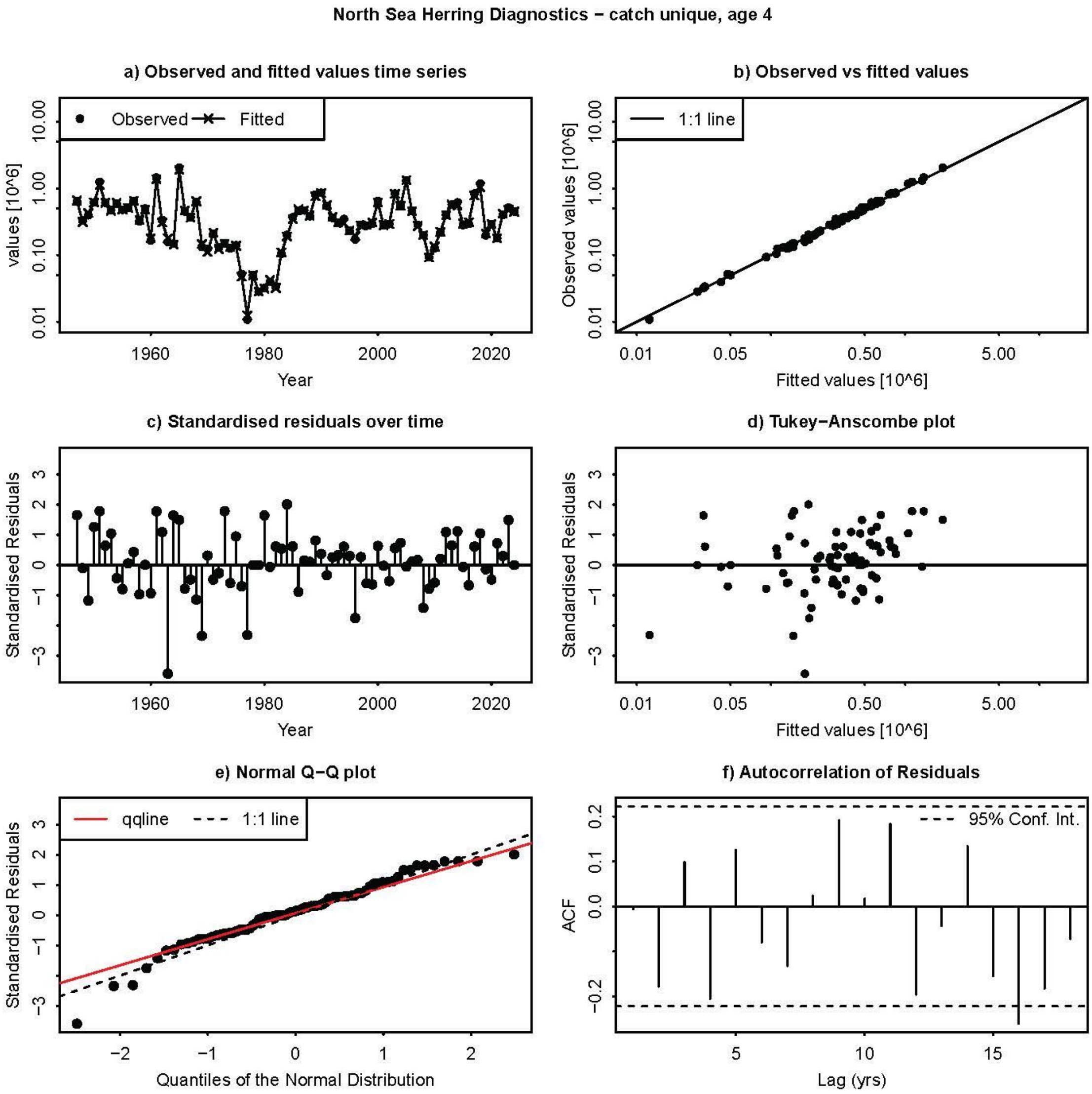


Figure A6.32. Assessment fit catch age 4.

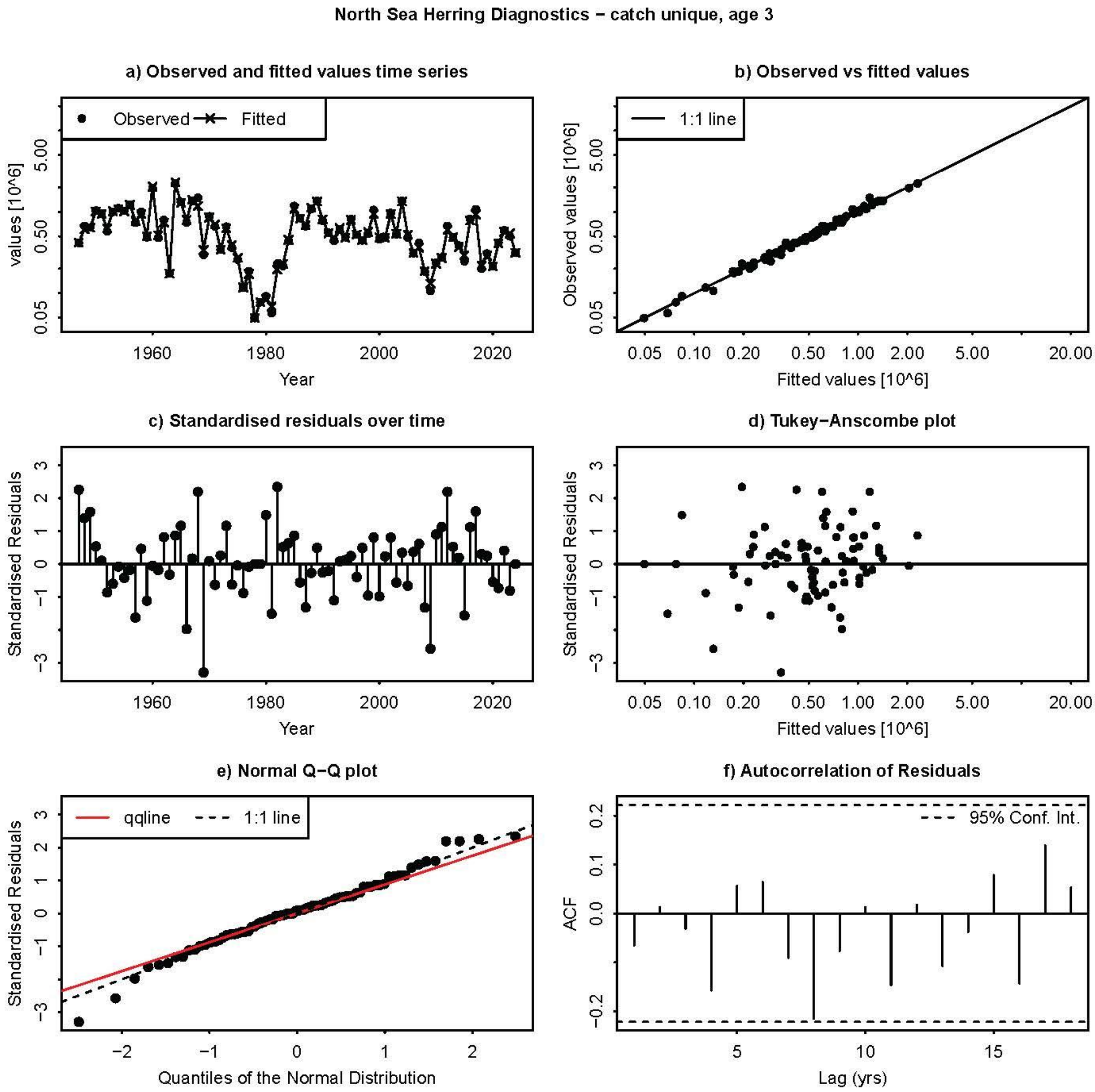


Figure A6.33. Assessment fit catch age 3.

