

# BENCHMARK WORKSHOP ON BALTIC PELAGIC STOCKS (WKBBALTPEL)

VOLUME 5 | ISSUE 47

ICES SCIENTIFIC REPORTS

RAPPORTS  
SCIENTIFIQUES DU CIEM



## 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](http://www.ices.dk)  
[info@ices.dk](mailto:info@ices.dk)

ISSN number: 2618-1371

This document has been produced under the auspices of an ICES Expert Group or Committee. The contents therein do not necessarily represent the view of the Council.

© 2023 International Council for the Exploration of the Sea

This work is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). For citation of datasets or conditions for use of data to be included in other databases, please refer to ICES data policy.



# ICES Scientific Reports

Volume 5 | Issue 47

## BENCHMARK WORKSHOP ON BALTIC PELAGIC STOCKS (WKBBALTPEL)

### Recommended format for purpose of citation:

ICES. 2023. Benchmark Workshop on Baltic Pelagic stocks (WKBBALTPEL).  
ICES Scientific Reports. 5:47. 350 pp. <https://doi.org/10.17895/ices.pub.23216492>

### Editors

Harald Gjøsæter • Johan Lövgren

### Authors

Mikaela Bergenius Nord • Massimiliano Cardinale • Julie Olivia Coad Davies • Maria Golovaneva •  
Nicolas Goñi • Julita Gutkowska • Stefanie Haase • Alex Hansell • Kristiina Hommik • Jan Horbowy  
Olavi Kaljuste • Francesco Masnadi • Maria Pierce • Jukka Pönni • Nuno Prista • Ivars Putnis • Eros  
Quesada • Tiit Raid • Katja Ringdahl • Malin Skog • Szymon Smolinski • Marie Storr-Paulsen • Maciej  
Tomczak • Vanessa Trijoulet • Morten Vinther



**ICES**  
**CIEM**

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

# Contents

i	Executive summary .....	iii
ii	Expert group information .....	v
1	Introduction.....	1
2	Herring ( <i>Clupea harengus</i> ) in subdivisions 25–29 and 32, excluding the Gulf of Riga (central Baltic Sea).....	2
	2.1 Issue list .....	2
	2.2 Stock ID and substock structure .....	5
	2.3 Ecosystem drivers .....	6
	2.4 Stock Assessment.....	6
	2.4.1 Data.....	6
	2.4.1.1 Landings and discards .....	6
	2.4.1.2 Biological Information.....	10
	2.4.1.3 Tuning indices .....	12
	2.4.2 Assessment model .....	12
	2.4.2.1 General description of the modelling framework .....	12
	2.4.2.2 Exploration of candidate reference model configurations and selection of the reference mode .....	13
	2.4.2.3 Additional diagnostics of the selected reference model .....	21
	2.4.2.4 Detailed description of the selected reference model configuration and diagnostics .....	26
	2.4.2.5 Trends in SSB, F and R of the reference model.....	36
	2.4.2.6 Additional runs tested .....	37
	2.4.3 Ensemble model .....	39
	2.4.3.1 Why use an ensemble model? .....	39
	2.4.3.2 Parameter levels .....	40
	2.4.3.3 Model weighting .....	41
	2.4.3.4 Delta-MVLN estimator .....	41
	2.4.3.5 Model results .....	41
	2.5 Short-term projections .....	44
	2.6 MSY reference points.....	44
	2.6.1 Performance Evaluation Criteria.....	46
	2.6.2 Results.....	47
	2.7 Roadmap for the future work on the stock assessment model for Central Baltic Herring. ....	55
	2.8 Progress Feco .....	58
	2.9 References .....	59
	2.10 Minority statement on the estimation of $B_{lim}$ for CBH stock .....	62
3	Sprat ( <i>Sprattus sprattus</i> ) in subdivisions 22–32 (Baltic Sea) .....	65
	3.1 Issue list .....	65
	3.2 Ecosystem drivers .....	66
	3.3 Stock Assessment.....	66
	3.3.1 Data.....	66
	3.3.1.1 Landings and discards .....	67
	3.3.1.2 Biological Information.....	69
	3.3.1.3 Tuning indices .....	73
	3.3.1.4 Summary of data input to the assessment model .....	73
	3.3.2 Assessment model .....	82
	3.3.3 Model settings .....	87
	3.3.4 Final model .....	88
	3.4 Short-term projections .....	96

	3.5	MSY reference points.....	97
	3.5.1	Settings .....	97
	3.5.2	Estimation of $B_0$ , $B_{lim}$ and $B_{pa}$ .....	98
	3.5.3	Estimating $F_{MSY}$ , $F_{p05}$ and $F_{pa}$ .....	99
	3.5.4	Estimating MSY $B_{trigger}$ .....	99
	3.5.5	Comparison with previous reference points.....	100
	3.6	Future research and roadmap .....	100
	3.7	Progress Feco .....	100
	3.8	References .....	100
4		Herring ( <i>Clupea harengus</i> ) in Subdivision 28.1 (Gulf of Riga) .....	101
	4.1	Issue list .....	101
	4.2	Stock ID and sub-stock structure .....	102
	4.3	Ecosystem drivers .....	102
	4.4	Stock Assessment.....	103
	4.4.1	Data.....	103
	4.4.1.1	Landings and discards .....	103
	4.4.1.2	Biological Information.....	105
	4.4.1.3	Tuning indices .....	111
	4.4.1.4	Summary of data input to assessment model .....	113
	4.4.2	Assessment model .....	114
	4.4.2.1	SAM model configuration .....	114
	4.5	Short-term projections .....	119
	4.6	MSY reference points.....	120
	4.6.1	Settings .....	120
	4.6.2	Estimating $B_{lim}$ , $F_{lim}$ and PA reference points .....	120
	4.6.3	Determining the SRR for $F_{MSY}$ calculations .....	121
	4.6.4	Estimating $F_{MSY}$ , $F_{p.05}$ and $F_{pa}$ .....	122
	4.6.5	Estimating MSY $B_{trigger}$ .....	123
	4.6.6	Comparison with previous reference points.....	124
	4.7	Future research and roadmap .....	124
	4.8	Progress Feco .....	124
	4.9	References .....	124
	4.10	Annex 1. Gulf of Riga herring SAM model full settings .....	125
Annex 1:		WKBBALPEL Resolution.....	131
Annex 2:		List of participants .....	133
Annex 3:		Reviewer report for WKBBALPEL .....	135
Annex 4:		Updated Biological Reference Points for Central Baltic Herring.....	144
Annex 5:		Working documents.....	146
Annex 6:		Updated stock annexes.....	350

## i Executive summary

Three pelagic stocks in the Baltic Sea, Central Baltic Herring (CBH; her.27.25-2932), Sprat (spr.27.22-32) and Gulf of Riga Herring (GOR, her 27.28) were examined during this benchmark. Work was prepared well in advance and 3 preparatory online meetings took place between the Data Evaluation Workshop and the actual benchmark to follow up progress. The benchmark meeting advanced really well until reference points were discussed. There were follow up online meetings on 3rd and 13th March to further discuss, and agree, on reference points for the three stocks.

All three stocks that was benchmarked were previously assessed with XSA, and hence a main point with the benchmark was to change the assessment model. Two of the stocks changed to SAM (GOR and sprat) whereas the CBH changed to stock synthesis (SS3). For the GOR and sprat the reference points were estimated following the ICES procedure using EQSIM. For the CBH the reference points were obtained using a Management strategy evaluations (MSE), a method previously used for one of the Northern shrimps stocks in ICES (pra.27.3a4a).

There was an extensive issue list for the CBH and although not all of the issues were addressed, significant progress was made. Additional survey indices, maturity at age, weight in catch, adding age reading error, landings corrected for misreporting (Danish landings only) were all investigated and updated time series were included in the new assessment. SS3 was chosen as main assessment method and there was extensive model development work carried out in advance and presented at the meeting.

An ensemble model was developed to incorporate uncertainty in natural mortality. Three scenarios of natural mortalities were developed from rescaling the natural mortalities from a multispecies model for the Baltic (SMS) with life history parameters. An objective weighing approach was used based on model performance statistics diagnostics to develop the final assessment.

The reference points were estimated using Management Strategy Evaluations (MSE).  $B_{lim}$  for this stock was set at 15% of  $B_0$ . The group agreed that, according to the results of the MSE, a combination of an  $F_{MSY}$  proxy at  $F_{35\%}$  combined with a  $B_{trigger}$  at 60% of  $B_{35\%}$  satisfied the criterion of being above  $B_{lim}$  with 95% probability, achieves a realised target  $F$  smaller than  $F_{MSY}$  and generates catches higher than  $MSY$ . It is important to note that several combination of  $F$  target and  $B_{trigger}$  achieve long-term catches equal to those under  $F_{MSY}$  and result in a median  $B$  that is significantly larger than  $B_{MSY}$  despite a lower  $F$ . However, these scenarios do not match with the ICES interpretation of  $MSY$ , for which the scenario with the highest  $F$  (which fulfils the MSE performance and ICES precautionary criteria) is selected.

This option results in an  $F_{MSY}$  proxy of  $F_{B35\%}$  and a  $MSY_{B_{trigger}}$  reference point of 60% of  $B_{35\%}$ .

At the meeting a roadmap towards the next benchmark for CBH was set up in order to deal with issues that were not covered at WKBBALPEL including, stock identity, multiarea assessment, catch data corrected from misreporting from more countries.

For the Sprat in the Baltic, the issue list was almost as extensive as for the CBH. Similarly, to the CBH stock, significant progress was made during the Benchmark, the progress includes:

- The natural mortalities were updated with new  $M$  values from the updated SMS runs.
- Include the Danish data that was corrected for misreporting
- Additional survey indices
- Moving from an XSA assessment model to a SAM model.

The reference points were calculated using EQSIM.  $B_{lim}$  was set to the biomass that produces 50% of maximal recruitment. There is one thing left on the issue list for the next benchmark also for this stock; to include catch data corrected for misreporting from countries other than Denmark.

For the GOR herring significant achievements were performed during the Benchmark meeting. The assessment model was changed from XSA to SAM, a trap net tuning series previously used in the assessment was investigated and excluded, maturity at age values were updated and the ages considered for  $F_{bar}$  were updated.

The reference points were calculated using EQSIM. There was no evident SR relationship, hence  $B_{lim}$  was defined based on  $B_{pa}$ .

## ii Expert group information

<b>Expert group name</b>	Benchmark Workshop on Baltic Pelagic stocks (WKBBALPEL)
<b>Expert group cycle</b>	Annual
<b>Year cycle started</b>	2022
<b>Reporting year in cycle</b>	1/1
<b>Chairs</b>	Johan Lövgren, Sweden Harald Gjøsæter, Norway
<b>Meeting venue and dates</b>	Online, on 14-18 November 2022 (data evaluation workshop) ICES Headquarters, Copenhagen, on 13-17 February 2023



# 1 Introduction

WKBBALPEL took place in Q4 2022 and Q1 2023, there was 32 participants from 9 countries including one participant from the fishing industry.

The two chairs were from Sweden and Norway, the reviewers were from Denmark and USA. There were two online sessions between the Data evaluation workshop (DEW) to decide on outstanding issues before the Benchmark.

In addition, two online meetings were held after the benchmark as well to finalise the discussions about the reference points

Participants were asked to declare any conflict of interest (none were declared) and reminded of meeting etiquette at the start of each meeting.

Eight working documents were presented by participants and the contents of these were either included in the main report sections or have been appended to this report.

## 2 Herring (*Clupea harengus*) in subdivisions 25–29 and 32, excluding the Gulf of Riga (central Baltic Sea)

The following working documents supports the summarised texts in this report.

- WD1\_MultiSpecies\_M for the central Baltic herring her.27.25-2932 and Baltic sprat spr.27.22-32
- WD2\_Correction of CBH catch data for the central Baltic herring stock her.27.25-2932
- WD3\_Sampling design, data storage and data edits for the central Baltic herring stock her.27.25-2932 in Sweden
- WD4\_Summary Report for the 2022 otolith exchange for the central Baltic herring stock her.27.25-2932 )ID 449)
- WD5\_WEST\_WECA for the central Baltic herring stock her.27.25-2932
- WD6\_Maturity Mat, M final for the central Baltic herring stock her.27.25-2932
- WD7\_Natural Mortality estimated using Life History based methods for the central Baltic herring stock her.27.25-2932
- WD8\_Working Document catch data of Baltic sprat and central Baltic herring\_all countries

### 2.1 Issue list

The issue list compiled for the meeting is detailed below in Table 2.1. An extra column 'Conclusions and outcomes' has been added to provide concluding remarks or outcomes for each issue.

**Table 2.1. Herring (*Clupea harengus*) in subdivisions 25–29 and 32, excluding the Gulf of Riga (central Baltic Sea). Issue list for the WKBBALPEL meeting.**

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed /are these available / where should these come from?	Re-search/WG input needed	Main Person	Conclusions and outcomes
Stock identity	Mixing of Western Baltic spring spawners and CBH components in SD 24–26.	Test the of different of methods	Genetic samples, morphometrics, otolith shapes etc.	Project from WKMIXHER	WKMIXHER	Results not ready in time for this benchmark.
Tuning series	BIAS data. Do we have new bias data from SD 32 that could be used in the assessment?	Produce index	Index produced by WGBIFS members	WGBIFS	Olavi Kaljuste (WGBIFS)	An index included SD 32 was provided and included some error corrections. This index starts in 2006 as opposed to the old index that started in 1991.  It was decided not to use the index calculated by StoX.

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed /are these available / where should these come from?	Re-search/WG input needed	Main Person	Conclusions and outcomes
		Investigate performance of new index			<u>Max Cardinale</u> , Mikaela Bergenius Nord	The new index including SD 32 performed well and replaced the old BIAS index in the assessment
Biological Parameters	Mean weight in the stock. Equals currently mean weight in the catch!	Compare WECA with WEST from the BIAS survey (base)	Mean weights at age from commercial and BIAS data		<u>Stefanie Haase</u> , <u>Szymon Smolinski</u>	Weight-at-age in the stock (WEST) and in the catch (WECA) generally showed a high correlation but differences exist in time, space and by age and in the work towards a spatially explicit assessment model further analyses need to be undertaken
	Maturity currently constant (time invariant) and potentially outdated	Investigate the need for a time variant maturity ogive. Data analysis and sensitivity analyses.	Maturity data from different countries – Data call		<u>Francesco Masnadi</u> Mikaela Bergenius Nord, Szymon Smolinski, Max Cardinale	The time-invariant maturity ogive previously used in the assessment was replaced by a time-varying maturity ogive
	Natural Mortality (base) currently constant over time and ages (SMS provides predation mortality)	New key from SMS (predation mortality)	WGSAM to produce a new key run	WGSAM	Morten Winter (WGSAM)	New SMS runs were produced in which the residual M was re-scaled according to changes in residual M estimates derived from the Barefoot Ecologist’s Toolbox
		Investigate age dependent natural mortality (residual M)	Biological parameters – review and data call		<u>Francesco Masnadi</u> Mikaela Bergenius Nord, Szymon Smolinski, Max Cardinale	A time-varying residual M vector was estimated using the Barefoot Ecologist’s Toolbox A significant breakpoint in the natural mortality time series was detected in 2000 and the mean value calculated before and after the breakpoint was used to re-scale the assumed annual residual M in SMS
Age reading	Quality of age reading	Comparison of age readings between countries	Reference otolith collection	Age reading experts	Julie Olivia Davies	Good overall agreement. No further action needed.  An age error matrix was provided

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed /are these available / where should these come from?	Re-search/WG input needed	Main Person	Conclusions and outcomes
Misreporting of herring and sprat.	Misreporting of herring and sprat in the mixed catches.	Misreporting estimates from relevant countries	Logbooks data, VMS data reports from landing controls	Project by RCG	ISSG	A new times series of catch data were produced by Denmark and included in the model. Sweden also produced alternative time series, but these were not yet evaluated due to the lack of misreporting information from some countries. The issue of misreporting was thus at large postponed and will be revisited as part of the road map for the next benchmark. The ISSG produced a working document in which most countries outline if their catch data need to be reconstructed due to misreporting or not.
Assessment method	A change to the SS model framework instead of the currently used XSA.	Configuration and subsequent testing of the SS model.	CANUM, WECA, maturity, mortality, etc	Model developer	<u>Max Cardinale</u> , Francesco Masnadi, Mikaela Bergenius Nord	A new SS model was produced and accepted at the benchmark. It includes three different models varying only by their assumptions of M, and stitched together in an ensemble approach

The natural mortality used varied between years and ages as an effect of cod predation and life history of the species. Specifically, length-at-age data from 1984 to 2021 has been used to derive VB parameters ( $L_{inf}$ ,  $k$  and  $t_0$ ) to be used as input parameters to derive proxy of M1 (residual mortality) to be used in SMS (WD1\_MultiSpecies\_M for the central Baltic herring her.27.25-2932 and Baltic sprat spr.27.22-32). Since length-at-age data reveals a decreasing trend all along the timeseries, the VB equation parameters per year were used as input value in the Barefoot Ecologist's Toolbox ([http://barefootecologist.com.au/shiny\\_m](http://barefootecologist.com.au/shiny_m)) to produce a time-varying natural mortality vector (WD6\_Maturity Mat, M final for the central Baltic herring stock her.27.25-2932). A significant breakpoint in the natural mortality time series was detected in 2000 so the mean value calculated before and after the breakpoint (M before 2000: 0.28, M after 2000: 0.38; WD6\_Maturity Mat, M final for the central Baltic herring stock her.27.25-2932) have been used to re-scale the assumed annual M1 in SMS (scenarios for "likely" M1 presented in WD1\_MultiSpecies\_M for the central Baltic herring her.27.25-2932 and Baltic sprat spr.27.22-32).