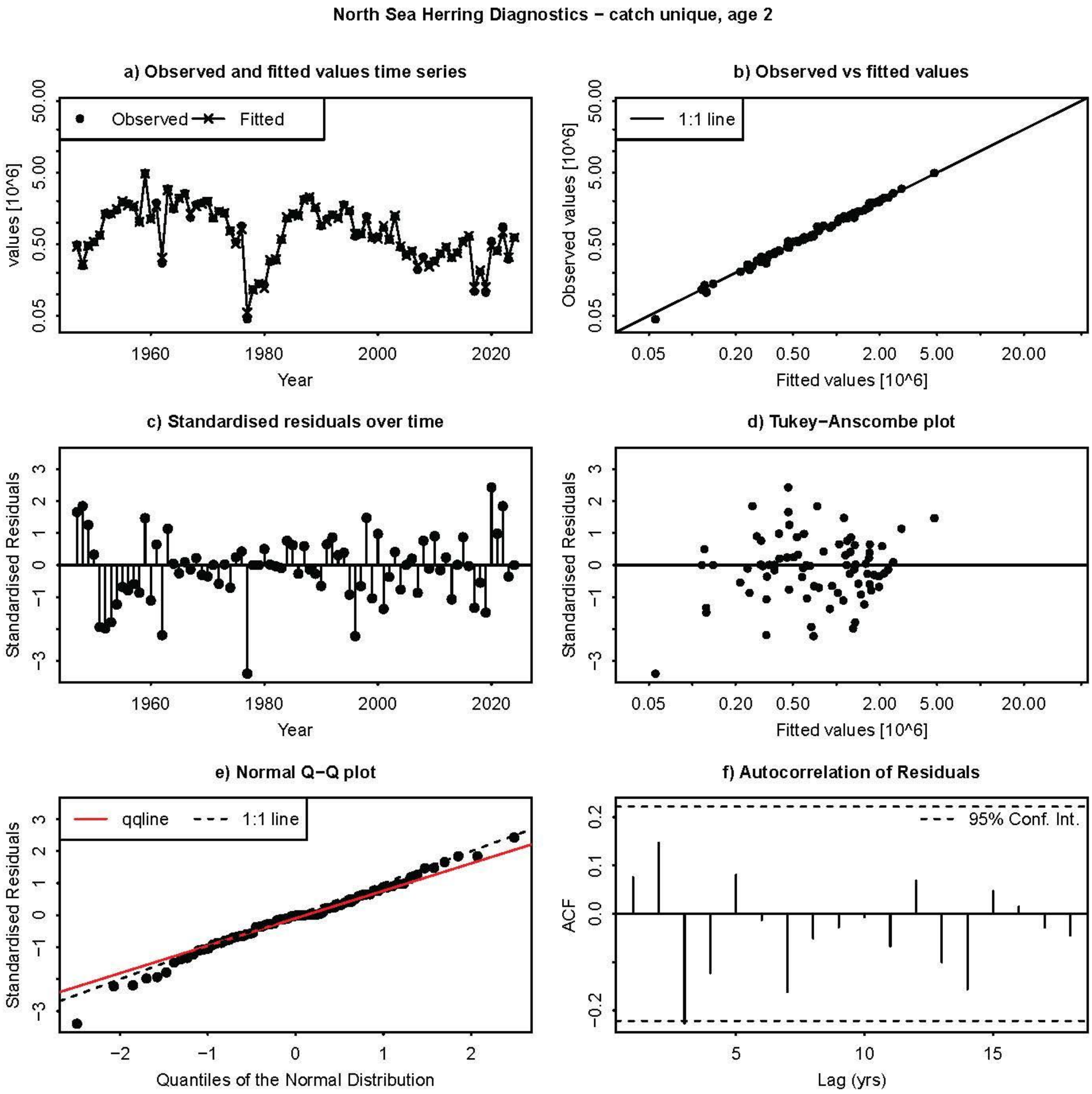


Figure A6.34. Assessment fit catch age 2.

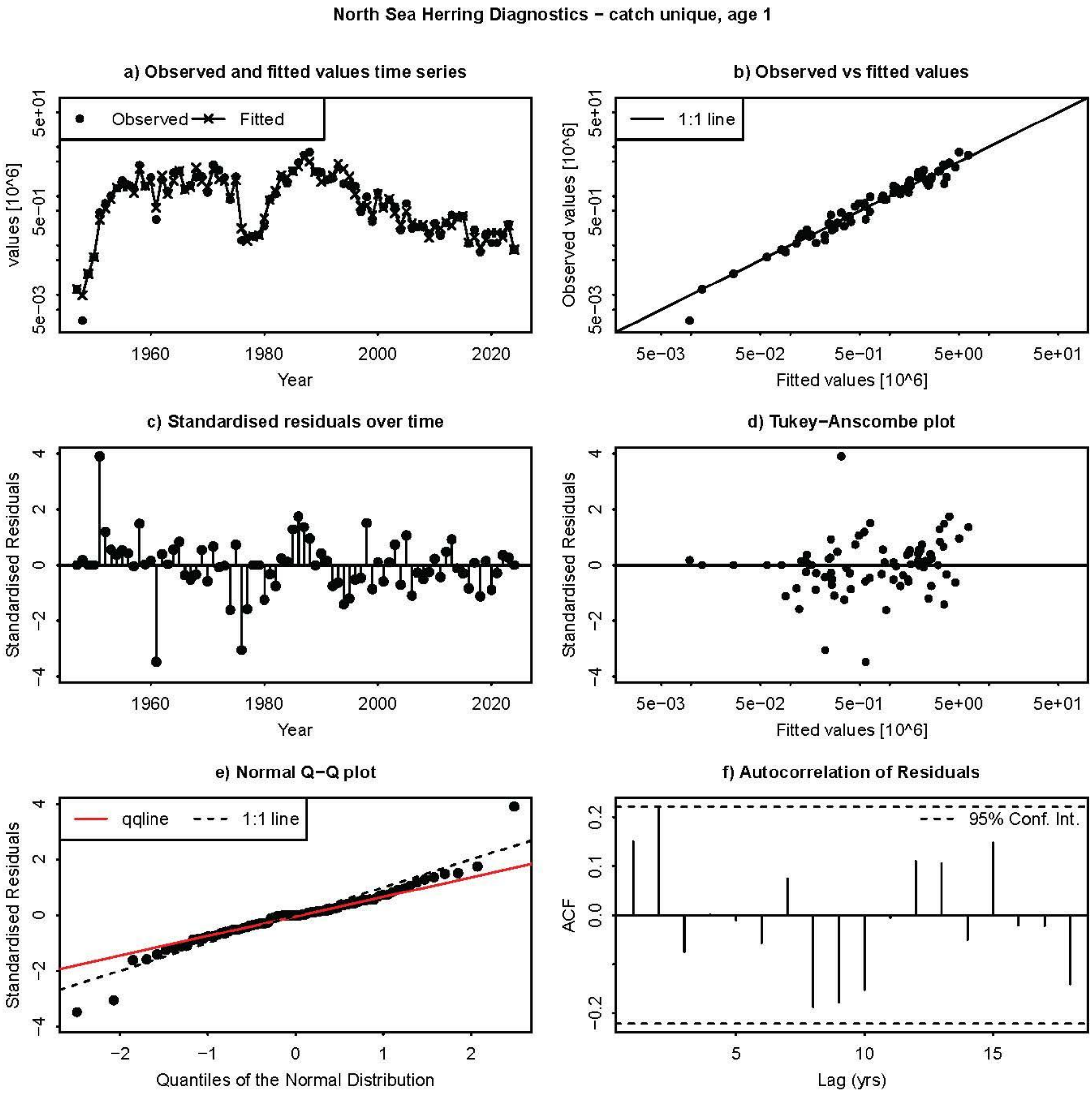


Figure A6.35. Assessment fit catch age 1.

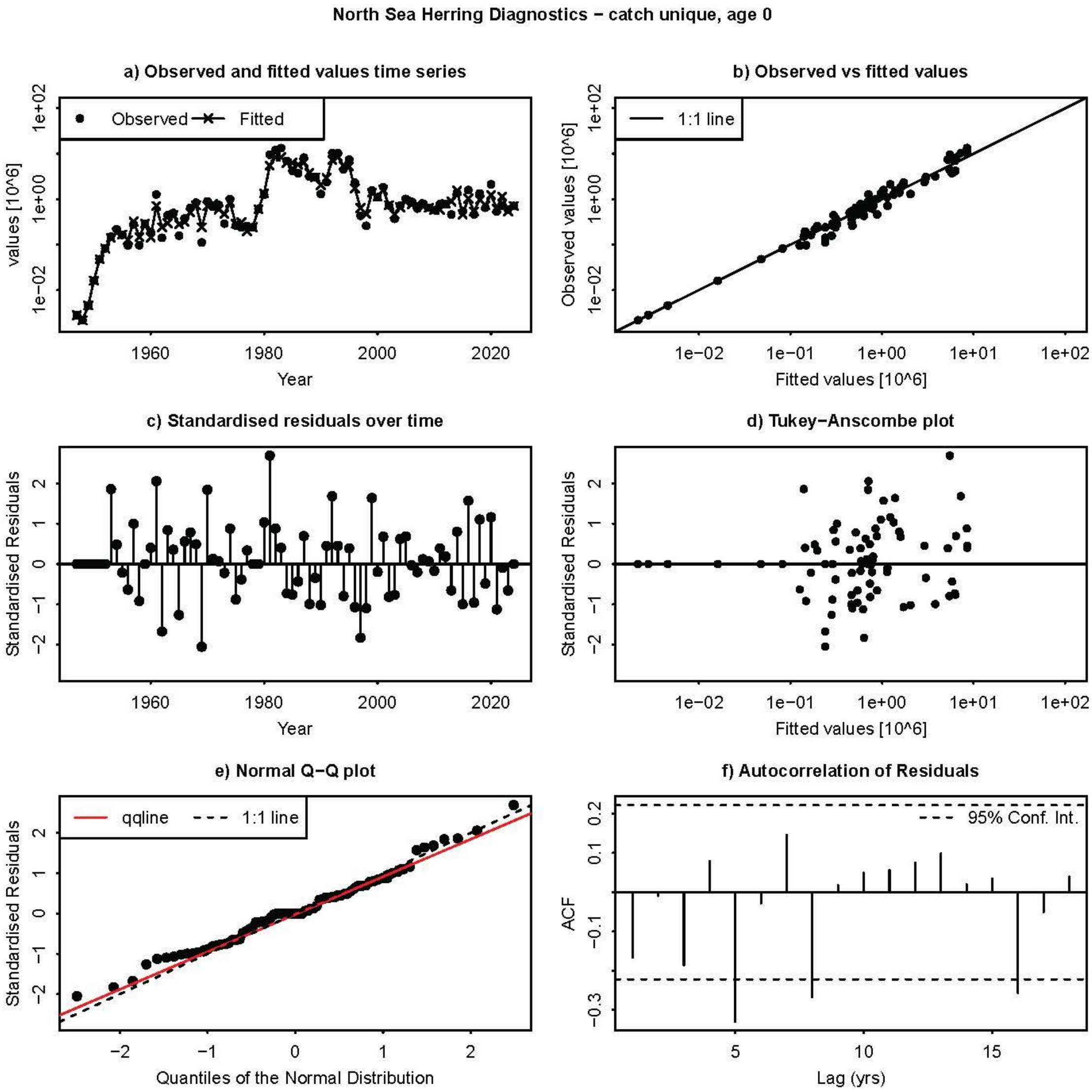


Figure A6.36. Assessment fit catch age 0.

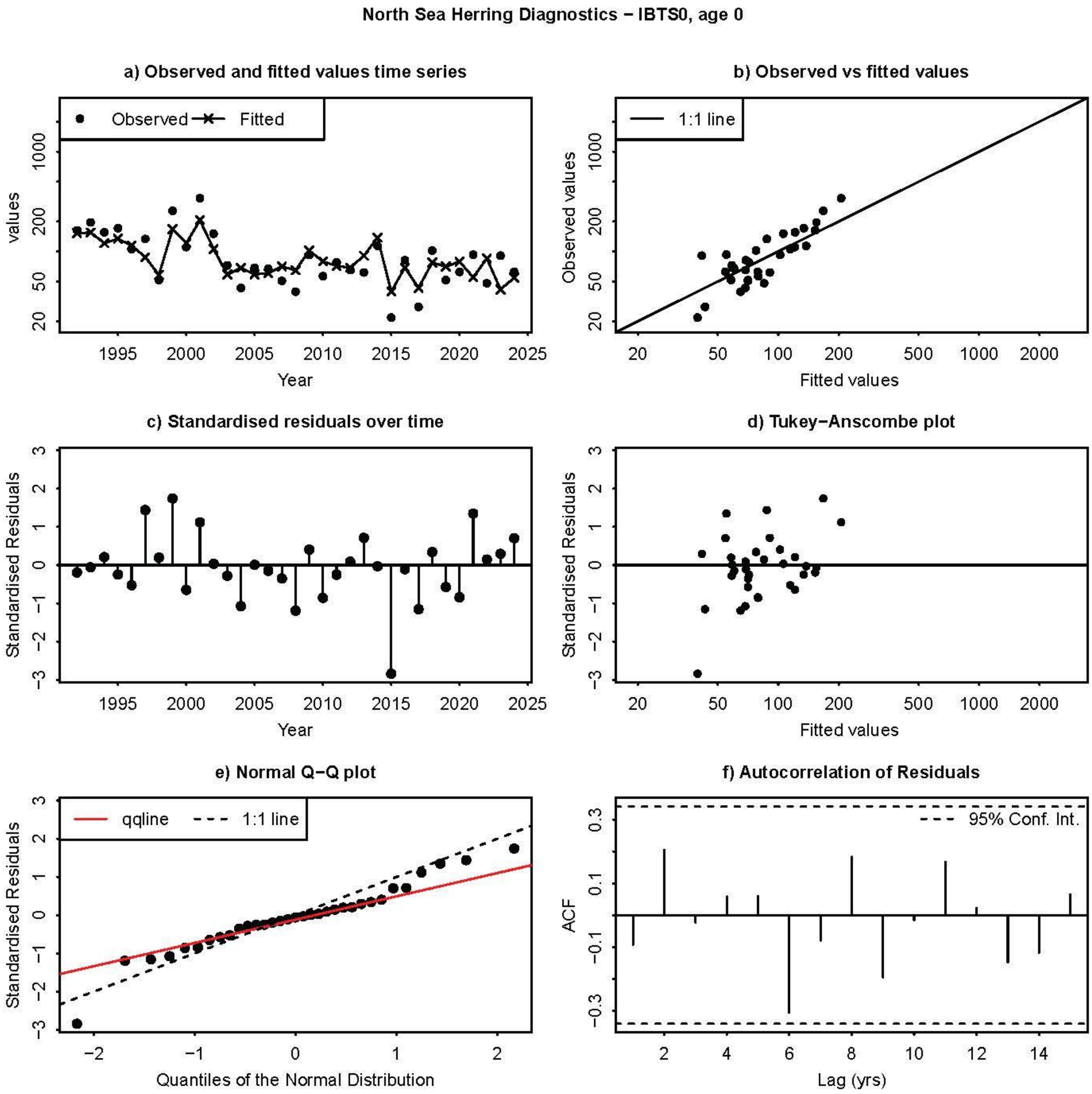


Figure A6.37. Assessment fit IBTS0 age 0.