## 2.2 Stock ID and substock structure

The herring in the Baltic Sea is divided into several spring- and a smaller number of autumn spawning populations (Ojaveer, 1981), which have also been shown to be genetically distinct (Han *et al.*, 2020). Springs- and autumn-spawning herring have different migration patterns, but in general, they move between coastal spawn- and nursery areas and feeding- and overwintering areas (Aro, 1989). Herring generally overwinter in deep water, below the halocline which is found at about 50-60 m depth, where the water is a little warmer in winter. Some groups of individuals, especially juvenile fish, may remain in the shallow waters of the archipelago (Ojaveer, 2003; Kaljuste *et al.*, 2009; Polte *et al.*, 2017). During the spring, the herring then migrates from the wintering areas in the open sea towards shallow coastal areas for spawning and to search for food. A compilation of tagging data shows that the herring along the Swedish Bothnian coast generally make rather short migrations out to the open sea, while the herring in the central Baltic Sea can sometimes take slightly longer migrations to more southern offshore areas where temperatures and food conditions are favourable in winter (Aro, 1989, 2002). There are also more stationary spawning stocks in some coastal areas (Aro, 2002).

The management unit (ICES subareas 25–29 and 32) consists of a number of smaller populations, which are more or less spatially separated and differ in, among other things, growth, and sexual maturity (Popiel, 1958; Ojaveer, 1989). Until 1990, ICES carried out separate stock estimates for two management units: herring in subdivisions 25-27 and herring in subdivisions 28-29+32 (ICES 2002). But they have since been merged, as it was not possible to collect biological information for all areas, and the possible genetic difference between the populations was not fully mapped. Analyses of the consequences of the pooling in the stock estimate for the smaller populations show, for example, that fishing mortality may be higher, and the relative biomass lower, in some of the populations compared to the values in the analysis of the whole stock (Raid *et al.* 2016). Until now, however, it has been considered that the complex stock structure in the central Baltic does not have a major impact on the view of the overall stock dynamics (ICES 2013). Repeated reports in recent years from the Baltic fishery about a decreasing supply of large herring in different areas of the Baltic Sea, combined with the generally decreasing average weight of the stock, have however raised the question of spatial coastal management of herring in both the central Baltic Sea and the Gulf of Bothnia. This in turn has once again raised the question of the herring's population structure and important spawning areas. The availability of new genetic technology enables the identification of spawning components with a finer local resolution than was previously possible (Han *et al.*, 2020). A research effort has therefore been started in Sweden (by Stockholm University in collaboration with SLU Aqua) to map the population structure of the herring in the eastern parts of the central Baltic Sea and the Gulf of Bothnia. Similar research

projects (e.g. PopHerr project) combining different methods ( e.g. genetics, morphometrics, parasitological analysis, otolith markers, etc.) for the discrimination of the stock components in the Southern Baltic are ongoing in Poland (led by the National Marine Fisheries Research Institute).

The central Baltic herring is known to be dominated by a northern and a southern component. A recent workshop (WKMIXHER, ICES 2018) showed how the latter shares numerous characters with the adjacent western Baltic herring stock. Its growth and otolith shape are more similar to those of herring of western origin than to fish from the northern component. Based on only growth, a high proportion of fast-growing herring is found in SD25 and especially in the westernmost rectangles but it remains unclear if those fish are part of the southern component of the central Baltic or if they are the results of extensive mixing with the western Baltic herring. Preliminary analyses suggest a progressive genetic differentiation along the entire southern Baltic coasts from SD24 to SD26 rather than a clear-cut division between different assessment units.

Herring from the Skagerrak, Kattegat, and the southwestern Baltic Sea mix with the stock in the central Baltic Sea and are thus caught there. On the opposite, the presence of the Central Baltic herring (albeit in small numbers) out of the Central Baltic - throughout the Kattegat–Skagerrak–North Sea was observed in a recent study (Bekkevold *et al.*, 2023). Data presented in this study supported the notion (ICES 2018) that mainly the southern component moves out of the Baltic Sea to feed, while the northern component, which grows more slowly, was rarely encountered outside the Baltic Sea (Bekkevold *et al.*, 2023).

The problem of separating the Central Baltic herring stock from the western Baltic spring spawning herring stock is related to understanding if the southern component should be considered together with the western Baltic herring, maintained with the central Baltic herring, or if it should be considered separately. Depending on the task, the methodologies reviewed for stock identification could be promising or insufficient. The stock discrimination between the Central Baltic herring and the Gulf of Riga herring is less problematic as these two stocks are more clearly distinguishable based on the body and otolith morphometrics and other biological features.

In conclusion, there is not sufficient evidence at the time of this benchmark to confirm the division of the current central Baltic herring stock into smaller distinct components that may require separate assessment and management. There is a need for a separate research project with the objective to validate herring assessment units and look for operational methods to separate them in mixed catches.

## 2.3 Ecosystem drivers

Not explicitly discussed, but ecosystem components (predation of herring by cod) are included in the natural mortality estimates used in the assessment (WD1\_MultiSpecies\_M for the central Baltic herring her.27.25-2932 and Baltic sprat spr.27.22-32).

## 2.4 Stock Assessment

### 2.4.1 Data

Below we outline fishery-dependent (landings and discards), fishery-independent, and biological data that are used as input data in the different assessment models.

#### 2.4.1.1 Landings and discards

The total reported catches by country used for this benchmark, which also include the fraction of the Central Baltic Herring that is caught in the Gulf of Riga (SD 28.1), were kept the same as

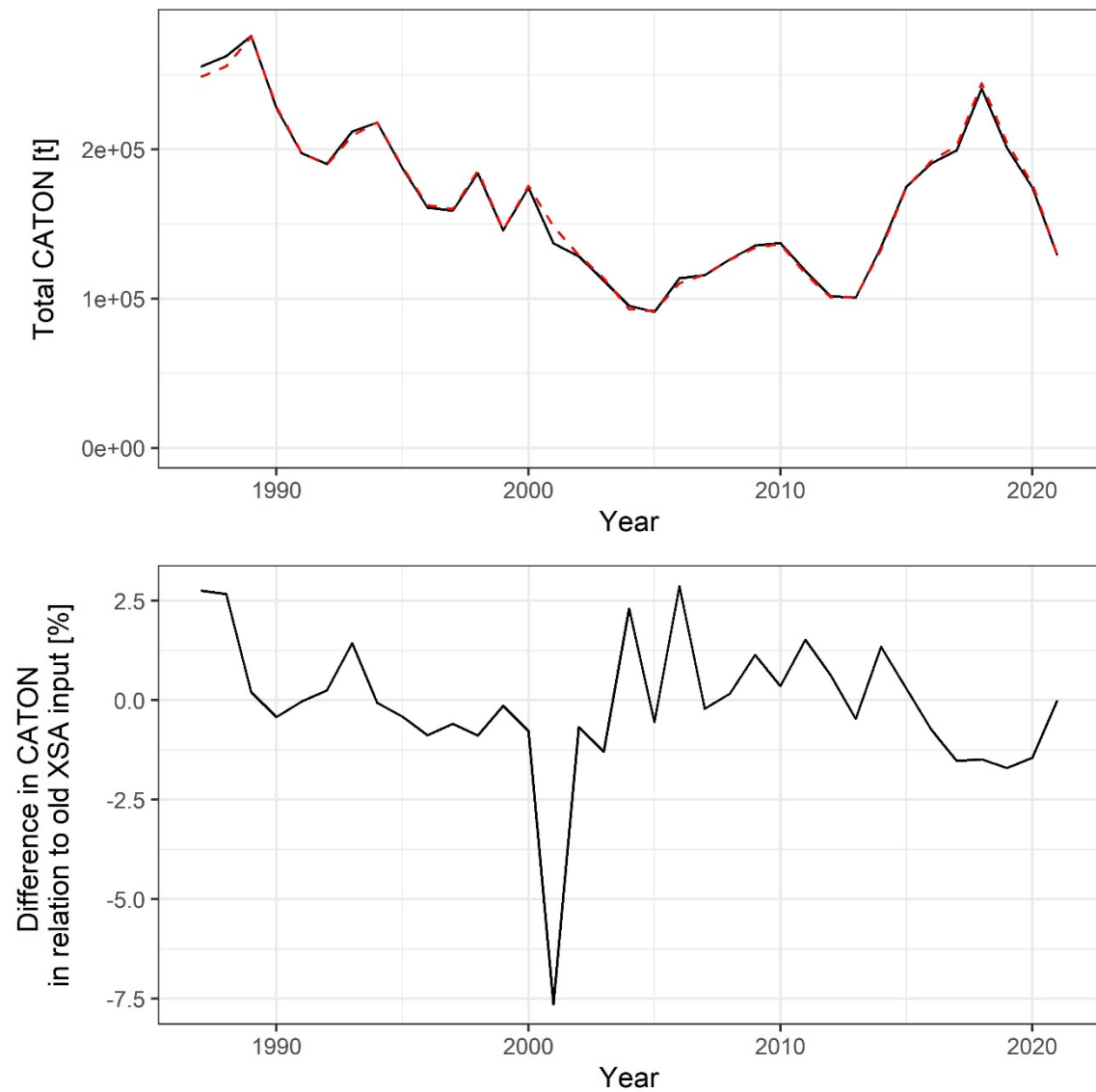
in the latest 2022 assessment (Table 4.2.1 in ICES 2022) with one exception (see section 3.4.1.1.1 about updated Danish Landings) (WD2\_Correction of CBH catch data for the central Baltic herring stock her.27.25-2932). This means that the catch data were not scrutinized or discussed in detail at this benchmark.

Discarding at sea is as in the previous assessments continued to be regarded as negligible.

There were intentions before the benchmark to gather catch data by year, and quarter from the different countries and include these in the new model. In meetings with the countries before a proposed data call it became apparent however, that much more preparation time would be needed for the countries to deliver data with this resolution. It was therefore decided to postpone this request and instead decide on a roadmap for this work to be undertaken over the next few years (see section 3.7).

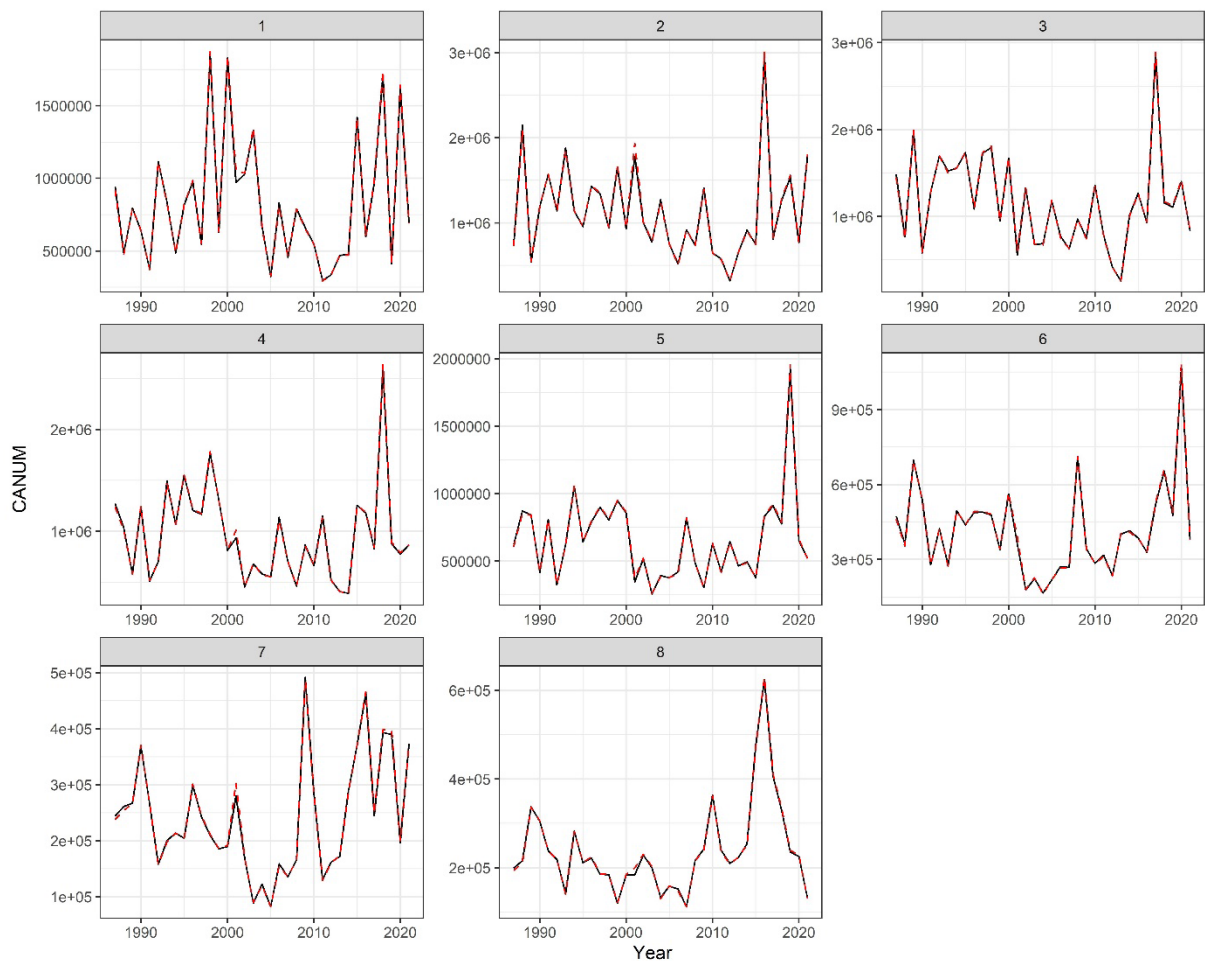
#### 2.4.1.1.1 Updated Danish Landings

Data provided by Denmark to the benchmark workshop represent old and corrected Danish catches from 1987 onwards. Thus, no changes have been made to the input data before 1987. Old XSA input files (CANUM, CATON from 2022 WGBFAS assessment) were used to recalculate CANUMs and CATON. Old CATONs were corrected using the difference between the old and corrected Danish catch time series. The ratio between the old and corrected CATON (including the corrected Danish catch time-series) was used to up-scale or down-scale the CANUMs proportionally. For most years, the correction of old CATON (old XSA input) was within the range  $\pm 3.0\%$  (Figure 2.1). Only in the year 2001, the CATON value was reduced by  $\sim 7.5\%$  in relation to the initial value. This year, Denmark reported initially 15786 tons, while the corrected value was 4462 tons, significantly lowering the contribution of Danish catches and lowering CBH CATON. With respect to the interannual variation of the time-series, the corrected CANUMs were similar to the initial time series and differences were visible only in certain years (e.g. 1987, 1988, 2001, 2006) and particular age groups dominating in the catches (Figure 2.2). Old and corrected CATON and CANUM data are given in Annex 5 Working Document 2 (WD2).



**Figure 2.1.** Total catch in tonnes (CATON) of the Central Baltic herring (upper panel) with indicated old (red, dashed line) and corrected (black, solid line) values. Difference between old and recalculated CATON (lower panel).





**Figure 2.2.** Total catch in numbers (CANUM) of the Central Baltic herring by age group with indicated old (red, dashed line) and corrected (black, solid line) values. Note that the lines are strongly overlapping.

#### 2.4.1.1.2 Swedish data

Working Document 3 (Annex 5) provides supplementary notes and methodological details related to Swedish commercial data on Central Baltic Herring (her.27.25-2932) provided to the 2023 ICES Benchmark Workshop on Baltic Pelagic Stocks (WKBBALPEL). Aspects such as changes through time in sampling designs used to select the samples and the methods used in biological analyses are detailed. Details are also given on data edits made prior to the submission, procedures used to fill missing values and a set of additional biological variables that were not requested but are available for exploration in future benchmarks. Finally, a set of analysis are highlighted for future improvement of biological estimation of commercial catches of the stock.

#### 2.4.1.1.3 Misreporting

For many years there has been discussions about the species misreporting of herring and sprat in the Baltic (ICES 2022). The ISSG consequently made an attempt to provide the benchmark with corrected time-series of catch data for which species misreporting had been corrected. It was concluded at the data compilation meeting of this benchmark that the issue of misreporting could not be addressed adequately by all the countries in time for the benchmark and that the issue need to be postponed. The working document WD8 (Annex 5) outlines the approach taken by countries so far to analyse if there are errors in the time-series of catch data due to inadequate reporting of species and/or other reasons and if the countries foresee that alternative time-series of catch should be provided. Denmark and Sweden provided alternative time-series of catch, of which the time-series of catch from Denmark was included in the benchmark assessment.

## 2.4.1.2 Biological Information

### 2.4.1.2.1 Catch in numbers

All countries except Germany (with less than 1% of the total catches) provide age compositions of their major catches (caught in their waters by quarter and subdivision). The non-sampled catches are generally assumed to have the same age composition as those sampled in the same subdivision and quarter.

The subsequently computed catch in numbers were kept the same as in the latest 2022 assessment (Table 4.2.5 in ICES 2022) with one small change due to the updates Danish Landings). With the exception of rescaling the original CANUM (WD2 in Annex 5) the data were not scrutinized or discussed in detail at this benchmark.

### 2.4.1.2.2 Calibration exercise

*WD4: SmartDots Summary Report for the 2022 exchange for the central Baltic herring stock her.27.25-2932*

This summary gives the results presented to the BWKBALTPPEL (ICES benchmark workshop on Baltic Pelagic stocks) 2023 data compilation meeting in November 2022. Age error matrices (AEM's) were provided following a request from the group. A single matrix per ICES Subdivision (SD) was provided as well as a combined AEM for all SD's.

The full report can be found <https://smartdots.ices.dk/ViewEvent?key=449>

The 2022 exchange for the central Baltic herring stock her.27.25-2932 took place via the SmartDots platform between May and October 2022. The exchange was organised following a request from WGBFAS and in preparation for the 2023 benchmark of the stock. Fifteen readers from nine countries took part (Denmark, Poland, Sweden, Germany, Latvia, Lithuania, Estonia and Finland); twelve "advanced" readers (providing age data for assessment) and 3 "basic" readers (do not provide age data for assessment). 163 otoliths images, covering ICES SD25, 26, 29 and 32 were provided by Poland and Finland and uploaded to the SmartDots platform. The aim was to include samples from all SD's included in the stock assessment but the otoliths from SD27 were not included due to lack of resources within the lab photographing the otoliths. Images of whole otoliths from SD's 25 (n = 27) and 26 (n = 30) were provided by Poland. For SD 29, images of sectioned and stained and whole otoliths from the same fish (n = 24) plus additional images (n = 18) of sectioned and stained otoliths were provided by Finland. For SD32, images of sectioned and stained otoliths (n = 40) were provided by Finland. The aim was to cover all areas, quarters and age groups for each ICES SD's used in the stock assessment but this aim was not reached.