Annex 7: Stakeholder engagement session

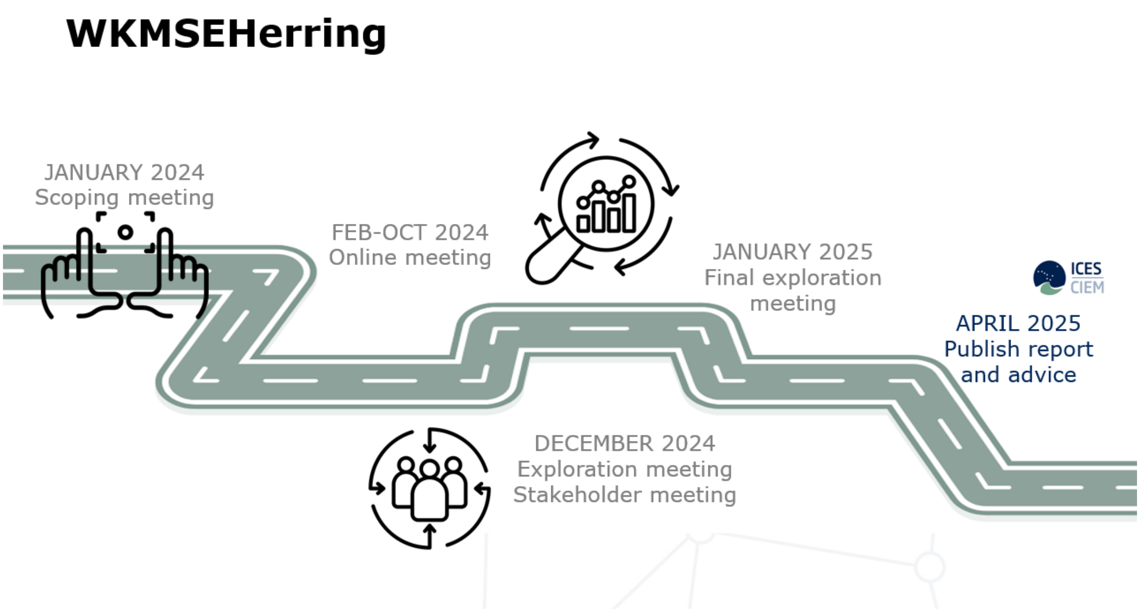
MSE provides the opportunity for meaningful stakeholder engagement. Participatory processes are critical to an MSE (Dichmont & Fulton, 2017; Miller et al., 2018), which offers multiple structures and processes to make it happen (Wilson et al., 2023).

The ICES community has called for developing MSE and harvest control rules in interactions between managers, stakeholders, and scientists (WKOMSE, 2009) and designed flow charts to analyze where and how it could occur (WKGMSE, 2013). However, stakeholder attendance, diversity and contributions to ICES MSE workshops are often limited (WKGMSE2, 2019), prompting recommendations for "stakeholder to play a more active role through the MSE process, and not just at the start and end of the process" (WKGMSE3, 2020).

In line with those recommendations, ICES has released the Stakeholder Engagement Strategy (ICES, 2023a) and the Implementation Plan (ICES, 2023b), which address the steps and challenges of how managers and ICES engage other stakeholders in MSE processes.

In this framework, WKMSEHerring held a Stakeholder Engagement Session (12 December) attended by 34 participants (18 in person, 16 online) from the EU, UK, and Norway. The profiles covered industry representatives, NGOs, Advisory Councils (PELAC and NSAC), government representatives, scientists, and ICES staff.

Figure A7.1. MSE herring timeline.



The session had two distinctive goals:

1. To present the preliminary results of the **ongoing process** regarding the Joint EU-UK-Norway request to ICES to advise on a long-term management plan for North Sea herring autumn spawners in North Sea, Skagerrak, Kattegat and Eastern English Channel.

2. To start a **forward-looking process** regarding the stakeholders' role in MSE processes within ICES.

Focusing on the ongoing process, the Chair explained the MSE approach, addressing the reliability of the North Sea herring autumn spawners data, model and knowledge. Communication tools make explicit important sources of uncertainty and facilitate the comprehension of the process, as evidenced by the questions raised by the participants (e.g. regarding climate change).

After establishing common ground, the request for advice and preliminary results were presented. The results illustrated the testing of various options across a broad range of potential scenarios for the fishery and its population. Participants sought clarity on how specific parameters were defined, such as mortality, who was involved in setting the advice request, and how to interpret the results, including the graphs. Overall, stakeholders expressed their satisfaction with the session and valued the clarity of the debate.

The forward-looking session built on the previous debate by examining the current ICES MSE process. It explored potential ways for stakeholders to engage, discussed these options, and assessed their feasibility. While the findings are beyond the scope of this report, they will inform future discussions about the next steps within the ICES community.

REFERENCES

ICES. 2023a. ICES Stakeholder Engagement Strategy. Version 01. ICES Guidelines and Policies. 12 pp. https://doi.org/10.17895/ices.pub.21815106

ICES. 2023b. Workshop on Implementation of Stakeholder Engagement Strategy (WKSTIMP). ICES Scientific Reports. 5:77. 68 pp. https://doi.org/10.17895/ices.pub.23507958

Dichmont C.M and Fulton E.A. (2017). Fisheries Science and Participatory Management Strategy Evaluation: Eliciting Objectives, Visions and System Models. In: Bunnefeld N, Nicholson E, Milner-Gulland EJ, eds. *Decision-Making in Conservation and Natural Resource Management: Models for Interdisciplinary Approaches*. Conservation Biology. Cambridge University Press; 2017:19-45.

Wilson, A.R., Miller, S.K and Galland, G.R. (2023) Management procedure development in RFMOs offer lessons for strategic and impactful stakeholder engagement and collaboration Front. Mar. Sci. , 18 Sec. Marine Fisheries, Aquaculture and Living Resources Volume 10 – 2023 https://doi.org/10.3389/fmars.2023.1112236

Annex 8: External Reviewers' Report

*Responses to questions raised by the reviewers have been inserted, where appropriate, clearly marked as "**Response**" and given in italics.*

Review of the Workshop on Management Strategy Evaluation for North Sea Herring (WKMSEHerring)

Carryn de Moor[3] and Tom Carruthers[4]

24 March 2025

Executive Summary

The report appropriately addresses the request, including the provision of tables of alternative combinations of control parameters which will satisfy the ICES precautionary criteria of long-term risk3[5]<=5%. The MSE undertaken to produce these results appropriately considered key uncertainties in the underlying conditions of herring population dynamics, subject to time-constraints. The amount of work covered was substantial and we would like to congratulate the Working Group, and in particular the analysts, for undertaking a substantial task in a constrained time period.

As noted in the report, the final selection of control parameters might not only be informed by projected average catch, as the difference is relatively small, but likely by other performance statistics/objectives including the average inter-annual variation in catch and whether stability constraints are preferred (i.e. a lower rather than higher $B_{trigger}$ control parameter, as these constraints are only assumed to apply when SSB is estimated to be above $B_{trigger}$). We strongly support the report's advice that the trade-offs between alternative combinations of control parameters be carefully considered by stakeholders. This set of results demonstrates how, ideally, there could be greater interactions between decision makers and analysts during the MSE to ensure that final results possibly cover a narrower range of options that satisfy all objectives. For example, instead of focusing on comparison of results with a HCR that gave maximum catch, decision makers might have preferred results corresponding to an alternative set of control parameters once the relatively small trade-off in average catch was observed.

Scope of this review

In contrast to 'standard' reviews elsewhere where reviewers are provided with a finalized output such as documentation and possibly code to review at the end of a process, reviewers were included in meetings (online and in person) from the early stages of WKMSEHerring. This written review does not include all verbal comments and/or advice given during this process (some of which has already been incorporated into the final report), but rather focuses on

[3] carryn.demoor@uct.ac.za. Marine Resource Assessment and Management (MARAM) Group, University of Cape Town, South Africa.

[4] tom@bluematterscience.com. Blue Matter Science Ltd., North Vancouver B.C. Canada.

[5] Max over years of $p(SSB(y)<SSB_{lim})$ over the time period (long term being 2034-2048 for this MSE).

documentation, presentations and discussion at the 10 - 12th December 2024 meeting at ICES headquarters, Copenhagen, and the final report (version provided on 12th March 2025).

Some examples of changes requested during the meetings are given in the Appendix.

In addition to this review, minor edits and comments that would not impact this review were provided in track changes and comments on a Word document version of the report (12th March 2025 version).

Documentation

Acknowledging the time constraints in this process and the comprehensiveness of the analyses that were undertaken, it is nonetheless recommended that any MSE process should have an accompanying 'trial specifications document' that provides a reproducible record of the methods applied, including the equations and model assumptions. One difficulty experienced by the reviewers was that the December meeting occurred before the methodology was documented in sufficient detail to support a comprehensive review.

Unfortunately, the equations provided in the final report were insufficient to ensure reproducibility in the methods. A package assessment was used (SAM) and a different package (FLR) was used to run the MSE. However, simply referring to those packages did not fully describe model assumptions and associated assumed processes. Annex 5 provided a 'model configuration' for SAM, but without associated documentation of what each number refers to - it would appear most of the numbers might simply refer to the estimated parameter number and indicates whether parameters were/not estimated.