This summary report presents the results based on advanced readers only (those who provide age date for stock assessment purposes); for SD 25 overall PA was 93%, CV was 8% and relative bias -0.04; for SD 26, overall PA was 85%, CV was 9% and relative bias -0.01; for ICES SD 29 overall PA was 89%, CV was 12% and relative bias 0.06; for ICES SD 32, (based on the ATAQCS analysis) overall PA was 70%, CV was 7% and relative bias 0.38. The analysis was carried out by ICES SD as not all readers are experienced in reading otoliths from all areas and the growth patterns observed in the otoliths vary greatly from north to south, meaning a correct interpretation by readers not experienced with samples from another SD would introduce bias in the results. *WD4: SmartDots Summary Report for the 2022 exchange for the central Baltic herring stock her.27.25-2932*

### 2.4.1.2.3 Weight-at-age

WECA (weight-at-age) files per ICES SD were extracted from Intercatch (IC) and selected for quarter 4 to ensure comparability between WECA and WEST

(<https://intercatch.ices.dk/login.aspx>). WEST was calculated based on the survey catches of the BIAS which were uploaded at the Acoustic Trawl Database (<https://www.ices.dk/data/data-ports/Pages/acoustic.aspx>). Individual fish, which are sampled length-stratified in the BIAS, were used to calculate an age-length key (ALK) per SD and year. Further, a length-weight relationship (LWR) was estimated per SD and year. The number of fish per length class in the catches was standardized to 30-minute hauls and total length distributions were calculated using standardized values. Length were transformed to weights by the LWR and an ALK was applied to get weight-at-ages per SD, year and species. WECA and WEST were compared for ICES SDs 25-32 and the years 2015-2021 as these years were available at the Acoustic Trawl Database.

Weight-at-age in the stock (WEST) and in the catch (WECA) generally shows a high correlation for herring (Figure 1, Figure 2 in WD5 in Annex 5). Differences between WEST and WECA are larger for younger age groups. The difference between WECA and WEST per age group and SDs over the years. Differences are particularly high in SDs 25 and 26 and for age group 1. In most years, WECA is larger than WEST in SD 25 and 26 (Figure 3 in WD5 in Annex 5). In contrast, WEST is larger in most years for age group 1 and 2 in SDs 27-32.

In conclusion, WECA and WEST in quarter 4 generally show a high correlation, differences, however, occur especially in ages 1 and 2 and for SDs in the southern Baltic Sea for herring and northern Baltic Sea for sprat. Weights in the catch are compared to weights in the survey only distinctively larger for herring in all age groups in SD 25 and 26.

Differences in weight between the survey and commercial catches might occur due to the different selectivity of the gears and unequal geographical coverage of fishing effort and surveys. One further explanation for the difference in weight could be that the BIAS samples are weighted and measured fresh while commercial samples are often frozen before they are measured and weighted. Clupeids shrink when they are frozen, and this effect can be higher for smaller fish than for larger specimens (e.g. Santos *et al.*, 2009).

A quarterly coverage of survey weights is currently unavailable. Therefore, only the weight-at-age from quarter 4 was compared. As SSB is calculated in the beginning of the year, weights from the quarter 4 BIAS could be used as WEST estimates. There is, however, a large difference in weights-at-age between the different SDs in both WEST and WECA. As there is currently no spatially resolved stock assessment in place, one WECA value is used for all SDs. Further investigation is needed on how to raise weight-at-ages from the survey based on the spatial distribution of the stocks to implement weight-at-age from the survey as WEST.

#### 2.4.1.2.4 Maturity

According to evidence of a spatial-temporal trend in maturation of herring stock, new analyses on maturity ogive were conducted on proportion of mature data from 1984 to 2021 using generalized linear mixed model (Annex 5 WD6). Based on observations, and in line with previous analyses (ICES, 2013), maturity ogive was produced based only on the spring spawning part of stock. Maturity ogives to be used in the stock assessment were produced as predictions by area and year from the best model presented in Annex 5 WD6. However, since the current stock assessment configuration did not allow results to be used by area, predictions were averaged over the total area using spatial distribution by BIAS survey to obtain final matrix of percentage of mature by age and by year (Table 3 in WD6 in Annex 5). Moreover, a revision of the dataset performed in preparation of the benchmark revealed several errors in the official data. Hence, further exploration of the identified issues (section 2 and 4.1 in WD6 in Annex 5) is strongly recommended. Nevertheless, time-varying component addressed in this benchmark are considered a good improvement compared to the constant maturity ogive previously adopted for the stock.

#### 2.4.1.2.5 Life history parameters and Natural Mortality

The natural mortality used varied between years and ages as an effect of cod predation and life history of the species. Specifically, length-at-age data from 1984 to 2021 has been used to derive VB parameters ( $L_{inf}$ ,  $k$  and  $t_0$ ) to be used as input parameters to derive proxy of M1 to be used in SMS (Annex 5 WD1). Since length-at-age data reveals a decreasing trend all along the timeseries, the VB equation parameters per year were used as input value in the Barefoot Ecologist's Toolbox ([http://barefootecologist.com.au/shiny\\_m](http://barefootecologist.com.au/shiny_m)) to produce a time-varying natural mortality vector (Annex 5 WD7). A significant breakpoint in the natural mortality time-series was detected in 2000 so the mean value calculated before and after the breakpoint (M before 2000: 0.28, M after 2000: 0.38; Annex 5 WD7) have been used to re-scale the assumed annual M1 in SMS (scenarios for "likely" M1 presented in Annex 5 WD1).

#### 2.4.1.3 Tuning indices

The data on Central Baltic herring stock size in the Baltic Sea, estimated by hydroacoustic methods, are collected annually by one internationally coordinated survey - the Baltic International Acoustic Survey (BIAS), which is conducted in autumn (October). The results from the individual national surveys are placed in the Access-database. This database file includes queries with the algorithms used to create the report tables and the calculation of the different tuning fleets.

WGBIFS (ICES 2023) provides in their latest report (Table 4.10) an updated tuning index for the assessment of the Central Baltic herring based on the BIAS herring abundance estimates in the ICES Subdivisions 25-29 per age-group (1-8+) for the years 1991-2021. Additionally, also the recruitment index for Central Baltic herring (age 0) is presented there (in Table 4.11). Compared to the previous tuning indices, used in assessment, even some historic corrections were made. Namely has Finland presented corrections for their 2016, 2018 and 2019 survey results, which were implemented in the BIAS database. As result the herring abundance estimates changed very slightly for those years. WGBIFS (ICES, 2023) recommends that, the updated and corrected BIAS index series can be used in the assessment of the Central Baltic herring stock with the restriction that the years 1993, 1995 and 1997 are excluded from the index series.

During the WGBFAS 2022 meeting requested stock assessors of Central Baltic herring that WGBIFS would provide WKBALTPEL with some alternative acoustic tuning indices (e.g. indices calculated with StoX and/or including data from Gulf of Finland). Additionally, it was requested that WGBIFS should provide number of hauls and survey variance values for Central Baltic herring tuning indices. These requests were discussed during the second meeting of WGBIFS in November 2022. Based on the results of the comparison exercises between StoX and traditional BIAS calculation methods WGBIFS concluded that the StoX calculated acoustic time-series cannot be used yet for the stock assessment of Central Baltic herring. WGBIFS (ICES 2023) provided (in Table 5.1.) a new Central Baltic herring tuning index, which also includes the survey data from the Gulf of Finland (SD 32). WGBIFS (ICES 2023) recommends that, the alternative BIAS index series (including data from SD 32) can be tested in the benchmark process of the Central Baltic herring stock with the restriction that the years 1999, 2001-2005 and 2008 are excluded from the index series.

Additionally, WGBIFS (ICES 2023) provided WKBALTPEL also with the numbers of BIAS hauls (Table 5.2) and survey variance estimates (in tables 4.23 and 4.24) per year for both index series.

## 2.4.2 Assessment model

### 2.4.2.1 General description of the modelling framework

The development of the reference model to be used in the ensemble of Central Baltic herring (*Clupea harengus*; CBH; her.27.25-2932) in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga

(Central Baltic Sea) was based on a virtual evolutionary process where the model evolved from an ancestor single sex, single area, single fleet, single survey age based XSA model towards a Stock Synthesis framework. The first Stock Synthesis model (SS3) has a very similar structure to the ancestor XSA model, so it is an age-based, single sex, single area, single fleet, single survey model which assumes logistic selectivity for both the fleet and the survey. From this first generation of the SS3 model, alternative model configurations were tested, which were used to select the reference model configuration to be used in the ensemble.

The different plausible reference model configurations tested were compared using model diagnostics based on recent papers from Carvalho *et al.*, 2017, 2021, Kell *et al.*, 2021 and Merino *et al.*, 2022). The key model diagnostics used were Runs test and RMSE, retrospective analysis, and hindcasting cross-validation (ICES, 2022). Once the candidate reference model configuration was selected, additional model diagnostics as jittering, MCMC, and trend in recruitment deviations were used to validate the selected reference model, which will be then used in the ensemble.

The text is split into four sections: (1) exploration of candidate reference model configurations and selection of the reference model (2) Additional diagnostics of the selected reference model, (3) detailed diagnostics of the reference model, (4) trends in SSB, F and R as estimated by the reference model and (5) additional runs tested. It is important to note that the reference model is selected based on diagnostics as in 3) and that estimates and trends in SSB, F, and R as described in 4) were used only to check the plausibility of the results.

#### **2.4.2.2 Exploration of candidate reference model configurations and selection of the reference mode**

##### **Model platform**

Assessment of Central Baltic herring (*Clupea harengus*; CBH; her.27.25-2932) in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea) was conducted using the Stock Synthesis (SS) model (Methot & Wetzel, 2013; Methot *et al.*, 2021). Stock Synthesis is programmed in the ADMB C++ software and searches for the set of parameter values that maximize the goodness-of-fit, then calculates the variance of these parameters using inverse Hessian and MCMC methods. The assessment was conducted using the 3.30.20 version of the Stock Synthesis software under the windows platform.

##### **Assessment model general configuration**

Here we describe the general configuration of the SS model that was common to all model configurations tested.

The assessment model of CBH is age-based, single sex, single area, single fleet, and single survey model with a population comprised of 8+ age classes (with age 8 representing a plus group). The model has a yearly time step with sexes combined (males and females are modelled together). Fishing mortality was modelled using a fleet-specific method (Methot *et al.*, 2022). Option 5 was selected for the F report basis as this option corresponds to the fishing mortality requested by the ICES framework (i.e. simple unweighted average of the F of the age classes chosen to represent the  $F_{\text{bar}}$  (age 3-6)).

##### **Landings**

Landings time-series starts in 1903, with 1903-1973 landings data obtained from the ICES historical database. From 1974 to 2021, landings are the same as used in the latest assessment (ICES 2022b) except for Danish landings data that have been updated (Annex 5 WD2).

### Uncertainty measures and likelihood

The total likelihood of a stock synthesis model is composed of several components, including the fit to survey and CPUE indices, tag recovery data (when tagging data are used), fishery and survey length frequency (LFD), age compositions and conditional age at length information, and catch data. There are also contributions to the total likelihood from the recruitment deviations and priors on the individual model parameters (if any). The model is configured to fit the catch so the catch component of the likelihood is generally small (although catch penalties might be created which allows the estimated catches to differ from the inputted catches). Details of the formulation of the individual components of the likelihood are provided in Methot & Wetzel (2013). The Hessian matrix computed at the mode of the posterior distribution was used to obtain estimates of the covariance matrix, which was used to compute approximate confidence intervals for parameters of interest.

### Samples sizes, CVs, data weighting

For the commercial fleet, the coefficient of variation (CV) of the catches (including the discards, discards were not estimated separately from landings as they are estimated to be negligible, i.e. less than 5% of the catches; ICES 2022b) was set to 0.1 for years from 1903 to 1973 and to 0.05 for years from 1974 to 2021. The CV of the initial catches of the commercial fleet was set to 0.2. The annual sample size associated with the age distribution data is reported as the number of trips sampled for commercial catches (as reported from national sources) and the number of hauls for the surveys. The CV of the acoustic survey index was estimated through bootstrapping (ICES 2023). Dirichlet-multinomial error distribution was used as an additional weighting of the age compositions (Methot *et al.*, 2022).

### Spawning stock biomass and recruitment

Spawning stock biomass was estimated at the beginning of the year and it was considered proportional to fecundity. In the model, recruitment was assumed to be a single event occurring at the beginning of the year. Recruitment was derived from a Beverton and Holt (BH) stock-recruitment relationship (SRR) and variation in recruitment was estimated as deviations from the SRR. Recruitment deviates were estimated for 1974 to 2020 as main recruitment deviations (47 annual deviations) and for 1968 to 1983 as early recruitment deviations (8 annual deviations). Recruitment deviates were assumed to have a standard deviation ( $\sigma_R$ ) of 0.5.  $\sigma_R$  is the stochastic recruitment process error and the estimation of this parameter within integrated models is generally recognised to be problematic (Kolody *et al.*, 2019). Consequently,  $\sigma_R$  for individual recruitment estimates were fixed at values large enough to prevent the SSR from constraining individual recruitment estimates (e.g. analogous to traditional VPA) (Kolody *et al.*, 2019). A meta-analysis of the estimation of  $\sigma_R$  performed outside the operative model (ISSF, 2011) yielded a median estimate between 0.2 and 0.5. This suggests that  $\sigma_R$  is often inflated in assessment models as it is generally set larger than 0.5. In all models tested,  $\sigma_R$  is fixed at 0.5. The steepness ( $h$ ) for the SRR and the autocorrelation of recruitment is also estimated within the model using a full Beta prior of 0.74 with a standard deviation of 0.113 as derived for herring in Myers *et al.* (1999).

### Weight and maturity

Empirical time-varying weight at age matrices for both commercial fleet and survey indices are provided as input for the model and are estimated using commercial data. A time-varying maturity at age matrix is also provided as input and derived from commercial data. Details on how weight and maturity at age were derived are included in the WD5 and WD6 in Annex 5.

### Natural mortality (M)

Estimates of natural mortality ( $M = M_1 + M_2$ ) at age and year (1974-2021) of herring were obtained from several runs using the SMS multispecies models (Annex 5 WD1) with alternative

configurations of food intake by cod and alternative values of M1 for herring estimated from growth (Annex 5 WD6). ICES WGSAM considered the key run made in October 2022 (WGSAM reference-ICES 2022) to provide the “best” estimate of M. However, the alternative runs supplied during this benchmark (Annex 5 WD1) provide similar results and in some cases a better model likelihood than the key run, even though the differences were small. The difference in total mortality (Z) estimated by the various model configurations is small, especially for the age 3+. The Z estimated from the key run for herring age 3+ were higher than Z estimated from catch curves (Annex 5 WD1), which may indicate that the key run Z might be too high for herring, and that alternative model configurations which produces a lower M are preferable for using in the Central Baltic herring ICES single species assessment. Thus, three alternative SMS configurations were selected based on AIC (Annex 5 WD1), which were M1\_010 (average annual M1 = 0.1, Quarterly M1(1974-1999) = 0.08/4 and M1(2000-2021) = 0.12/4), M1\_020\_average annual M1 = 0.2, Quarterly M1(1974-1999) = 0.17/4 and M1(2000-2021) = 0.23/4 and lim\_10 (10% quantile of the parameter a and b for food consumptions (ignoring correlation). M1\_010 was used for the development of the different plausible reference model configurations while M1\_010, M1\_020 and lim\_10 were used as alternative hypotheses of the M dimension of the ensemble. Natural mortality at age between 1903 and 1973 was assumed to be equal to the values estimated in 1974. This assumption is justified as M of herring is assumed to be dependent on the abundance of Eastern Baltic cod, for which the SSB was practically constant between the beginning of the century and 1974 (Eero *et al.*, 2008).

### Fishery dynamics

Fishery selectivity of the reference model is assumed to be age-specific and time-invariant. For both commercial fleet and surveys, a random walk selectivity was used. This selectivity pattern provides for a random walk for  $\ln(\text{selectivity})$ . For each age  $a \geq A_{\text{min}}$ , where  $A_{\text{min}}$  is the minimum age for which selectivity is allowed to be non-zero, there is a selectivity parameter,  $p_a$ , controlling the changing selectivity from age  $a - 1$  to age  $a$ . All data inputs are summarized in Table 2.2.

**Table 2.2. Herring (*Clupea harengus*) in subdivisions 25–29 and 32, excluding the Gulf of Riga (central Baltic Sea). Input data used in the Stock Synthesis assessment models.**

Type	Name	Year range	Range
Catches	Catches in tonnes for each year	1903-2021	
Age compositions	Catch in numbers (thousand) per age class	Commercial fleet: 1974-2021 Acoustic survey: 1991- 2021	0 – 8+
Maturity at age	Empirical maturity at age estimated from commercial data	1903-2021	0 – 8+
Weight at age	Empirical weight at age estimated from commercial data	1903-2021	0 – 8+
Natural mortality	Time varying natural mortality by age class for the entire time-series derived from SMS (WD1 and WD7 in Annex 5)	1903-2021	0 - 8+
Surveys indices	Abundance index from acoustic survey	Acoustic survey: 1991- 2021	

### Exploratory runs

The following alternative configurations were explored (Table 2.3):

Those were:

1. A model which uses survey covering SDs 25-29 (i.e. excluding SD 32) starting in 1991, and including number at age for 1-8+ individuals as estimated from the acoustic survey. The model assumes logistic selectivity for both the fleet and the survey for age 5 to 8+.
2. A model which uses survey covering SDs 25-29 (i.e. excluding SD 32) and including number at age for 0-8+ individuals as estimated from the acoustic survey. The model assumes logistic selectivity for both the fleet and the survey for age 5 to 8+.
3. A model which uses survey covering SDs 25-29 (i.e. excluding SD 32) and including number at age for 1-8+ individuals as estimated from the acoustic survey. Selectivity is estimated for all age classes.
4. A model which uses survey covering SDs 25-32 starting in 2000 and including number at age for 1-8+ individuals as estimated from the acoustic survey. Years 1999, 2001-2005 and 2008 were excluded as recommended by WGBIFS. The model assumes logistic selectivity for both the fleet and the survey for age 5 to 8+.
5. A model which uses the acoustic survey as two pseudo-surveys, one covering only SDs 25-29 (i.e. excluding SD 32) from 1991 to 1999 and a second covering SDs 25-32 starting in 2000. Both surveys include age classes 1-8+ as estimated from the acoustic survey. Years 1999, 2001-2005 and 2008 were excluded as recommended by WGBIFS. The model assumes logistic selectivity for both the fleet and the survey for age 5 to 8+.



**Table 2.3. Herring (*Clupea harengus*) in subdivisions 25–29 and 32, excluding the Gulf of Riga (central Baltic Sea). Description of exploratory runs.**

Name	Brief description	Reason
<i>Reference_run</i>	Survey covering SDs 25-29 from 1991, ages 1-8+ assuming logistic selectivity for both the fleet and the survey	Model emulating previous XSA
<i>Reference_run_age0</i>	Survey covering SDs 25-29 from 1991, survey ages 0-8+ assuming logistic selectivity for both the fleet and the survey	Include age 0 from the acoustic survey
<i>Reference_run_DomShSel</i>	Survey covering SDs 25-29 from 1991 and age 1-8+ and selectivity estimated for all age classes for both the fleet and the survey	Model that does not assume logistic selectivity for the fleet and the acoustic survey and selectivity is freely estimated
<i>Reference_run_SD32_survey</i>	Survey covering SDs 25-32 from 2000 and age 1-8+ assuming logistic selectivity for both the fleet and the survey	Survey covering all SDs but starting in 2000
<i>Reference_run_Survey_split</i>	Acoustic survey treated as two pseudo-surveys, one covering only SDs 25-29 from 1991 to 1999 and a second covering all SDs 25-32 starting in 2000 with age 1-8+ assuming logistic selectivity	Acoustic survey split in 2 pseudo-surveys

## Model selection

It is good practice that an objective methodology for selecting, pruning and weighting hypotheses is pre-agreed, to overcome artifacts and biases introduced by a “cherry-picking” approach (Pechlivanidis *et al.*, 2018). This is particularly important since divergent views and opinions mean that uncertainties can be used to support stakeholder positions and to strengthen or weaken management measures (Fromentin *et al.*, 2014).

Stock assessment models are deeply scrutinised for model misspecification during their development within benchmark workshops. Traditionally in ICES, diagnostics have been based on retrospective and visual analysis of the residuals. However, recent papers by Carvalho *et al.*, (2021) showed that when several diagnostic tests are considered together, the power to detect model misspecification improves without a substantial increase in the probability of incorrectly rejecting a correctly specified model (Carvalho *et al.*, 2017, 2021). Consequently, several available diagnostics should be applied routinely during benchmarks. When the criterion for rejecting a model is a failure of at least one of the diagnostic tests, nearly 90% of most model misspecifications are detected with no real increase in the probability of false detection (Carvalho *et al.*, 2017, 2021). For example, residual analyses were easily the best detector of misspecification in the observation model, while the retrospective analysis had low rates of detection of mis-specified models (Carvalho *et al.*, 2017, 2021), although retrospective analysis is effective in detecting unmodelled temporal variation (Hurtado-Ferro *et al.*, 2015). Also, as opposed to the widely used maximum-likelihood estimator, MCMC gives clear warning signs when non-identifiable parameters are used for fitting (Siekman *et al.*, 2012).

## Model diagnostics

The different plausible reference model configurations to be used in the ensemble were compared using model diagnostics based on recent papers from Carvalho *et al.*, 2017, 2021; Kell *et al.*, 2021 and Merino *et al.*, 2022). The key model diagnostics used were convergence (which includes checking of parameters at the bounds, final gradient and inversion of the Hessian matrix for

uncertainty estimation), runs test and RMSE, retrospective analysis, and hindcasting cross-validation (ICES 2022b). Once the most plausible reference model configuration has been selected, additional model diagnostics such as jittering, MCMC, analysis of surplus production trend, and trend in recruitment deviations were used to further validate the reference model which will be then used in the ensemble. Estimates and trends in SSB, F, and R were used at the last step and only as a plausibility check.

### Convergence

The first step for checking model convergence is to verify if parameters are estimated at a bound, which can suggest problems with data or the assumed model structure. The second is checking that the final gradient of the model is relatively small (e.g.  $\leq 1.00E-04$  or smaller). The third is to determine whether the Hessian (i.e. the matrix of second derivatives of the log-likelihood concerning the parameters, from which the asymptotic standard error of the parameter estimates is derived) is positive definite (Carvalho *et al.*, 2021). Other convergence diagnostics include (i) examining the correlation matrix for highly correlated (e.g.  $> 0.95$ ) parameter pairs; and (ii) examining parameters for excessively high variance as an indication that they do not influence the fit to the data (Carvalho *et al.*, 2021).

### Residuals test

A non-random pattern of residuals may indicate that some heteroscedasticity is present, or there is some leftover serial correlation (serial correlation in sampling/observation error or model misspecification). Several well-known nonparametric tests for randomness in a time-series include: the runs test, the sign test, the runs up and down test, the Mann-Kendall test, and Bartel's rank test (Gibbons and Chakraborti, 1992). Standardized residuals are commonly used, although recent analysis showed that one-step-ahead (OSA) should be used instead in stock assessment model diagnostic (Trijoulet *et al.*, 2022) because the use of compositional distributions create correlation between observations, which are propagated in the residuals if estimated as Pearson while OSA residuals are independent, and standard normally distributed for correctly specified models. Here we used the runs test to evaluate whether residuals of the surveys, and the length frequency distributions, were normally distributed and/or displayed time trends. The runs test was chosen as this test has recently been used to diagnose fits to indices and other data components in other assessment models (e.g. FAO-GFCM, 2021; Winker *et al.*, 2018; Carvalho *et al.*, 2021; ICES 2022b).

The RMSE runs test (Carvalho *et al.*, 2021 for details) could indicate the presence of a random pattern in the length frequency distributions and in the survey indices. The RMSE plot is frequently used as a tool for identifying trends in residuals, and if the standard deviation is small on a given year this means the fleets included in the model agree, even if not fitting well, which is a useful diagnostic. Its purpose is to visualize multiple residuals at once, pick up on periods of substantial data conflicts (width of boxes), and systematic departures in median residuals (loess smoothers). The fit is considered satisfactory if no residuals are larger than 1 and the RMSE is below 30%.

### Retrospective analyses

Retrospective analysis is a diagnostic approach to evaluate the reliability of parameter and reference point estimates and to reveal systematic bias in the model estimation. It involves fitting a stock assessment model to the full dataset. The same model is then fitted to truncated datasets where the data for the most recent years are sequentially removed. The retrospective analysis was conducted for the last 5 years of the assessment time horizon to evaluate whether there were any strong changes in model results. Given that the variability of Mohn's rho index depends on life history, and that the statistic appears insensitive to F, Hurtado-Ferro *et al.* (2015) proposed the following rule of thumb when determining whether a retrospective pattern should be

addressed explicitly. Values of Mohn's rho index higher than 0.20 or lower than -0.15 for long-lived species (upper and lower bounds of the 90% simulation intervals for the flatfish base case), or higher than 0.30 or lower than -0.22 for short-lived species (upper and lower bounds of the 90% simulation intervals for the sardine base case) should be cause for concern and taken as indicators of retrospective patterns. However, Mohn's rho index values smaller than those proposed should not be taken as confirmation that a given assessment does not present a retrospective pattern, and the choice of 90% means that a "false positive" will arise 10% of the time. In both cases, model misspecification would be correctly detected more than half the time.

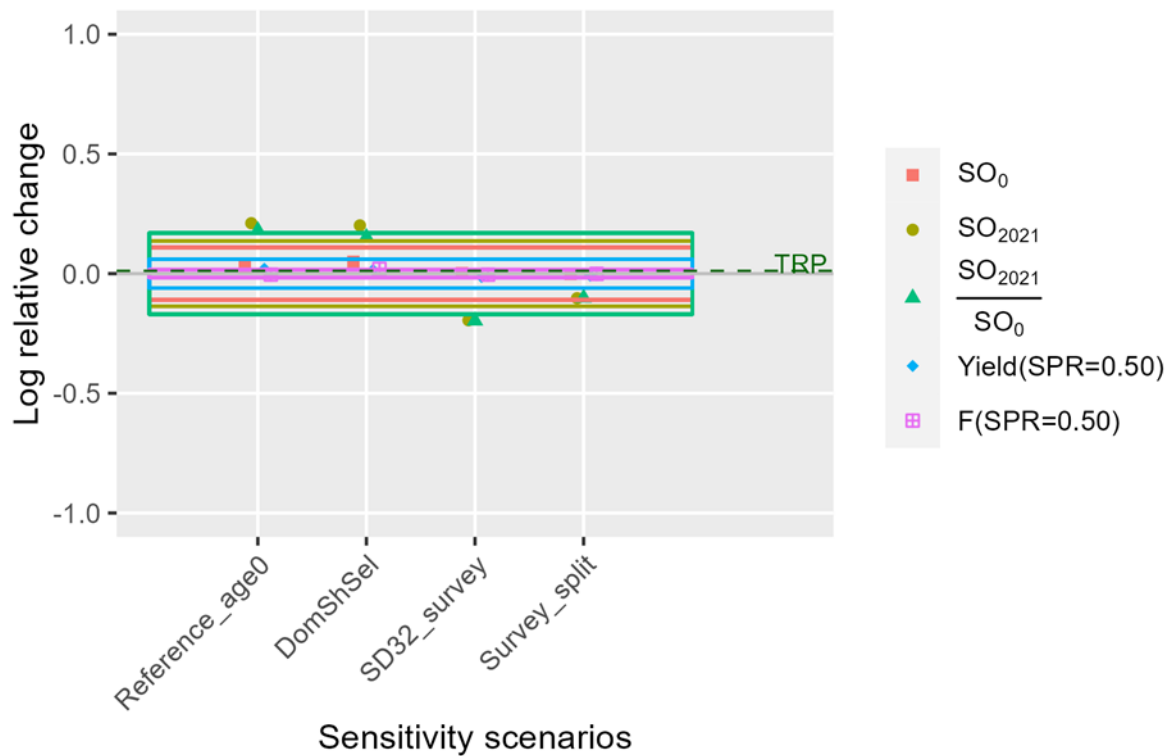
### Hindcasting

The provision of fisheries management advice requires the assessment of stock status relative to reference points, the prediction of the response of stock to management, and checking that predictions are consistent with reality. A major uncertainty in stock assessment models is the difference between model estimates and reality. To evaluate this uncertainty, it is common for several scenarios to be considered, whereby scenarios correspond to alternative model structures and/or dataset choices (Hilborn, 2016). It is difficult, however, to empirically validate model predictions, as fish stocks can rarely be observed and counted. Various criteria are available for estimating prediction skill (Hyndman and Koehler, 2006). One commonly used measure is a root-mean-square error (RMSE). RMSE, however, is an inappropriate and misinterpreted measure of average error (Willmott and Matsuura, 2005). On the other hand, mean absolute error (MAE) is a more natural measure of average error, and unlike RMSE is unambiguous. Scaling the average errors using the Mean Absolute Scaled Error (MASE) allows forecast accuracy to be compared across a series at different scales. MASE values greater than one indicates that in-sample one-step forecasts from the naïve method perform better than the forecast values under consideration. MASE also penalizes positive and negative errors and errors in large forecasts and small forecasts equally.

Kell *et al.* (2016, 2021) and Carvalho *et al.*, (2021) showed that hindcasting can be used to evaluate the model prediction skill of the CPUE time-series. When conducting hindcasting, a model is fitted to the first part of a time-series and then projected over the period omitted in the original fit. Prediction skills can then be evaluated by comparing the predictions from the projection with the observations using, for example, the MASE indicator (Hyndman and Athanopoulos, 2013). If a model is used for prediction, the specific tool used for model selection is less important than the approach used to validate predictions. Quantifying predictive skills using independent data in ecology is therefore essential (Tredennick *et al.*, 2021).

### Conclusions and selection of the reference model configurations to be used in the ensemble

The alternative model configurations are very similar in terms of stock status as key productivity parameters are within or just at the 95% confidence intervals of the reference model (Figure 2.3).



**Figure 2.3.** Herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Sensitivity to alternative model configurations as described in the text. Boxes correspond to the 95% confidence interval of a derived quantity (indicated by colour) in the Reference\_run. Values outside the box would indicate significant differences from the uncertainty provided in the Reference\_run. The metric used were: SSB0 ( $SO_0$ ), SSB ( $SO$ ) in 2021,  $SO_{2021}/SO_0$ , yield at  $SPR_{50\%}$  (Spawner per recruit) and  $F$  at  $SPR_{50\%}$ .

Table 2.4 summarizes all diagnostics used to evaluate candidates' reference model configurations to be used in the ensemble. The table is an attempt to sum up a multidimensional space and thus it needs to be seen as a guidance more than as a definitive result. All models invert the hessian matrix and have a good to moderate convergence. The ordinary runs test was passed for all components tested and for all models (Table 2.4) except for the age compositions runs test of Reference\_run\_SD32\_survey. On the other hand, the RMSE is below 30% for both the survey and the age compositions only for Reference\_run\_SD32\_survey.

The estimated Hurtado-Ferro *et al.* (2015) variant of the Mohn's rho indices was inside the bounds of recommended values for long-lived species for model Reference\_run and Reference\_run\_SD32\_survey, failed only for forecast  $F$  for model Reference\_run\_DomShSel and did not pass for model Reference\_run\_age0 and Reference\_run\_Survey\_split (Table 2.4). Hindcasting was conducted for all models (Table 2.4). The results showed the Reference\_run, Reference\_run\_DomShSel, Reference\_run\_SD32\_survey and Reference\_run\_Survey\_split have good prediction skills given that the MASE value is lower than the 1.0 threshold when predicting the index one year ahead. On the other hand, Reference\_run\_age0 failed the MASE test for the acoustic survey index. All MASE values for the age compositions (including the joint MASE) are well below the 1.0 threshold (Table 2.4).

From Table 2.4 the models that perform best are Reference\_run\_SD32\_survey and Reference\_run. Both models perform well in most of the key diagnostic tests performed although Reference\_run\_SD32\_survey has better performances in terms of retrospective and prediction skills.

Thus, Reference\_run\_SD32\_survey was proposed as the reference model to be used to integrate the key dimensions of uncertainty in the ensemble (see Section 1.1.2.4).

**Table 2.4. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (central Baltic Sea). Summary table of the diagnostics of the model tested as a candidate of the reference model of the ensemble. Values in red refer to tests that failed and in green those that were passed.**

Run	Reference_run	Reference_run_age0	Reference_run_DomShSel	Reference_run_SD32_survey	Reference_run_Survey_split
Convergence	3.37E-04	6.57E-04	1.05E-04	4.90E-04	1.12E-03
Total_LL	1313	1413	1288	1073	1309
N_Params	69	69	75	69	74
Runs_test_cpue1	Passed	Passed	Passed	Passed	Passed
Runs_test_cpue2	NA	NA	NA	NA	Passed
Runs_test_age1	Passed	Passed	Passed	Failed	Passed
Runs_test_age2	Passed	Passed	Passed	Passed	Passed
Runs_test_age3	NA	NA	NA	NA	Passed
RMSE_Perc	31.7	31.0	31.9	25.5	30.1
RMSE_Perc_2	6.1	8.9	6.1	6.0	6.1
Retro_Rho_SSB	0.11	0.22	0.19	0.08	0.28
Forecast_Rho_SSB	0.17	0.34	0.26	0.12	0.38
Retro_Rho_F	-0.11	-0.20	-0.15	-0.09	-0.25
Forecast_Rho_F	-0.14	-0.29	-0.19	-0.12	-0.30
MASE_cpue1	0.94	1.56	0.95	0.83	NA
MASE_cpue2	0.94	1.56	0.95	0.83	0.99
MASE_cpue3	NA	NA	NA	NA	0.99
MASE_age1	0.41	0.62	0.38	0.42	0.55
MASE_age2	0.53	0.88	0.47	0.68	NA
MASE_age3	0.47	0.91	0.42	0.55	0.61
MASE_age4	NA	NA	NA	NA	0.58

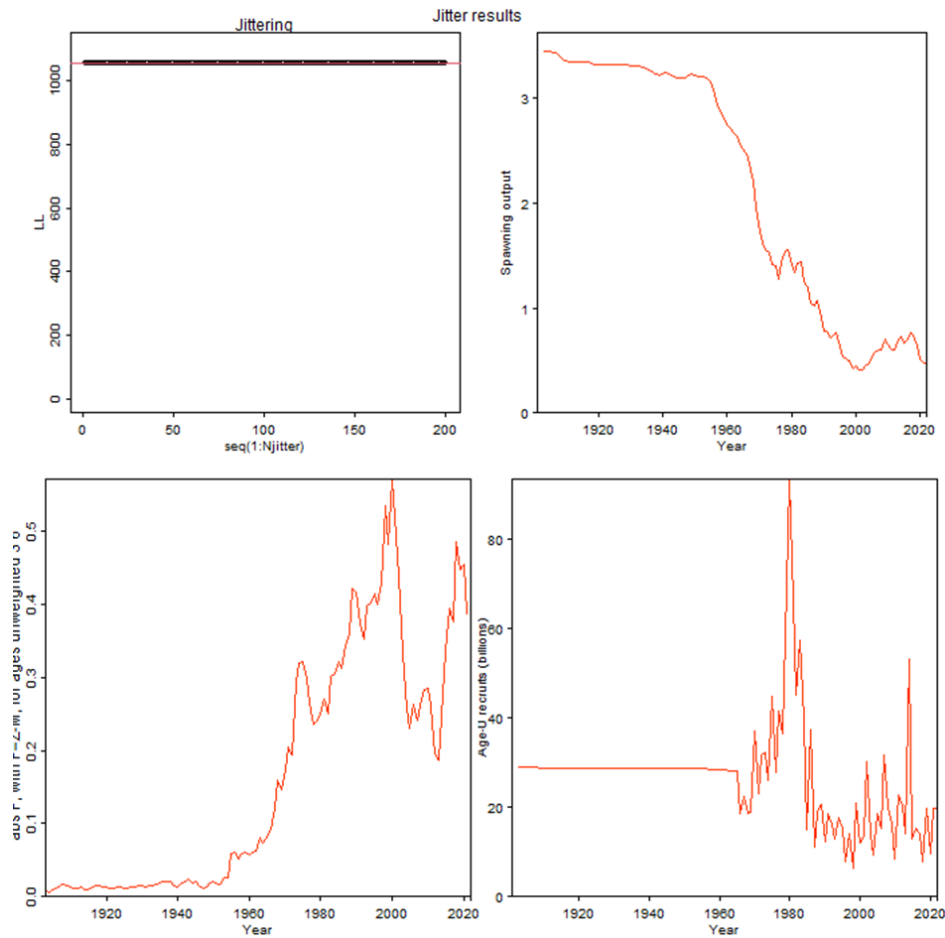
### 2.4.2.3 Additional diagnostics of the selected reference model

Before using the selected Reference\_run\_SD32\_survey model configuration as the reference model for the ensemble, additional key diagnostics were run. Those were jittering, MCMC with NUTS algorithm, test for trends in recruitment deviations and analysis of surplus production trend. A minor change was done compared to the original Reference\_run\_SD32\_survey model configuration in that a parameter for estimating extra standard deviation of the acoustic survey was introduced. This was done as the estimated CV of the acoustic survey is small, always below 0.2. Also, initial F was originally estimated at the lower bound (i.e., 0) and thus was fixed at the estimated value of 0.009. This improves both retrospective and hindcasting (see results below).