A number of questions remain, for example:

- If there were any constraints on parameter values etc.

  Response: *Yes, those are given in the model configuration (Annex 5). A description of this model configuration can be found in the report of the IBPNSHerring workshop*[6]

- How large was the survey bias estimated to be? Figure 7.1 appears to indicate it is close to 1 for low ages and slightly overestimates numbers-at-ages 3+?

  Response: *This is shown through survey catchability in Figure 7.2, which is just above 1.1 over ages 3+.*

- How exactly is M(y,a) calculated. From Section 3.3 it appears that $M(y,q,a) = M_1 + M_2(y,q,a) + M_{add}$. How is M(y,a) calculated from M(y,q,a)?

  Response: *It is correct though natural mortality is only time at age varying; it is not clear what variable q relates to. Natural mortality M is the summation of background natural mortality M1 (0.05 for age 0 and 0.1 for ages 1+) and predation mortality (estimated by the SMS multi-species model). Further additive scaling is applied.*

- Understanding the first paragraph of section 4.4 (Fishing selectivity) would be vastly improved if equations of what was done were shown.

[6] ICES. 2021. "Inter-Benchmark Protocol of North Sea Herring (IBPNSHerring)." *ICES Scientific Reports. 3:98*.

Response: The process followed the one used in the SAM stock assessment model. We suggest referring to the publication describing the model[7].

- To clarify, is $Zprespwn_{a,y} = 0.67\ Z_{a,y}$ (Section 4.8.5)? Is it reasonable to assume that fishing mortality occurs evenly through the year, which is what this equation would assume?

Response: Zprespwn comprises both natural and fishing mortality before spawning. The fishing pressure on NSAS herring is not even through the year and is mainly concentrated in Q1-3, e.g. low in Q4.

- What is the difference between 'Percentage inter-annual change in catch/SSB' and 'Average annual variability in catch/SSB' (Section 5.1). It appears that only the latter is used in the remainder of the report.

Response: The difference is the unit; whilst one is the change in %, the other one is in absolute catch amount. The performance metrics not used were deleted from the list in Section 5.1.

- The first set of empirical HCRs are not clearly defined; the equations do not match the text or Figure 7.3. In addition, it is not explained how the HCR calculates a TAC from 'mult'.

Response: The descriptions provided in and consistency of Section 7 have been improved.

The report does not provide equations or detailed text to explain how the future fishing effort for fleet A, E_A, is generated from the TAC advice. Future fishing effort for the remaining fleets depends on this E_A.

Response: The fishing effort of fleet A is computed in accordance to targets described in Section 4.9. No more than four targets are needed in accordance to the number of fleets in the MSE model.

In some places in the report, the EM and OM are described together (eg section 4.5), presumably since both were SAM models. However, the OMs only should be described in section 4, and the EMs should be described in section 5 as part of the Management Procedure.

Response: This aspect is acknowledged and specific comments were addressed. However, time was lacking to reshape the different sections in depth.

It is not clear how WKMSEHerring distinguish between sensitivity tests and robustness tests (e.g. section 3.5, table 3.2 and 3.3). It may be that they used robustness tests to refer to uncertainties about the underlying science, but sensitivity tests to alternative candidate MPs (e.g. without constraints; including banking and borrowing). Typically uncertainty in an assessment is tested with sensitivity tests, while robustness of an MSE to uncertainties is considered through robustness tests (alternative OMs). Either way, this distinction should be clarified in the text, or all referred to as robustness tests.

Response: Sensitivity tests refer to specific scenarios in the request and the associated OM conditioning for fishing exploitation patterns deviated from the reference set of OMs. In contract, robustness sets shared similar assumptions to the reference set for fishing exploitation patterns but tested specific assumptions not covered in the reference set.

[7] Nielsen, A., & Berg, C. W. (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, *158*, 96–101. https://doi.org/10.1016/j.fishres.2014.01.014

Software

The R statistical software (R core team, 2024) was used to condition operating models using the SAM stock assessment framework (Nielsen et al. 2014; 2021), specify MSE calculations and management procedures using FLR (Kell et al. 2007), and then conduct closed-loop simulation projections using FLR and Flasher (Scott and Mosqueira 2023). Prior to the December 2024 review meeting code was provided demonstrating the various steps of MSE calculations allowing reviewers to interrogate the operating models.

Presentation

At the December meeting, presentations were provided on this MSE request, the terms of reference, operating model specification, MP configuration, observation error models and projection models.

Operating models

WKMSEHerring opted to include key uncertainties about natural mortality and the relationship between recruitment and spawning stock abundance through a range of 'baseline' Operating Models, hereafter referred to as the Reference Set (RS) of Operating Models (OMs). Further uncertainties were tested with robustness tests. This follows best practice. The robustness tests included models that were either less plausible or lower priority than those included in the RS. The candidate Management Procedures (MPs) were tuned to the RS of OMs which were weighted equally between the categories (base OM, stock recruitment OMs and natural mortality OMs). Given time constraints, only the 'optimal' candidate Management Procedure (that corresponding to maximum average long term catch) was tested against the robustness tests. These choices are considered very appropriate, particularly under the given time constraints.

Stock structure is frequently a key uncertainty in population dynamics models when the modelled population does not consist of a single homogeneously distributed stock within the management area(s). This MSE considered management of North Sea Autumn Spawner Herring (NSAS) which is distributed in the North Sea and ICES division IIIa. There is mixing with the Western Baltic Spring Spawning (WBSS) herring in ICES division IIIa and along the Norwegian coast and thus 'herring' catches consist of a mix of both stocks. These landings were previously separated between stocks using vertebral counts and, more recently, using otolith microstructure. The large majority of landings consist of NSAS herring and the NSAS herring stock is estimated to be an order of magnitude larger than the WBSS herring stock. The impact of this uncertainty w.r.t. the mixing between stocks on management of NSAS is thus expected to be small. (In contrast, this stock mixing should be considered a key uncertainty in the management of WBSS herring.)

One key uncertainty selected by WKMSEHerring was the historical and future rate of natural mortality (M). For the base model, M was informed by output from a multispecies model (the 2023 SMS key run. However, there have been substantial changes in the absolute value of M over time as the multispecies model has been updated (Section 4.5.2). Thus treating M as a key uncertainty through alternative OMs (M2-M5) is a reasonable priority.

The other key uncertainty selected by WKMSEHerring concerns the stock recruitment relationship that will apply in the future. The base model assumed future productivity would reflect that of the recent (~20yr) past, while robustness (SR2) was tested assuming future productivity would reflect that of the full historical period (>74yrs). The base model assumed future recovery dynamics would reflect that of the full time period, while robustness (SR1) was tested assuming future recovery dynamics would reflect that of the more recent period. The choice between the base and these robustness tests was reasonable. Depensation (SR3) and autocorrelation (SR4) were also considered, although for SR3 the estimated depensation effect was relatively weak.

While Figure 4.20 showed the estimated stock-recruitment relationships, no figures were provided showing the fitted curves together with the data points, preventing evaluation of goodness-of-fit. For example, section 4.6 refers to the poor statistical fit being the reason why the segmented regression relationship was not used for SR3, but this is not shown. Similarly, it is difficult to evaluate the depensation OM (SR3) without comparing its fit to the data to that of SR2.

The operating model is a multi-fleet model (Fig 3.14, section 3.5), excluding the LAI index from the likelihood (Section 3.4) (The estimation model is a single-fleet model). Thus all conditioning for parameter values etc and demonstration that conditioning has been adequately achieved for the OM must be for a multi-fleet model which matches that of the OM. Annex 6 should thus show the model diagnostics for the multi-fleet model excluding the LAI index, and all parameter values sampled for use in the OM should correspond to this same model.

- The large positive residuals for all years for age 7 of Figure A6.2 is concerning and should be explained. Might this be due to model mis-specification, or point to the need for higher Ms for older age classes?

  Response: *Figure A6.2 shows deviations from the survivor equation from the last true age to the plusgroup. The large values observed at age 7 is due to the plus group. This is not a model issue but a plotting glitch.*

- It is intriguing that 'catch unique2-6' and HERAS 3 are fitted with a smaller error than that observed (Figure A6.10); it would be worth explaining why.