#### Jittering

The jittering procedure allows users to verify the stability of a model and its parameter estimates by examining the effect that small changes in its starting values have on model results. An accurate model should converge on a global solution (i.e., not be stuck in local minima of the likelihood surface) across a reasonable range of starting values for all input parameters. In this case, 200 runs were performed considering a 10% jitter of all initial parameters. In practice, this means that a small random jitter is added to the initial parameter values and the model is rerun. Starting values are jittered based on a normal distribution based on the  $\text{pr}(\text{PMIN}) = 0.1\%$  and the  $\text{pr}(\text{PMAX}) = 99.9\%$ .

All jittered runs resulted in the same results as the reference run (Figure 2.4) and no local minima were observed as no runs had a likelihood lower than the reference run. It is, however, important to stress that the absence of a local minima when running jittering is not a guarantee that the model is not indeed stuck in a local minimum, although its absence does reduce the risk that this occurs (Subbey, 2018).



**Figure 2.4** Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Results from jitter using 200 iterations and an average jitter of 10%.

### MCMC

Markov chain Monte Carlo (MCMC) methods comprise a class of algorithms for sampling from a probability distribution. It is used in integrated models for detecting misspecification in key fixed parameters or issues with the estimation of the parameters. By constructing a Markov chain, it is possible to obtain a sample of the desired distribution by observing the chain after several steps. The more steps there are, the more closely the distribution of the sample is expected to match the actual desired distribution. MCMC methods create samples from a possibly multi-dimensional continuous random variable, with probability density proportional to a known function. These samples can be used to evaluate an integral over that variable, as its expected value or variance. Practically, an ensemble of chains is generally developed, starting from a set of points arbitrarily chosen and sufficiently distant from each other. Those are then used to estimate the posterior distribution of the parameters of interest within the model.

We performed an MCMC run as diagnostic (i.e. thus not for inference, for which a much larger number of iterations might be necessary) using the NUTS algorithm with 3 chains of 5 000 000 iterations each. We discounted the first 500 000 of the iterations as a burn-in period and no thinning. The results showed that the MCMC is almost identical to the MLE estimated, which is an indication of the robustness of the model (Figure 2.5a). NUTS algorithm in MCMC (Monnahan *et al.*, 2019) was also used to regularize the model, i.e. to check that all parameters are identifiable. MCMC run with NUTS algorithm confirmed that all parameters of the model are identifiable including steepness (Figure 2.5b). The high correlation between steepness and  $R_0$  does not affect

the estimation of steepness as estimate from the MLE and the MCMC NUTS runs are practically identical (i.e. 0.78 and 0.782).

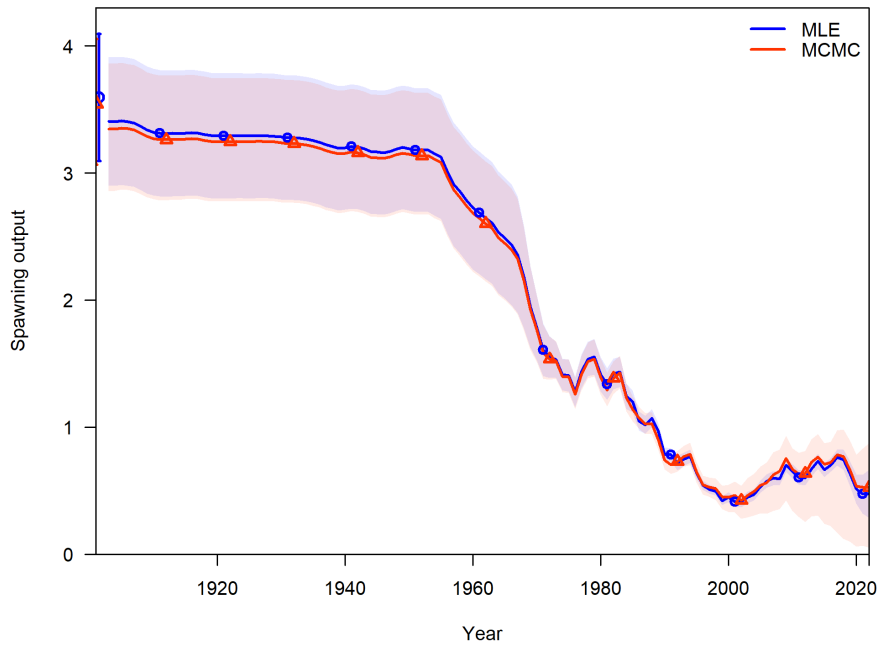
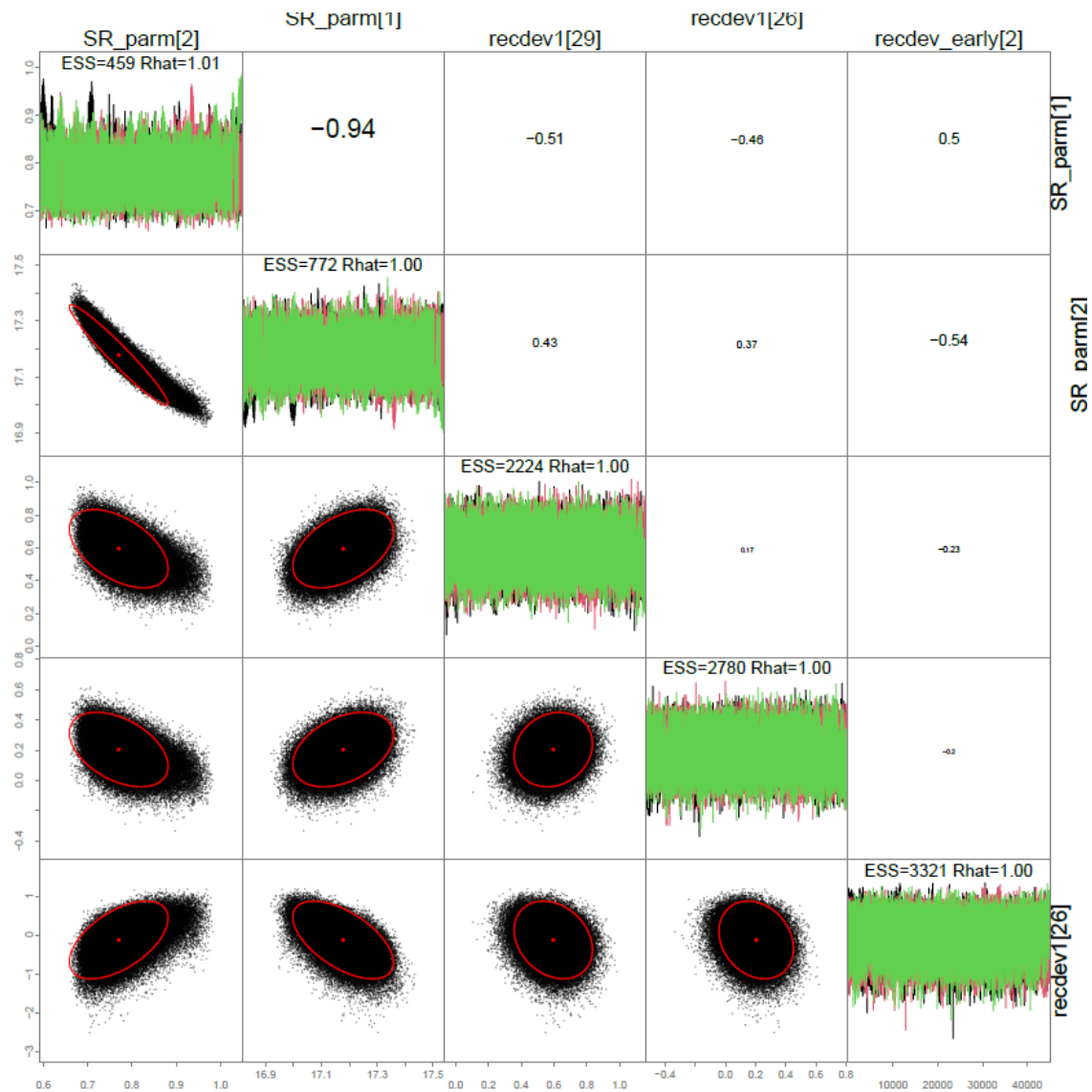


Figure 2.5a. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Results of the MCMC analysis in terms of SSB, R and F compared to the MLE model.



**Figure 2.5b.** Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Comparison between MLE (red points) against posteriors of the reference model obtained by an MCMC with 5 000 000 iterations, 3 chains, run with NUTS algorithm, with 500 000 iterations as burn in and thinning every 100 for the 5 slowest mixing parameters. Red ellipse is 95% confidence interval, points are posteriors draw and lines shows chain traces.

### Trends in recruitment deviations

Merino *et al.*, (2022) have described and applied a novel model diagnostic to identify trends in the process error in recruitment deviation estimates within integrated assessment models. Significant trends in recruitment deviates can be caused by misspecification of the biological parameters used as fixed values in integrated assessment models. The process error diagnostic described here can provide a statistical criterion in support of hypotheses and assumptions when using best case or ensembles of models to develop fisheries management advice. No significant temporal trend for the main recruitment deviations was found.

### Analysis of surplus production trend

Estimates of Surplus Production (Walters *et al.*, 2008) can provide a check of whether predictions of changes in biomass can be made reliably based on catch and current biomass (clockwise or linear behaviour) or whether there has been non-stationarity in production processes, i.e. are dynamics driven by climate and oceanic conditions (counter clockwise, Figure 2.6). This is



important, for example, for the development of management procedures (MPs) in the MSE process. In the case of Central Baltic herring, figure 6 shows a general clockwise pattern indicating that changes in stock biomass can be made reliably based on catch and current biomass. Moreover, the production function (Figure 2.7) is typically left skewed and flat-topped, which implies that fishing at the naïve  $F_{MSY}$  brings the stock close to  $B_{lim}$  with very reduced theoretical gain in yields compared to a more conservative proxy as  $B_{40\%}$ .

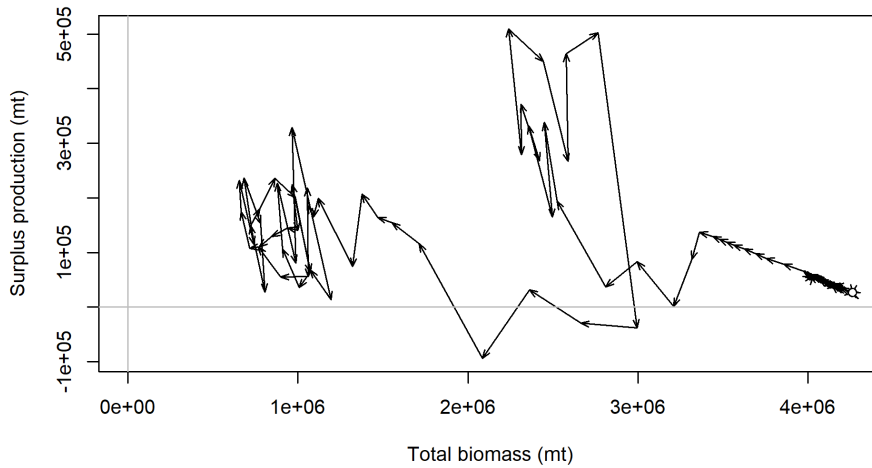


Figure 2.6. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Surplus production against biomass plot of the *Reference\_run\_SD32\_survey* model. The round circle represents the first year of the time-series (1903).

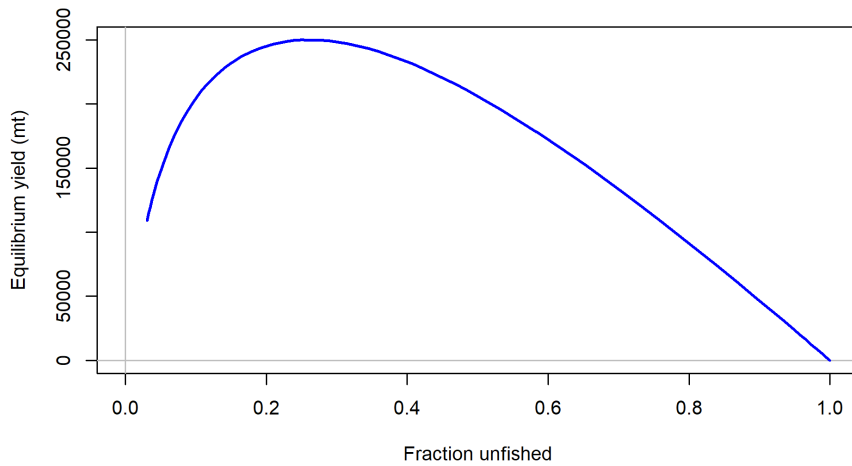
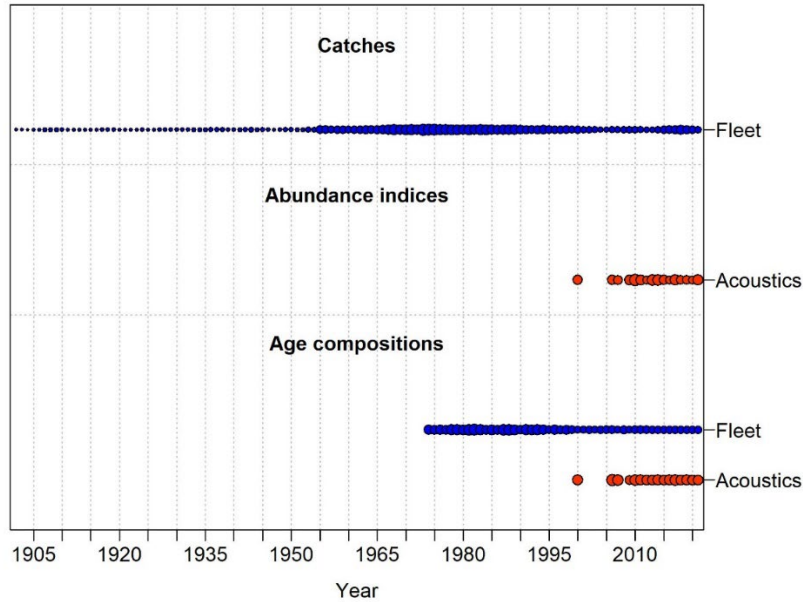


Figure 2.7. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (central Baltic Sea). The production function of the *Reference\_run\_SD32\_survey* model.

### 2.4.2.4 Detailed description of the selected reference model configuration and diagnostics



F

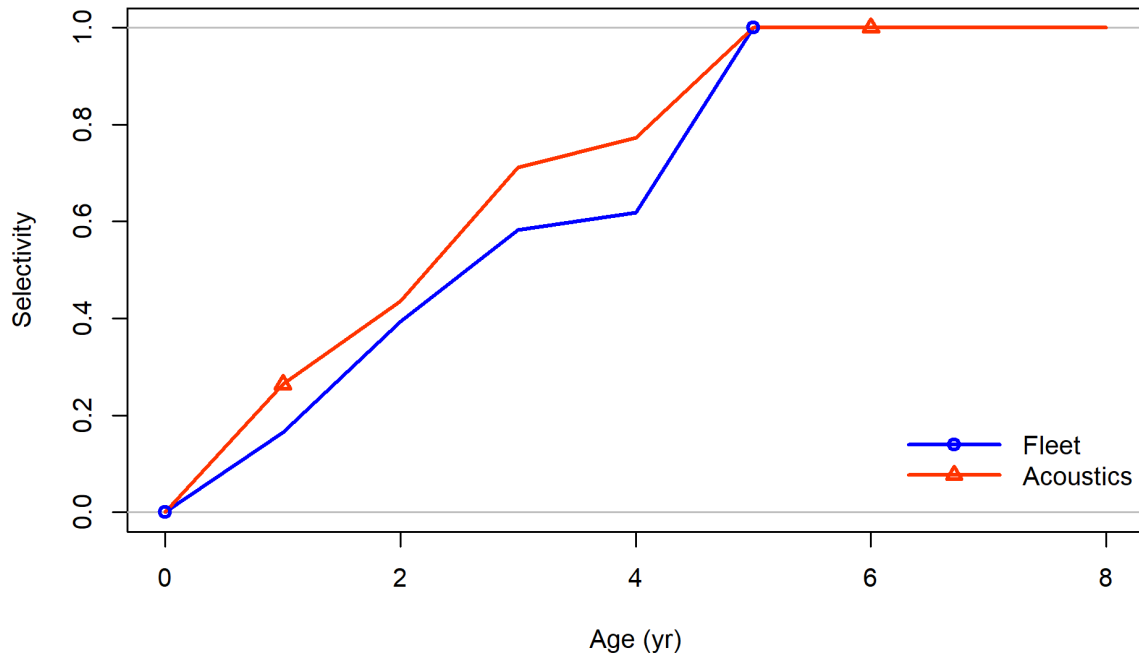
Figure 2.8. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Summary of the input time-series included in the *Reference\_run\_SD32\_survey* model.

Table 2.5. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Settings of the Stock Synthesis *Reference\_run\_SD32\_survey* model configuration. The table columns (left to right) show: the parameter name, the number of estimated parameters, the initial values (from which the numerical optimization is started), the intervals allowed for the parameters, the priors used (value and standard deviation), the value estimated by the model and its standard deviation. Parameters in bold are set and not estimated by the model. \* indicates that the parameter is close to the bound.

Parameter	Number of parameters estimated	Initial value	Bounds (low, high)	Prior and standard deviation	Value (MLE)	Standard deviation
<u>Natural mortality</u> (age classes 0-8+)		Time varying derived from SMS 1903-1973 = values estimated in 1974.				
<u>Stock and recruitment</u>						
$\ln(R_0)$	1	17.76	(16, 25)	No_prior	17.18	0.067
Steepness ( $h$ )	1	0.74	(0.1, 1)	0.74 (0.113)	0.78	0.041
Recruitment variability ( $\sigma_R$ )		<b>0.50</b>				
$\ln$ (Recruitment deviation): 1974 - 2020	47					
Recruitment autocorrelation**	1	0	(0,1)		0.18	0.09

Parameter	Number of parameters estimated	Initial value	Bounds (low, high)	Prior and standard deviation	Value (MLE)	Standard deviation
<u>Initial catches</u>		Average of 1903-1905				
Commercial fleet**	1	0.009	(0.001, 1)	No_prior	0.009*	0.002
<u>Selectivity</u>						
<b>Commercial fleet</b>						
<i>Change from age2 to age3</i>	1	0.79	(-5, 9)	No_prior	0.87	0.050
<i>Change from age3 to age4</i>	1	0.37	(-5, 9)	No_prior	0.39	0.050
<i>Change from age4 to age5</i>	1	0.19	(-5, 9)	No_prior	0.06	0.060
<i>Change from age5 to age8+</i>	1	0.09	(-5, 9)	No_prior	0.48	0.060
<b>Acoustic Survey</b>						
<i>Change from age2 to age3</i>	1	0.69	(-5, 9)	No_prior	0.50	0.08
<i>Change from age3 to age4</i>	1	0.50	(-5, 9)	No_prior	0.49	0.08
<i>Change from age4 to age5</i>	1	0.24	(-5, 9)	No_prior	0.08	0.09
<i>Change from age5 to age8+</i>	1	0.01	(-5, 9)	No_prior	0.26	0.10
<u>Catchability</u>						
<b>Acoustic survey</b>						
<i>Ln(Q) – catchability</i>		<b>-5.95741</b>				
<i>Extra standard deviation</i>	1	0	(0,1)	No_prior	0.156	0.050
<i>Dirichlet-multinomial error distribution (Fleet)</i>	1	5	(-5,5)	1.813 (6)	3.64	0.83
<i>Dirichlet-multinomial error distribution (Acoustic survey)</i>	1	5	(-5,5)	1.813 (6)	1.13	0.89

\*\*Estimated and then fixed at the estimated value in the final Reference\_run\_SD32\_survey model configuration used for the ensemble.



**Figure 2.9.** Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Age-based selectivity of the commercial fleet and acoustic survey of the *Reference\_run\_SD32\_survey* model configuration.

All parameter and variances were well estimated (i.e.  $CV < 1$ ) and within the bounds except for initial  $F$  which is estimated at the lower bound (Table 2.5). The selectivity of the fleet and the acoustic survey is well estimated (Figure 2.9). The fitting of the model was satisfactory, with the aggregated age compositions well reconstructed (Figure 2.10). The Pearson residuals are generally low, above -2.0 and below 2.0, and without clear patterns (Figure 2.11). One step ahead residuals were also calculated (Trijoulet *et al.*, 2022; OSA). The plots showed that those are quite similar to the Pearson residuals (Figure 2.12a,b), being OSA slightly larger and ACF plots showing some correlations in time for both fleets. The commercial fleet shows also positive lag 1 correlations within cohort and negative correlations for lags 3-6.

Overall, the model does provide a moderate fit to the survey, with a better fit in the latest years compared to the beginning of the time-series (Figures 2.13).

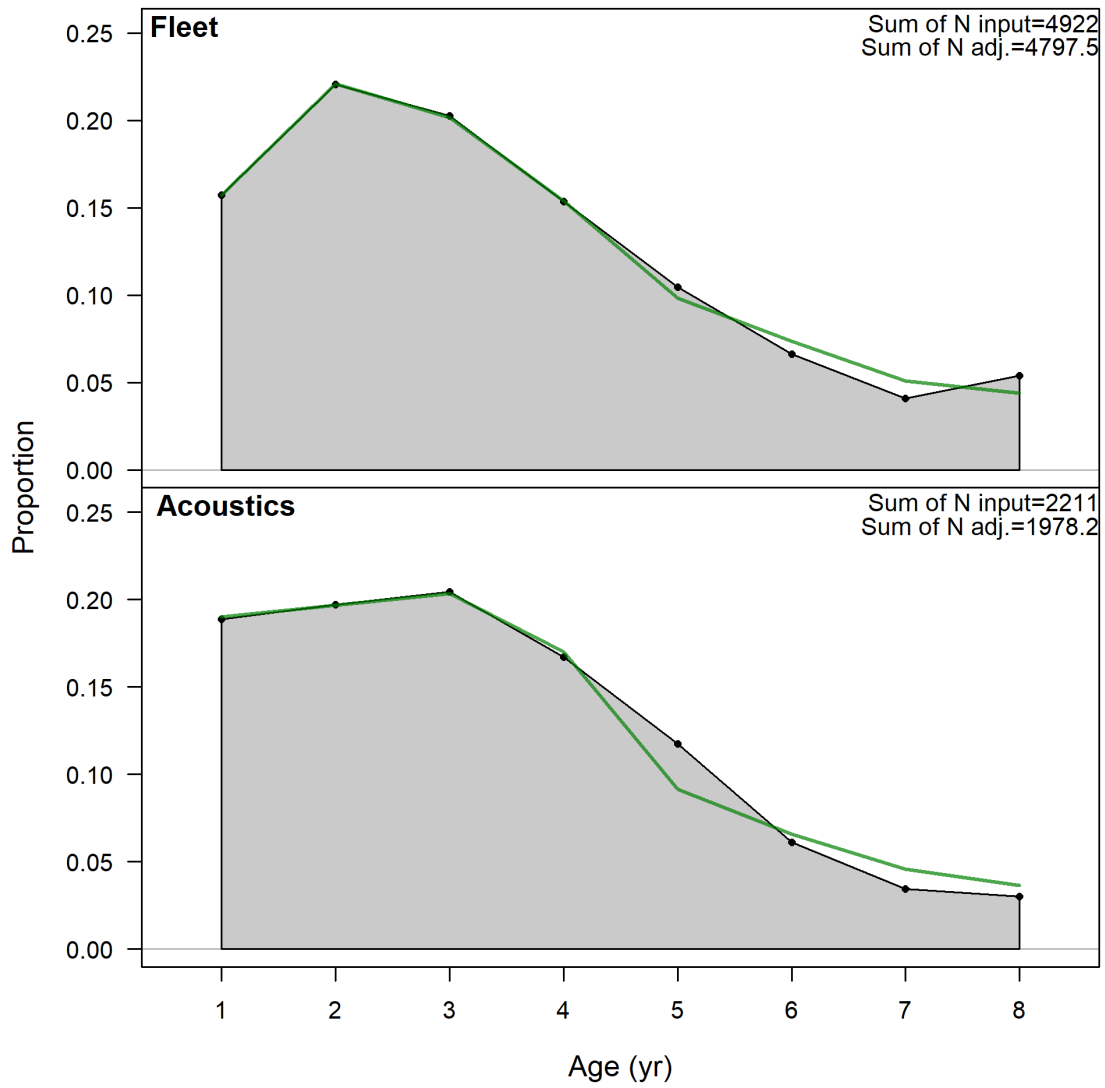


Figure 2.10. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (central Baltic Sea). Model fits to age composition data of the *Reference\_run\_SD32\_survey* model configuration.

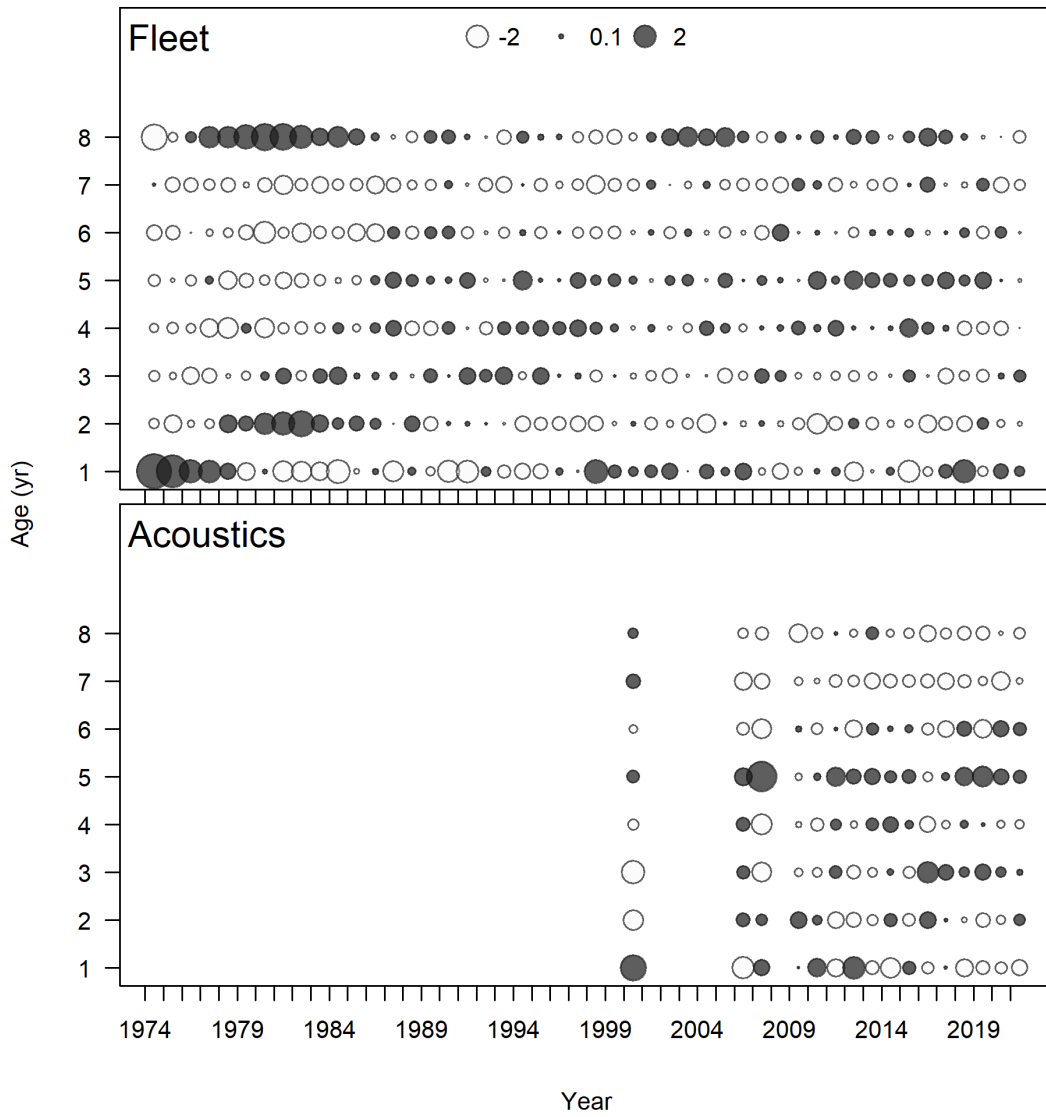


Figure 2.11. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Residuals of fits to age composition data for the different fleets of the *Reference\_run\_SD32\_survey* model configuration.

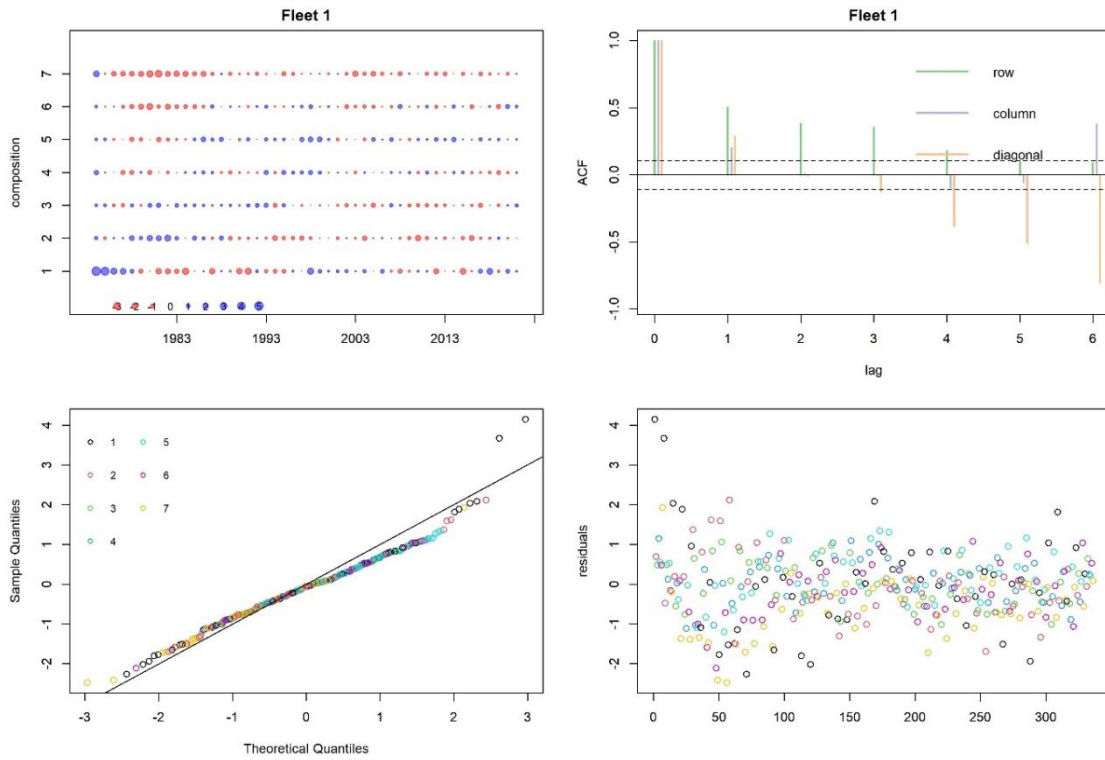


Figure 2.12a. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). One-step-ahead (OSA) residuals diagnostic plots for the age compositions of the commercial fleet. The top left plot shows the bubble plot of the residuals scaled to their size. The top right plot is the autocorrelation function (ACF) of residuals in three directions: row (correlation in time), column (correlation among ages), diagonal (correlation within cohort). The bottom left plot is a quantile-quantile (Q-Q) plot of the residuals. The bottom right plot is a simple plot of the residuals.

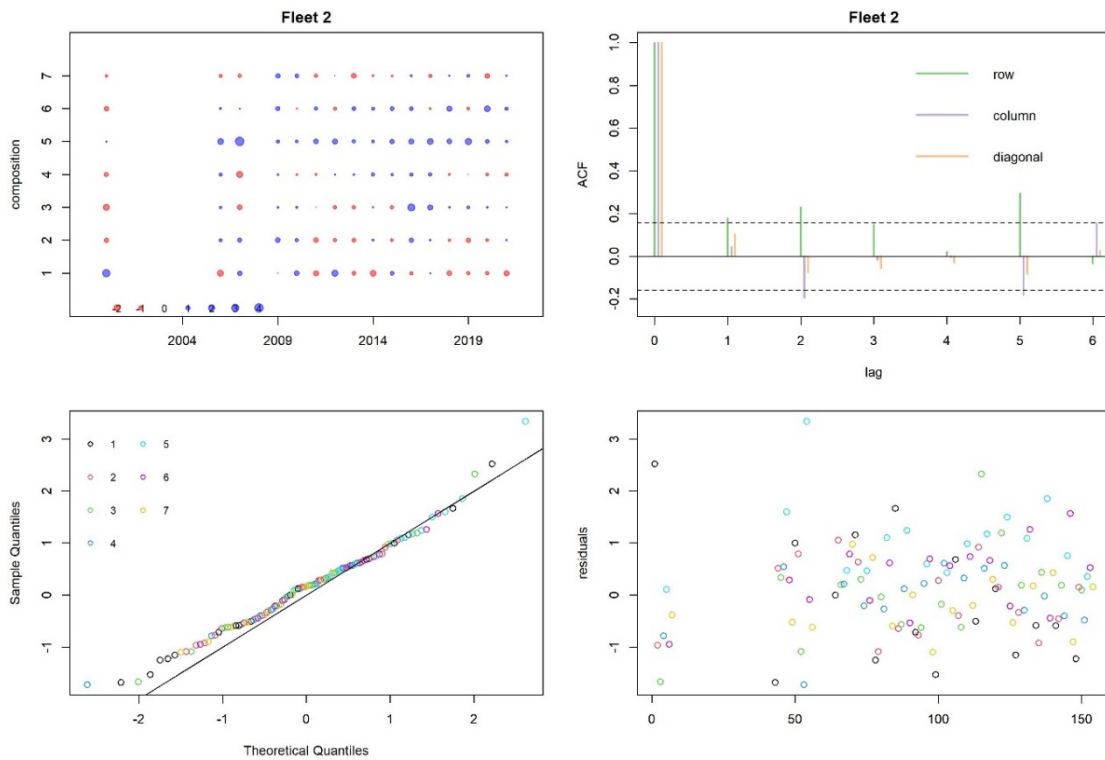
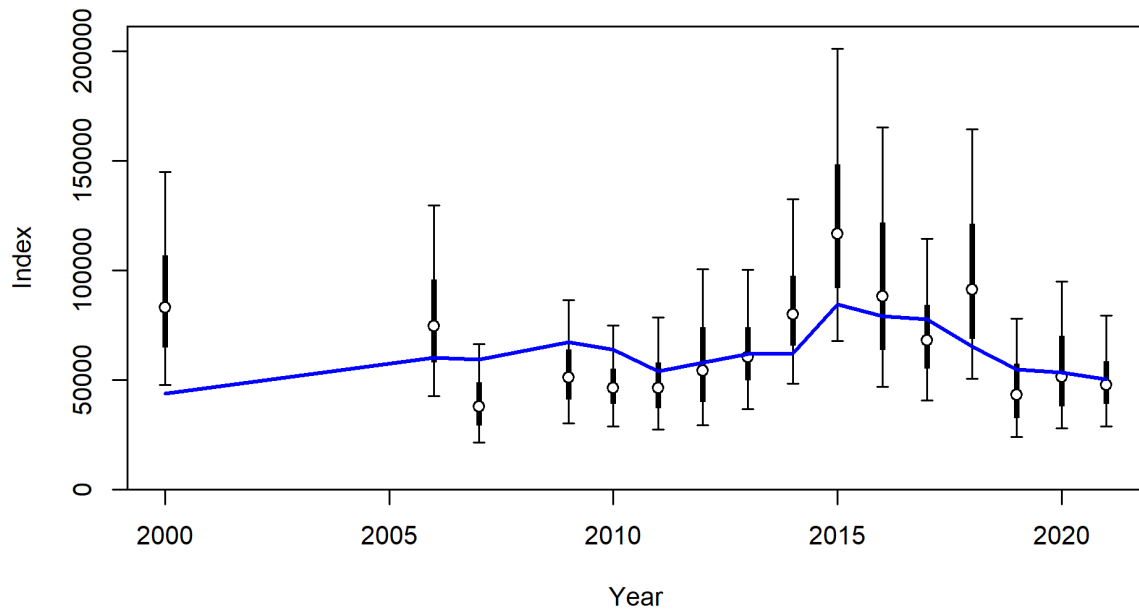


Figure 2.12b. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). One-step-ahead (OSA) residuals diagnostic plots for the age compositions of the tuning fleet. See Figure 10a for the details on each plot.