  Response: *There may be confusion as to what this Figure represents. It is the estimate of observation variances, and the uncertainty of these estimates (reflected as CVs). This is not showing fitted versus observed.*

- The very low observed recruitment in the final year of Fig A6.19a is a concern, with the associated poor fit. Is this a real observation?

  Response: *This last point is from the IBTS-Q1 age 1 index in 2024 which is the lowest in the time series. Despite the apparent anomaly in the context of the entire time series, the survey was considered appropriate by survey leaders. It is suspected that this low abundance is due to large predation of haddock and whiting on juvenile herring.*

In addition to not being able to verify that conditioning of the OM was adequately achieved, there are two further comments on the OMs:

- Section 4.6 reports that only 500 replicates were sampled for SR3. Yet 1000 replicates were used for each OM (Section 4.1, 5.5). Is this a typo, or were the 500 replicates doubled? It would be best to sample 1000 replicates from the Beverton Holt function for SR3.

  Response: *SR3 is indeed comprised of 500 iterations, to comply with the 50/50 mix between Beverton-holt and segmented regression (and therefore 500 replicates for each) in all OMs. It would have been possible to extend this to 1000 replicates but it was not done considering that SR3 was only part of the robustness set.*

- The report also notes the 'overshoot' of the TAC in recent years, apparently due to TAC allocated to the C fleet being taken by the A fleet instead (section 3.2). While it is not clear from the report why this occurs (the total TAC given from the C fleet should have matched that gained by the A fleet), if this 'overshoot' is substantial and likely to continue into the future, it should have been included in the Implementation Error.

Response: *That's an important point. No overshoot was implemented in the MSE but in recent years the overshoot is ~3% of total TAC due to no accounting for TAC transfer taking place from area 3.a to the North Sea.*

Management procedures

MSE closed-loop testing focused on model-based management procedures based on a SAM estimation model with various hockey-stick harvest control rules.

Although not part of the original request, some empirical MPs were also considered. These are faster to simulate with MSE as well as quicker to implement in practice as the TAC advice from the harvest control rule is directly dependent on observed data and no estimation model is included.

MSE results

B_{lim} was calculated based on per-simulation estimation of the breakpoint in the hockey-stick ('segreg') stock-recruitment function. Section 4.10 and Figure 4.25 indicate B_{lim} is similar across the base OM and OMs SR1-4. However, Figure 4.20 shows the B_{lim}s for OMs SR2 and SR4 would be similar to the base, but that for SR1 would be higher. Additionally, the report does not describe how B_{lim} was calculated for SR3 given this is selected from the segreg fit (Section 4.10), but the segreg fit was deemed unsuitable (Section 4.6). This is particularly important given the 'buffer' provided by the flatter BH curve estimated for SR3 compared to SR2 where recruitment is assumed to remain about an almost unchanged level as biomass decreases until ~1.5 million t and thus projections under this OM would be less risky, i.e. less likely to fall to undesired levels of biomass.

This approach leads to the counter-intuitive result that B_{lim} values are higher with higher natural mortality rate (Figure 4.25, 4.26). For example, scenarios M2 and M5 have M values 7% and 10% higher than base values, leading to estimated B_{lim} values that are 25% and 50% higher than base levels. This is not the case for definitions of B_{lim} used elsewhere such as Canada and U.S. that are based on fractions of SSB_0 and SSB_{MSY} which are lower with increased M, accounting for the expectation of a smaller more productive stock. The approach used here sets a higher bar for evaluating risk in high M scenarios than in MSE processes elsewhere (see further comment on risk below).

It would be useful to see plots of future projections against historical years, e.g. Figure 5.12, to compare how the projected SSB and recruitment compare with that estimated historically.

Response: *This is partly shown in Figure 5.4 for SSB/recruitment pairs, but see Figure A8.1 given below.*

The analysts ran the MSE for a fixed number of $B_{trigger}$-F_{target} combinations and then interpolated results over the remaining combinations. The results from an initial selection of grid cells before interpolation resulted in further cells being selected to ensure more cells close to the 'risk boundary' were calculated rather than interpolated. This choice to include interpolation was appropriate given the amount of time required for simulating each MP-OM combination, particularly given the model-based MPs used.

The 'optimal' combination of the control parameters $B_{trigger}$ and F_{target} selected was based on average long-term total catch, as requested. However, the range of values achieved for this performance statistic (average catch) was low (4%) over a fairly substantial range of $B_{trigger}$ (800 000t to 1.7 million t) and F_{target} (0.18 to 0.39). In contrast, the difference in the performance statistic measuring the inter-annual variation in long-term catch was far larger (87%). Thus the requested selection for the 'optimal' combination of control parameters may thus not adequately reflect stakeholders' needs. In addition, high $B_{trigger}$-F_{target} combinations (like that selected for the 'optimal' combination) imply that in many simulations TACs are set using an F value from the descending limb of the HCR and not at F_{target}. Constraints on inter-annual variation in TAC only apply for $B>B_{trigger}$, and thus apply infrequently for high $B_{trigger}$-F_{target} combinations.

We thus strongly support WKMSEHerring's suggestion that stakeholders consider the trade-offs between alternative $B_{trigger}$-F_{target} combinations. Ideally, there could be greater interactions between decision makers and analysts during the MSE to ensure that final results possibly cover a narrower range of options that satisfy all objectives. For example, instead of focusing on the comparison of results with a HCR that gave maximum catch, decision makers might have preferred results corresponding to an alternative set of control parameters once the relatively small trade-off in average catch was observed with initial results.

The MPs spanned the principal trade-off between what is taken (catch) and what is left (biomass) but it was not clearly communicated that the latter is related to average expected catch rates which determine fishing efficiency and hence profitability (which is typically of interest to industry stakeholders). For example, in many fisheries, industry would readily forgo 4% yield for a 13% increase in expected catch rates (as implied across the MPs in Figure 5.20).

The risk under OM SR4 is substantially higher than (almost double) that for the other individual OMs. If future reality appears to reflect that of the assumptions of SR4 (highly autocorrelated recruitment) then the MP selected so that the risk for the reference set is precautionary may no longer be precautionary in practice. In this case one may consider declaring Exceptional Circumstances (e.g. de Moor et al. 2022). To assist in determining if Exceptional Circumstances exist, future indices or survey observations - particularly, in the context of SR4, those measuring incoming recruitment - should be monitored to see if future indices/observations more closely match that of the reference set or that of SR4 (cf Section 5.5).

In terms of the results/comparisons requested (Annex I):

i) "The recent exploitation pattern with F0-1=0 from above contrasted with exploitation patterns moved one year older and one year younger fish (three scenarios)."

> This was implemented by assuming all fishing would be from fleet A only and none from fleets B-D (section 6.1.3). While this effectively sets F0-1=0, it does also slightly reduce the selectivity of ages 2-6 as well (Figure 4.10).

ii) "All alternatives should be evaluated with and without a constraint on the inter-annual variation of TAC"

> Section 6.2.1. demonstrates the difference between including/excluding the constraints for 4 combinations of control parameters for the base OM only. The request might have implied results should have been shown for all 7 reference set OMs and the combination thereof, as well as comparisons with/out constraints for the sensitivity tests to different exploitation patterns etc.

iii) "The constraint mechanism shall be tested separately from and in combination with 10% banking and borrowing mechanism."

> Section 6.2.2 demonstrates the difference between including/excluding [one version of] the Banking and Borrowing scheme for 4 combinations of control parameters for the base OM only, assuming constraints on the inter-annual variation of the TAC apply if $SSB>B_{trigger}$. The request might have implied results should have additionally been shown

for all 7 reference set OMs and the combination thereof, also excluding the constraints on the inter-annual variation of the TAC.

Response: *Given time, all combinations requested would have been presented (including a full grid of $F_{tartget}$-$B_{trigger}$ combinations), but we had to prioritise the most important elements, and here that would be including TAC constraints as a default, and treating banking and borrowing as a sensitivity test.*

While additionally including some empirical MPs as part of the MSE is commendable, the conclusions reached in section 7 were not supported by results included in the report. The MP dependent on the most recent survey estimate of SSB contained a number of control parameters. It appears some of these may have been pre-fixed and not tuned to optimise the MP performance - this needs explanation. The empirical MPs considered were fairly complicated; complicated empirical HCRs often evolve over time from simple HCRs, given requirements for a specific situation. It may be more beneficial to begin with simpler rules (see below) and then modify further as required.