**Figure 2.13.** Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Model fits to the acoustic survey index of the *Reference\_run\_SD32\_survey* model configuration.

The results of the runs test are presented in Figures 2.14 and 2.15. The RMSE runs test (Carvalho *et al.*, 2021 for details) indicated that the fit was satisfactory as no residuals were larger than 1 and the RMSE was below 30%, indicating the presence of a random pattern in the length frequency distributions and in the survey indices. The RMSE plot is frequently used as a tool for identifying trends in residuals, and if the standard deviation is small on a given year this means the fleets included in the model agree, even if not fitting well, which is a useful diagnostic. Its purpose is to visualize multiple residuals at once, pick up on periods of substantial data conflicts (width of boxes) and systematic departures in median residuals (loess smoothers). The ordinary runs tests were all passed.



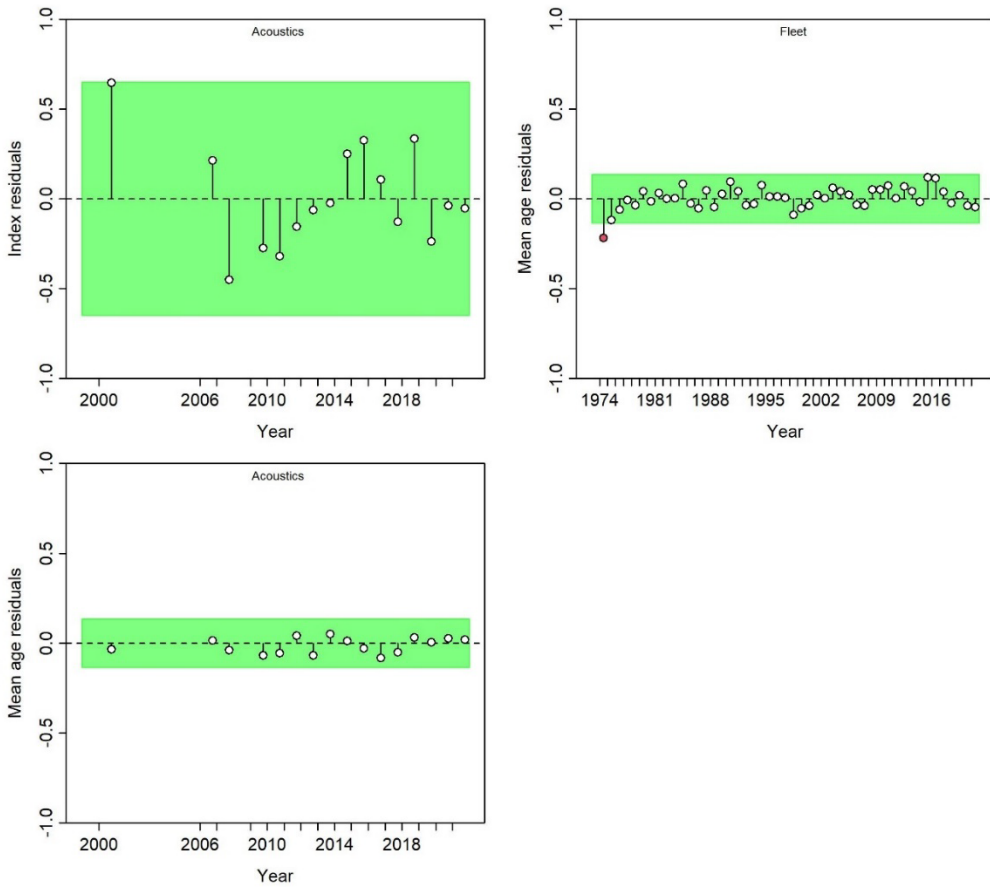


Figure 2.14. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (central Baltic Sea). Residuals from runs test analyses for the fit to the acoustic survey index and age distributions of commercial fleet and survey.

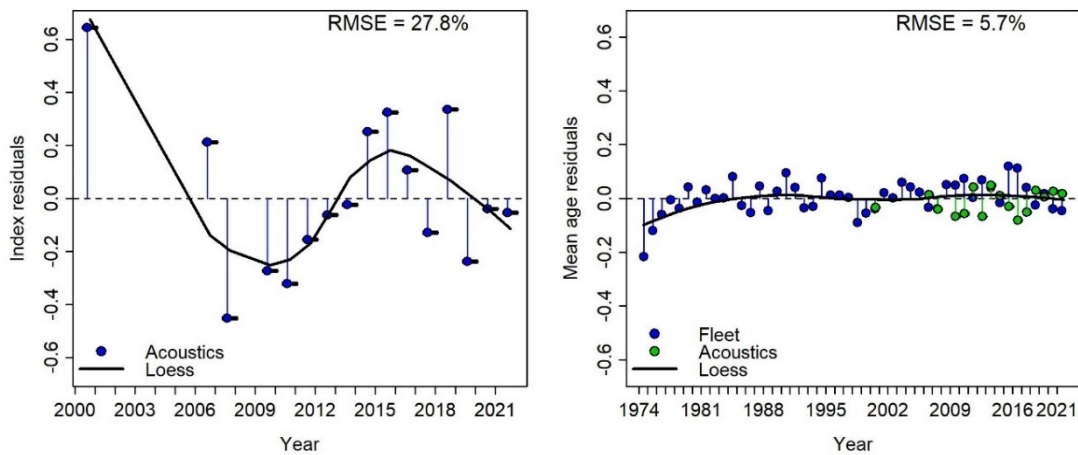
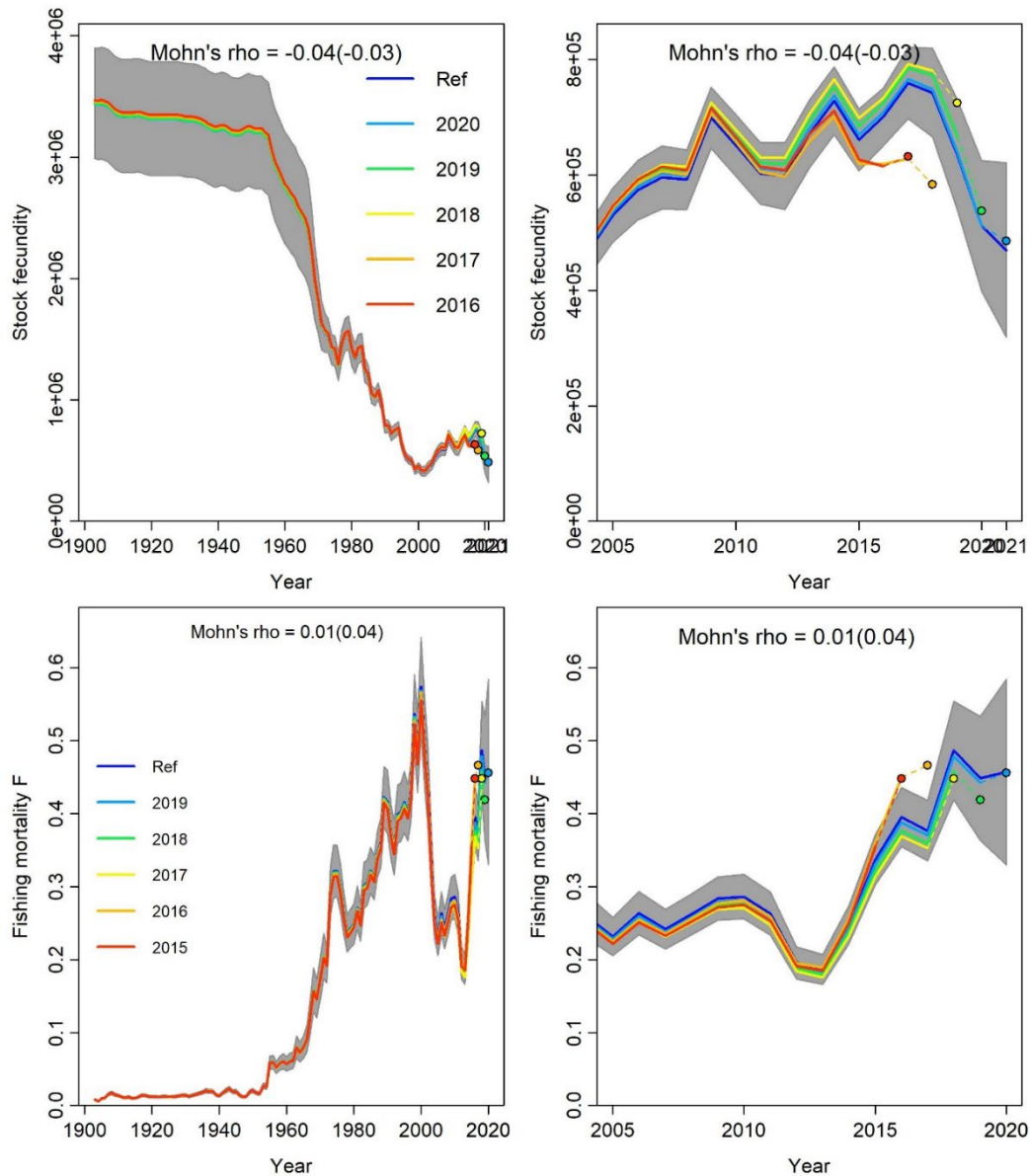


Figure 2.15. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Residuals from the RMSE runs test analyses for the age distributions and the fit to the acoustic survey index.

The retrospectives of the reference model were rather stable (Figure 2.16). The estimated Hurtado-Ferro *et al.* (2014) variant of the Mohn’s rho indices were inside the bounds of recommended values for long-lived species for both SSB (0.03) and F (0.03).



**Figure 2.16. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Retrospective analyses of the base case model.**

The results of the hindcasting showed that the survey performs well given that the MASE value is lower than the 1.0 threshold when predicting the index one year ahead. All MASE values for the age compositions of the catches and the acoustic survey (including the joint MASE) are well below the 1.0 (Figure 2.17 and 2.18).

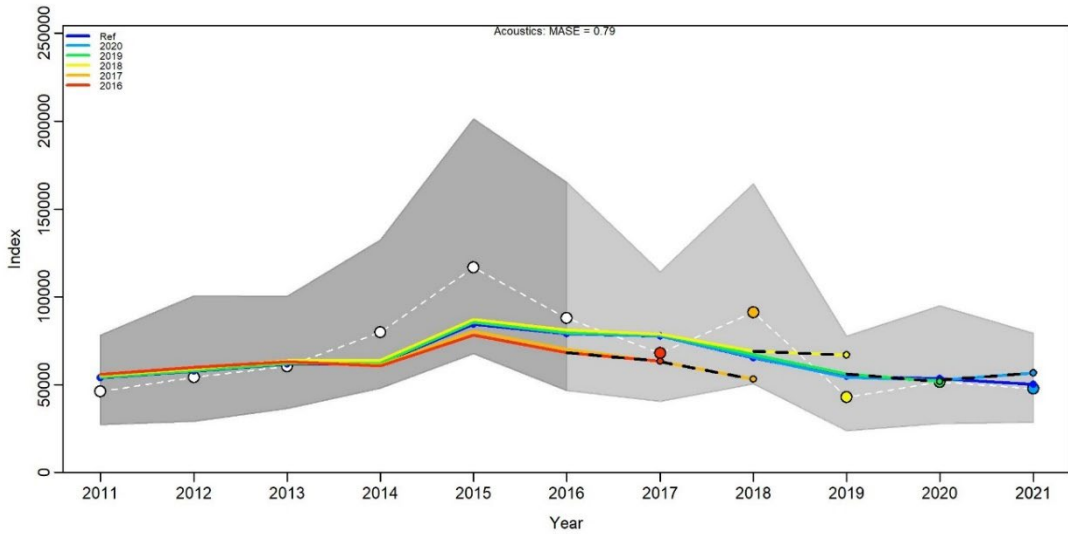


Figure 2.17. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Hindcasting results for survey showing observed (large white points connected with dashed line), fitted (solid lines) and one year-ahead forecast values (small terminal points). Hindcasting was performed using one reference model (Ref equal to last year data 2021) and 5 hindcast runs (solid-coloured lines) relative to the expected index. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95 % confidence intervals (light-grey shading). The mean absolute scaled error (MASE) score associated with the acoustic survey index is denoted in each upper part of the panel.

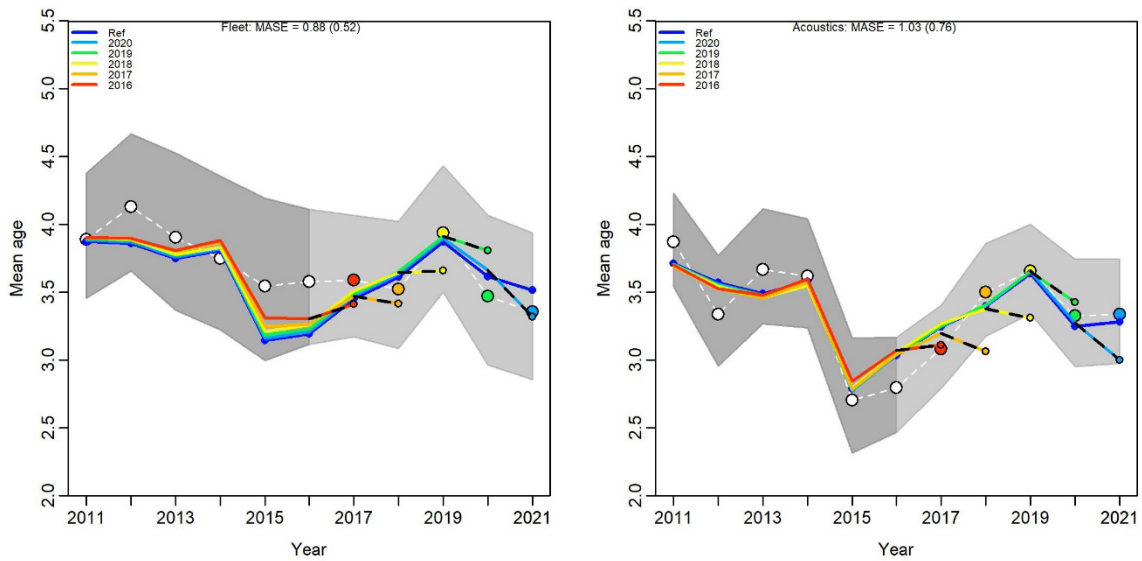


Figure 2.18. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Hindcasting results for age compositions showing observed (large white points connected with dashed line), fitted (solid lines) and one year-ahead forecast values (small terminal points). Hindcasting was performed using one reference model (Ref equal to last year data 2021) and 5 hindcast runs (solid-coloured lines) relative to the expected index. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95 % confidence intervals (light-grey shading). The mean absolute scaled error (MASE) score associated with the acoustic survey index is denoted in each upper part of the panel.

### 2.4.2.5 Trends in SSB, F and R of the reference model

The stock status and the trends in SSB, R and F are based on the MLE model. The probabilistic Kobe plot with stock status in the last year of the assessment (2021) showed that the stock is overfished and subject to overfishing with 99.7% probability (Figure 2.19). It is however important to note that the reference point is set to  $B_{30\%}$  only for representation and that reference points will be estimated through Management Strategy Evaluation (MSE). The spawning stock biomass (SSB) showed a declining trend from the beginning of the 1960s to the 2000s, it then increased during the beginning of the 2000s, but declined again from 2015 and onwards. Fishing mortality (F) showed a similar pattern to SSB, and it has increased markedly from beginning of the 1960s to the 2000s, with a decrease between 2000s and 2015s and has increased hereafter to remain at high levels in recent years. Recruitment (R) shows a decreasing trend in the last 20 years. In 2014 a very strong year class appeared (Figure 2.20).