Response: *Section 7 was additional work not requested in the joint request, and was therefore given less priority and attention. Nevertheless, it provides useful analyses for avenues to explore in future MSE exercises.*

Projected recruitment does not closely follow the range of values estimated for the historical period. For example, (Figure 5.4). For SR2 and SR4 future recruitment can occasionally be substantially higher than those estimated historically for the same level of SSB. For the remaining stock-recruitment relationships recruitments appear to be lower and less variable than those estimated historically for the same level of SSB.

Response: *This point is acknowledged, but noting that Figure 5.4 compares future SR pairs under an F=0.2 assumption, whereas past SR pairs were subject to a wider range of Fs and SSBs (see e.g. Figure 4.2). Figure A8.1 shows an extension of Figure 5.11 (application of the MP with control points Btrigger=1.7e6 and Ftarget=0.34 across all OMs) to include historical years for recruitment. See also Figure 4.21 which shows predefined recruitment deviances used in projections for each SR OM compared to a subset of historical years, and Figure 5.1, which shows recruitment for the historical and projection periods for each SR OM under no fishing.*

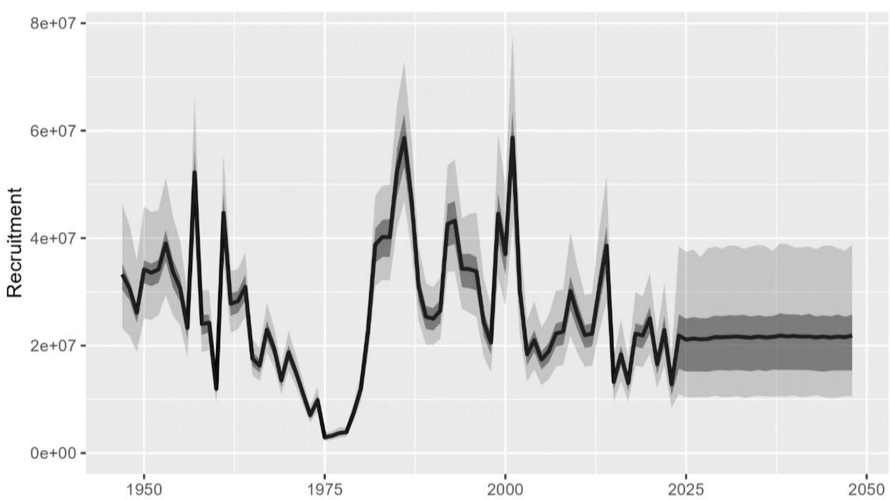


Figure A8.1. Repeat of the recruitment plot in Figure 5.11 (bottom left plot), but extended back in time to show the historical years (see caption to Figure 5.11 for more details).

For best practice, robustness should be tested against the reference set of OMs and not only the base OM. Figure 6.2 (Section 6.1.1), Figure 6.4 & 6.5 (Section 6.1.2), Figure 6.7 (Section 6.2.1), Figure 6.8 (Section 6.2.2) only show results for the base OM. Similarly, for best practice, robustness of the likely final MP should be tested. However, in this MSE, robustness was tested for the 'optimum' high $B_{trigger}$-F_{target} control parameter combination only, which may not be that which stakeholders select.

Response: *We could not produce sensitivity tests for all combinations of F_{target}-$B_{trigger}$, and made the decision to do it for one cell only in the grid, and for the base OM only. The choice of cell was the "optimum" in terms of catch in the precautionary zone – essentially arbitrary, but we had to select something, and that was the obvious one. Ideally, these sensitivity tests should be repeated for the F_{target}-$B_{trigger}$ combination eventually selected.*

Possible improvements for future ICES MSEs

- Acknowledging the tight time constraints of this MSE and the intensive and efficient work of the authors, to maximize the effectiveness of future reviews, a complete trial specifications document and commented code should be available at least two weeks prior to a review meeting. The trial specifications document should be sufficiently detailed to ensure reproducibility of methods.
- It is standard practice in software development that prior to use, computer code is checked for errors by an independent code reviewer. This should happen well before the review of an MSE framework so that the authors can make any appropriate changes and recalculate results. At times during this review, minor errors were found in plotting code leading to the presentation of results that were not representative. Even if these errors are ignorable qualitatively, there is the potential to undermine the review process and negatively impact confidence in the work.
- In order to organize the roles of participants, data collection, processing, OM scoping, MP exploration, code review, MSE review etc, it would be beneficial to develop a *modus operandi* for MSE development (a more detailed roadmap, e.g. Carruthers 2024).
- It is worth noting that a fixed risk level is not always appropriate as it does not take into account the differences in natural fluctuations in the population between alternative OMs. Some OMs have underlying assumptions that correspond with the population naturally dropping to a low level, even in the absence of catch. It would be useful to additionally see Figure 5.2 (annual risk) under F=0. This might indicate higher risk under a no catch scenario for some OMs. Of interest, in particular, is if risk for SR4 would be <5% under a no catch scenario. Taking the risk under F=0 into account might be more appropriate, by, for example, requiring Risk(Catch) < Risk(no catch) + 0.05, i.e. 5% above that which would 'naturally' be expected under a no future catch scenario.
- In cases such as herring where F_{MSY} can be expected to be relatively stable and a relatively precise and accurate abundance survey is available, there may be few benefits to using an assessment as the basis for providing advice every year (i.e. model based MP). A simpler approach is to assume that the survey calibration (q) is constant and simply provide annual TAC recommendations that are a fixed fraction of the most recent survey index (e.g. Butterworth and Rademeyer 2022; ICCAT 2024). Alternatively, control points can be mapped to index levels (via q) and a hockey stick harvest control rule applied that has recent index value on the x-axis (independent

variable) and TAC per index on the y-axis (dependent variable). This is the empirical equivalent of the estimation-model based MP that was the subject of this analysis.

- Exploring the 2-dimensional space of control points can be difficult to communicate in tables. For example, it can be hard to provide an intuitive presentation of the performance trade-offs. In this application risk equivalent MPs were identified at a given risk threshold. Although not necessarily this exact approach, it would be beneficial to establish a *modus operandi* for presenting MSE results for HCRs with multiple control points. Additionally it is desirable to establish standardized MSE results outputs (e.g. box plots, quilt plots) across MSE processes, preferably including a 1-page summary that is familiar to the clients.
- Although it was not explored during this meeting, it is considered best practice to establish Exceptional Circumstances protocols at the time that an MP is adopted. Where possible these should follow a standardized process (comparable principles, data types, probabilities and time horizons).

Opportunities and areas for further investigations

Similar stocks in the region that are subject to important uncertainties might benefit from the MSE approach to identify robust management procedures. For example, the western Baltic herring stock that may have time varying subsidization (immigration / mixing) from the much larger ('10x') North Sea stock.

An established MSE framework provides a flexible and powerful basis for investigating aspects relevant to a changing ecosystem and environmental conditions. This could include an improved characterization of time-varying natural mortality rate arising from the SMS multi-species model (it was noted that there was relatively low variance among simulated M time series).

As explained above, MSEs have often been used to develop and adopt simple empirical MPs that can obtain suitable management performance robustly given a range of plausible uncertainties in system dynamics (e.g. de Moor et al. 2011, Fischer et al. 2023). The use of estimation models within MPs can lead to confusion between aspects of assessment (estimated risk) from those of MSE (known risks calculated from the simulated system). Assessments are often not easily understood by a wide range of stakeholders and when run on newly observed data there can be counter-intuitive impacts on advice. Empirical MPs offer a simpler, more intuitive alternative that could be presented to requestors (e.g. Butterworth 2008, Rademeyer et al. 2007).

The WKMSEHerring MSE framework represents a synthesis of scientific understanding that could be used for directing future research. For example, the cost-benefits of alternative survey designs (the impact of precision and bias on expected performance of an adopted MP). By varying inputs to the operating models and identifying those that have the largest impacts on management performance it may be possible to identify which system uncertainties are most in need of research.

References

Butterworth, D.S. 2008. Some lessons from implementing MPs. In: Fisheries for Global Welfare and Environment, 5th World Fisheries Congress. Eds: K. Tsukamoto, T. Kawamura, T. Takeuchi, T.D. Beard, Jr., and M.J. Kaiser. Pp:381-397. Terrapub.

Butterworth, D.S. and Rademeyer, R.A. 2022. BR CMP as of June 2022. Collect. Vol. Sci. Pap. ICCAT, 79(3): 587-597. Available here.

Carruthers, T.R. 2024. An updated roadmap for MSE development. Collect. Vol. Sci. Pap. ICCAT, 81(2). Available here.

de Moor, C.L., Butterworth, D.S. and De Oliveira, J.A.A. 2011. Is the management procedure approach equipped to handle short-lived pelagic species with their boom and bust dynamics? The case of the South African fishery for sardine and anchovy. ICES Journal of Marine Science 68:2075-2085. https://doi.org/10.1093/icesjms/fsr165.

de Moor, C.L., Butterworth, D.S. and Johnston, S.J. 2022. Learning from three decades of Management Strategy Evaluation in South Africa. ICES Journal of Marine Science 79:1843-1852. https://doi.org/10.1093/icesjms/fsac114.

Fischer, S.H., De Oliveria J.A.A., Mumford, J.D., and Kell, L.T. 2023. Risk equivalence in data-limited and data-rich fisheries management: An example based on the ICES advice framework. Fish and Fisheries 24:231-247. https://doi.org/10.1111/faf.12722.

ICCAT. 2024. Draft Recommendation by ICCAT on conservation and management measures, including a management procedure, for North Atlantic swordfish. PA4_810B/2024. Available here.

Kell, L. T., I. Mosqueira, P. Grosjean, J-M. Fromentin, D. Garcia, R. Hillary, E. Jardim, et al. 2007. “FLR: an open-source framework for the evaluation and development of management strategies.” ICES Journal of Marine Science 64 (4): 640–46. https://doi.org/10.1093/icesjms/fsm012.

Nielsen, Anders, and Casper W. Berg. 2014. “Estimation of time-varying selectivity in stock assessments using state-space models.” Fisheries Research 158 (October): 96–101. https://doi.org/10.1016/j.fishres.2014.01.014.

Nielsen, Anders, Niels T Hintzen, Henrik Mosegaard, Vanessa Trijoulet, and Casper W Berg. 2021. “Multi-fleet state-space assessment model strengthens confidence in single-fleet SAM and provides fleet-specific forecast options.” ICES Journal of Marine Science, July. https://doi.org/10.1093/ICESJMS/FSAB078.

R Core Team, 2024.R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Available from: https://www.R-project.org/.

Rademeyer, R.A., Plaganyi, E.E., and Butterworth, D.S. 2007. Tips and tricks in designing management procedures. ICES Journal of Marine Science 64:618-625.

Scott, F, and Mosqueira, I. 2023. “FLasher FLasher_reference Vignette.” https://github.com/flr/FLasher/blob/master/vignettes/FLasher_reference.Rmd. 2023b. “FLasher r Package.” https://github.com/flr/FLasher.

Appendix to Annex 8

Table 1. Requested changes to the MSE during the December workshop and January 2025 virtual meeting.

Change	Context
Include survival process errors in MSE projections	Projected operating model population dynamics should match the historical reconstruction of the conditioning model that includes process error in the numbers at age. Process errors are now included in the projections of all operating

Change	Context
	models.
OM M1 better accounts for uncertainty from SMS	While the justification for this OM is sound since it provides varying trends in historical M from the multispecies model, the initial implementation led to a relatively narrow range of fits in terms of biomass, numbers and recruitment that was, mostly within the range of outcomes of the base model. The M1 OM was updated to appropriately account for SAM estimation error and more variable M estimates arising from the SMS.
Include a robustness OM with a lower value of natural mortality rate M.	Likelihood profiles with respect to M revealed that lower values were supported by the data and model structure than were included in the reference set of operating models. A lower value of M was determined by likelihood ratio test (i.e. approximately 2 log likelihood points higher than that of the maximum posterior density.
Use post-hoc estimation of lag-1 autocorrelation for OM SR4.	The join point of the segmented regression was estimated at a level of spawning biomass that was considered to be much higher than levels that are typically assumed to correspond to recruitment impairment. This was attributed to numerical instability in the approach for estimating stock recruitment model parameters including the lag-1 autocorrelation. The segmented regression was re-estimated without a lag-1 autocorrelation parameter and this was determined post-hoc from recruitment residuals (a relatively common approach elsewhere).
Relegate OM SR3 (recruitment depensation) to a robustness test.	SR3 indicated little empirical support for depensation and so was removed to the robustness set to avoid duplication (and essentially doubling the weight of those simulations).
Weighting of reference set of OMs by OM group (base, M, SR)	The new reference set of OMs includes a Base model, OM.M1, OM.M2, OM.M3, OM.SR1, OM.SR2, OM.SR4. Weighting is now by category: 1/3 each to base, set of 3 M OMs, and set of 3 SR OMs. Equal weighting will be applied within a category.
Combine all samples from reference set before calculating performance statistics	Initial results included taking the performance statistic from each OM 'category' and then weighting each by ⅓. This would not give correct statistics for percentiles such as risk. Statistics will be recalculated by

Change	Context
	combining all samples from the three categories, with an equal number of samples from each and then calculating, for example, risk from that total sample.
Ensure that the MP does not receive perfect information.	For projected weight-at-age and maturity-at-age: pass average of replicate-specific pre-generated values (generated using the block-sampling procedure) from the OM to the MP. For M: keep most recent HAWG M-at-age (or some recent average) constant for the projection period.
Consider the use of less elaborate harvest control rules for empirical MPs.	The demonstrated empirical MPs included harvest control rules that have multiple discontinuities and control points. This seems unnecessary given that they are not dissimilar from a constant exploitation rate policy. Additionally these were derived from a management setting in South Africa where the shape of the curve was determined by external policy considerations rather than purely on a performance basis.
Recalculate risk percentile based on weighted quantile function rather than the mean of the three groups of operating models	This is the correct calculation.
Summarize absolute performance across MPs of varying control points (Btrigger, Target combinations) along the 5% risk threshold.	By exploring a risk contour it is much easier to show the performance trade-offs among risk-equivalent sets of control points.
Create an accompanying table of metrics for the MPs along the 5% risk threshold	Provides a clearer representation of performance trade-offs.

Table 2. Additional diagnostic checks requested during the December workshop to assist with the review.

Diagnostic	Rationale
SAM fits to data	In order to have confidence that OMs are plausible and consistent with empirical observations, it is necessary to see model fitting. These fits were provided during the December meeting.
Check consistency in SR estimation for OM SR2	The model estimates of historical spawning stock

Diagnostic	**Rationale**
(e.g. MLE fit too precise a range)	biomass and recruitment produce a relatively poorly defined stock recruitment relationship. The segmented regression model is estimated from the MLE fit to those data and those MLE fits may lead to a relatively narrow range of SR relationships relative to the uncertainty represented in the historical model estimates of SSB and recruitment. This is less of an issue since B_{lim} reference points are not derived directly from the mean stock-recruitment curve.
Simulation test the post-hoc approach to quantify lag-1 autocorrelation in recruitment deviations.	Simulation testing revealed that the post-hoc derivation of lag-1 autocorrelation provided suitably precise and accurate estimates of the true simulated value.
Arbitrarily increase observation error in survey indices	Confirm that the management procedures respond as expected to changes in the quality of simulated observed data.
Plot the MP estimation model's estimates of observation error over the projection.	Check that the estimator of the MP is working as expected and that estimated observation errors are stable and approximately match the simulated error.
Include plots with a timeline from beginning of assessment to end of projection period, showing historical estimates (+variance) and projections (median + e.g. 95% CI + a few worm plots).	MSE projections should not show obvious discontinuities from historical reconstructions to future projections in terms of e.g. biomass / recruitment etc. Future simulated data should be within the range of that observed historically if observation and process errors are correctly accounted for.
Check inexplicable trends in plots	It was a plotting error.

Annex 9: WKMSEHerring Scoping Report

The scoping report can be found as a second document attached to this item on the ICES library, downloadable separately.