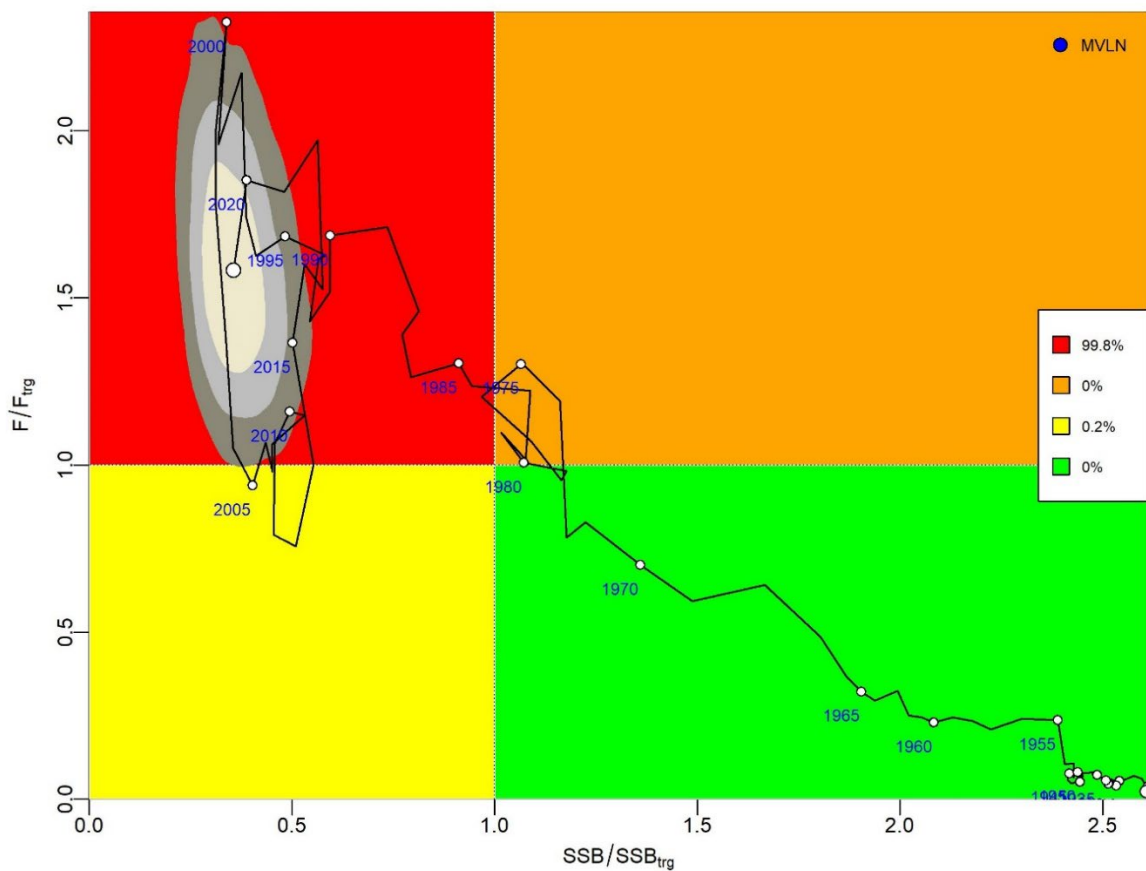


Figure 2.19. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Probabilistic Kobe plot with stock status in the last year of the assessment (2021). The legend indicates the estimated probability of the stock status being in each of the Kobe quadrant. The reference point is set to  $B_{30\%}$  only for representation.

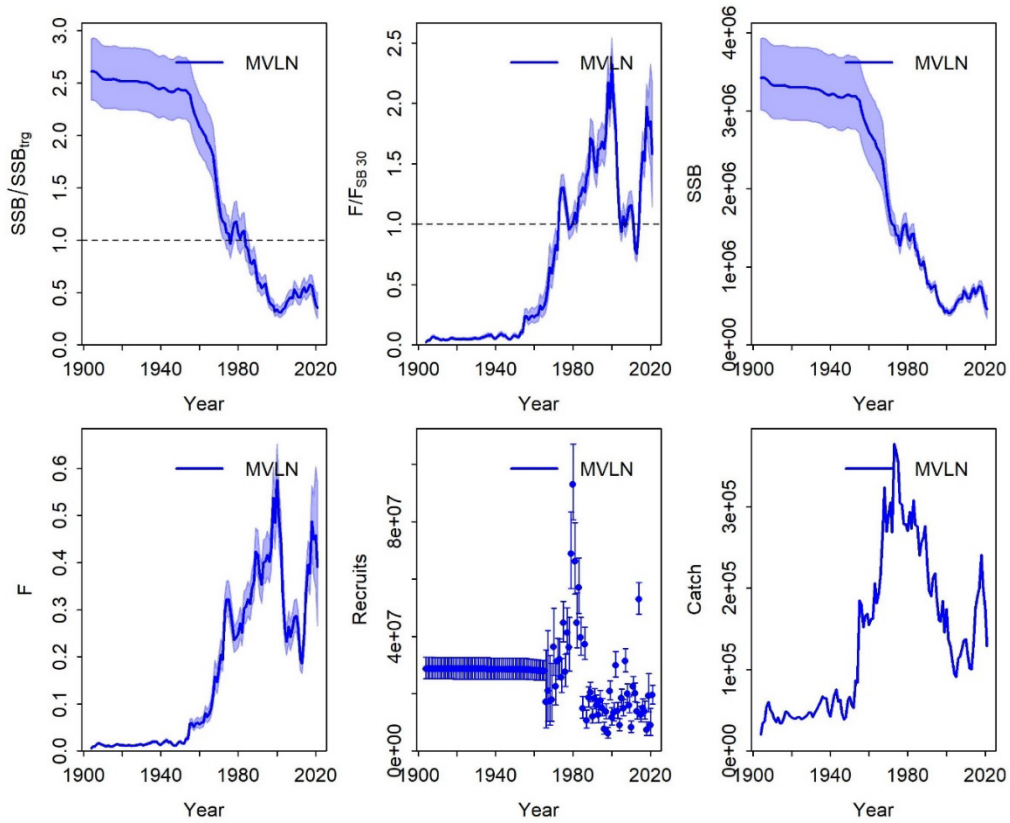


Figure 2.20. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Summary of the stock assessment. SSB, F and R with 95% confidence intervals. Catches by fleet and SSB are in tonnes R is in thousands of individuals. The reference point is set to  $B_{30\%}$  only for representation.

#### 2.4.2.6 Additional runs tested

The reference model, which has selectivity of the ages 5-8+ fixed to be logistic, was compared to a model configuration that allows the selectivity of ages 5-8+ to be freely estimated. This improves, as expected, the fit of the age compositions (Figure 2.21). We analysed the OSA residuals of this alternative model configuration, which are shown in Figure 2.22. There is an improvement in the OSA, especially for ages 1 to 4. When the two models are compared in terms of diagnostics, all testes are passed, and while the alternative model configuration has a smaller retrospective bias, the MASE of the ages is still better for the reference run (Table 2.5). Based on the MASE of the ages, it was decided to keep the reference run as the basis for the ensemble.

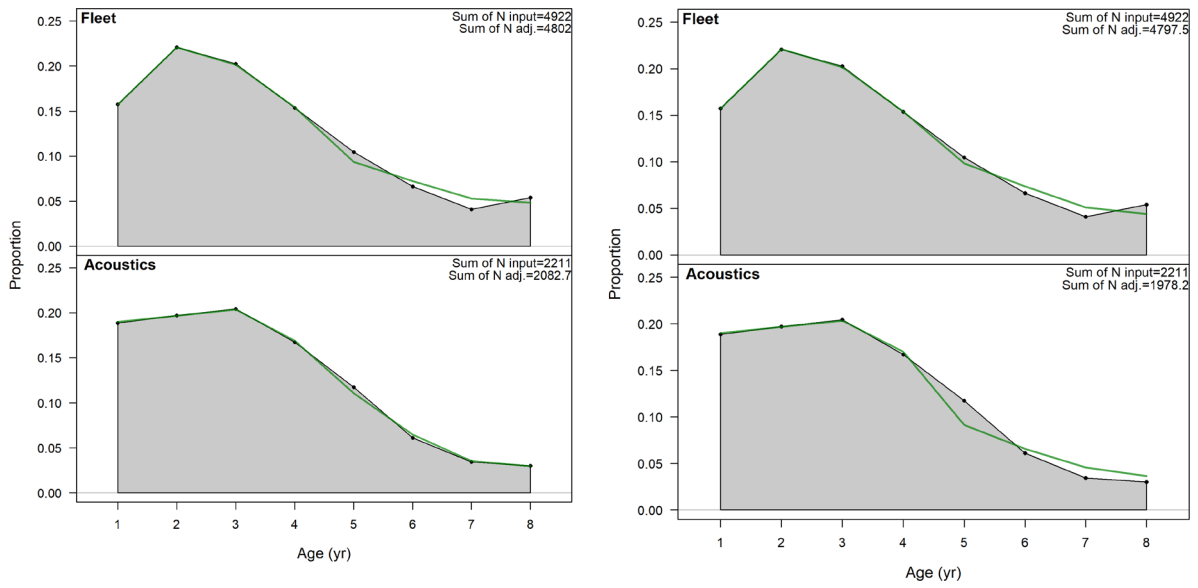


Figure 2.21. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (central Baltic Sea). Model fits to age composition data of the *Reference\_run\_SD32\_survey* model configuration (right) against the model configuration that allows the selectivity of ages 5-8+ to be freely estimated (left).

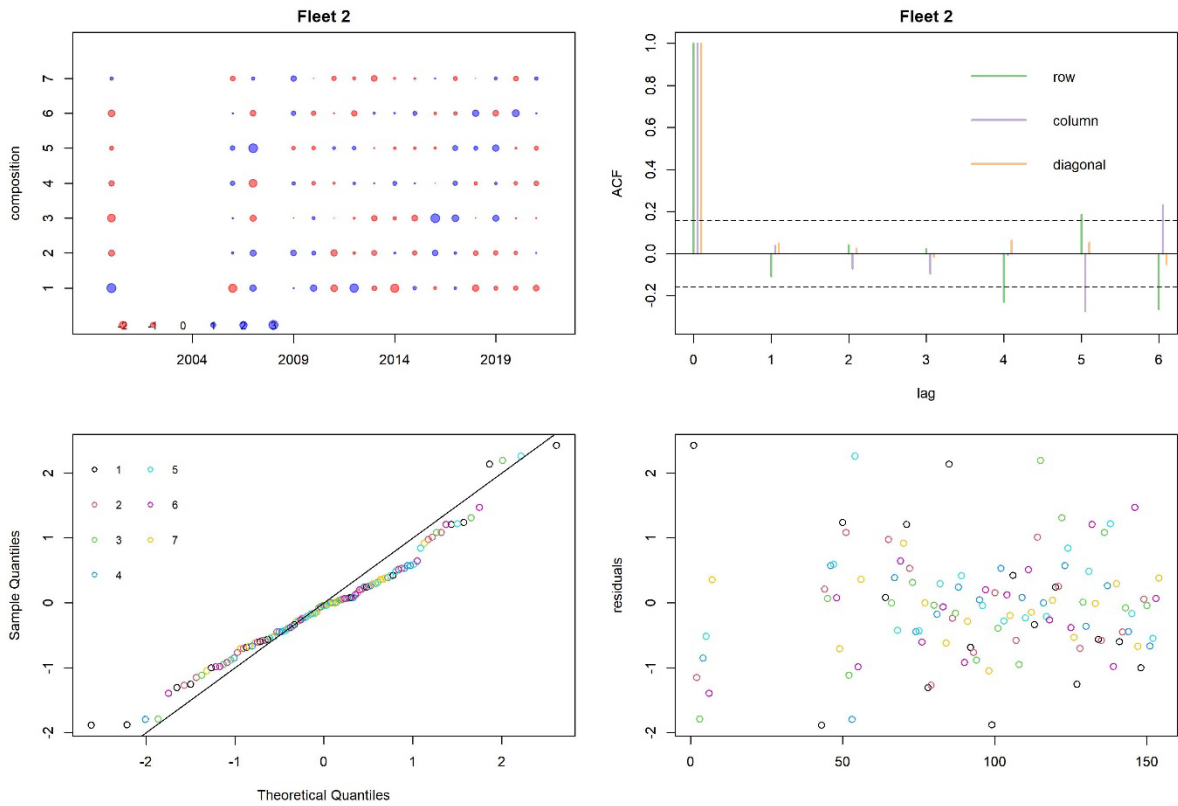


Figure 2.22. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). One-step-ahead (OSA) residuals diagnostic plots for the age compositions of the tuning fleet for the model configuration that allows the selectivity of ages 5-8+ to be freely estimated. See Figure 2.12a for the details on each plot.

**Table 2.5. Central Baltic herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (central Baltic Sea). Summary table of the diagnostics of the reference run against the model configuration that allows the selectivity of ages 5-8+ to be freely estimated. Values in green refer to best test although all tests were passed for both models.**

Run	Reference run	Dome-shaped selectivity survey
Convergence	0.0010	0.0004
Total_LL	1033.7	1022.18
N_Params	68	71
Runs_test_cpue1	Passed	Passed
Runs_test_age1	Passed	Passed
Runs_test_age2	Passed	Passed
RMSE_Perc	27.8	29.6
RMSE_Perc_2	5.7	5.4
Retro_Rho_SSB	-0.04	0
Forecast_Rho_SSB	-0.03	0.01
Retro_Rho_F	0.04	-0.02
Forecast_Rho_F	0.07	0
MASE_cpue1	0.79	0.78
MASE_cpue2	0.79	
MASE_age1	0.52	0.70
MASE_age2	0.76	0.92
MASE_age3	0.64	0.82

## 2.4.3 Ensemble model

### 2.4.3.1 Why use an ensemble model?

The main input parameters of a stock assessment are often uncertain. This means that stock assessors are often faced with a range of model formulations and/or alternative management scenarios which should be scrutinized before decisions are made (Mannini *et al.*, 2021). In this context, when discussing which could be the best model used in assessing stocks, Hilborn and Walters (1992) recalled an adage that “the truth often lies at the intersection of competing lies”. This uncertainty in ‘what is the best model?’ necessitates a comparison of a range of alternative models.

The biggest novelty used in this benchmark assessment is that, instead of comparing multiple model outputs and selecting a single final model, an ensemble modelling approach (Dietterich, 2000) was used to present results with a quantitative criterion for weighting several model predictions. Ensemble methods provide a promising approach when decisions have to be made despite the presence of multiple and potentially conflicting estimates of stock status (Anderson *et al.*, 2017). Ensemble models have been proven to be more accurate and less biased than the choice of an individual model, as they can effectively tease apart the conditions under which various model assumptions result in the most accurate predictions. In general, an ensemble approach will better encapsulate the variability and uncertainty of model predictions because instead of choosing a single set of fixed parameter values, you can explore a contrasting but plausible range of values. (Dietterich, 2000; Tebaldi & Knutti, 2009). This is crucial when the reliability of single fixed parameters is in question. The objective when using an ensemble model is therefore to quantify the total uncertainty across all plausible models, where the structural uncertainty is likely to be much greater than the within-model uncertainty. For example, ensembles are often helpful because modellers need not decide on dome versus asymptotic fisheries selectivity (e.g. Sampson & Scott, 2012, FAO-GFCM, 2021), or whether to fix or estimate natural mortality (e.g. Johnson *et al.*, 2015). Moreover, ensemble forecasting has been proven to improve forecast

accuracy, robustness in many fields, particularly in weather forecasting where the method originated (Wu and Levinson, 2021).

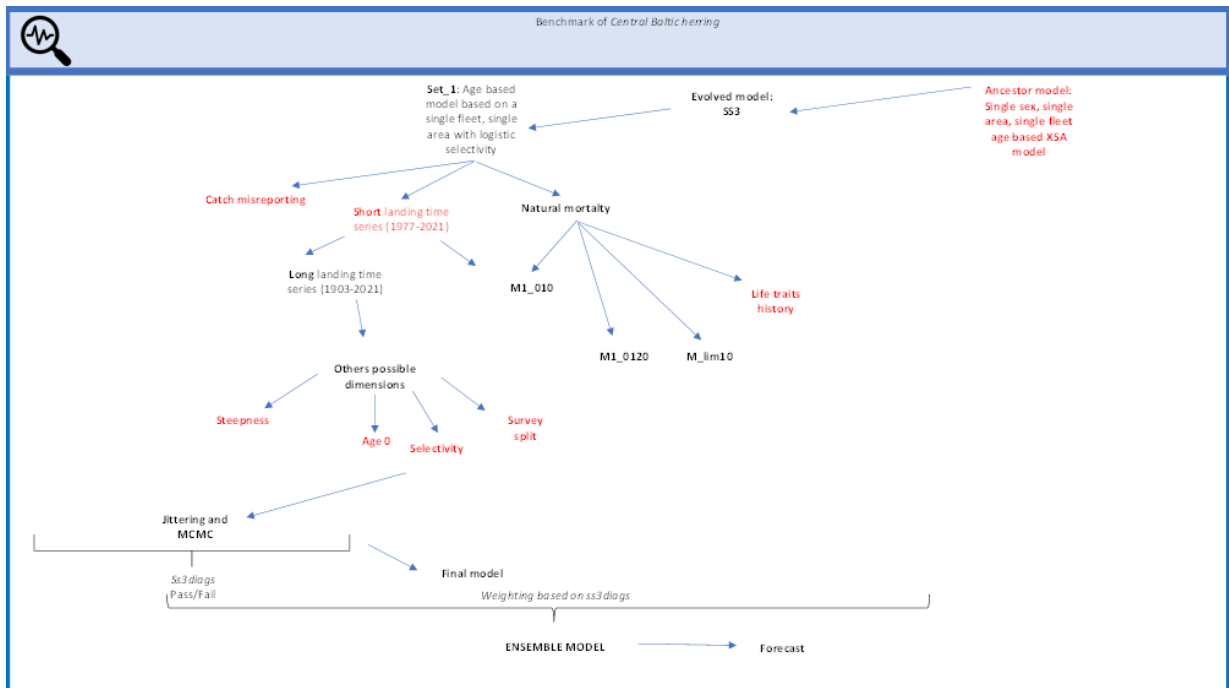
### 2.4.3.2 Parameter levels

Based on the importance of considering both structural and parameter uncertainty, an ensemble approach was selected as the best solution by experts. This is because an ensemble can theoretically represent all plausible “states of nature” of the stock under analysis, based on selected main sources of uncertainty, which in this case was identified in natural mortality (M; see section YYY and Morten WD). Therefore, alternative hypotheses on M are reasonable, given that M is considered one of the most difficult parameter to estimate, yet most influential parameters in stock assessment (Mannini *et al.*, 2021). The final model grid for the ensemble included three alternative values for M as listed in Table 2.6. Input files for each of the runs can be found in the official ICES SharePoint (WKPELA).

**Table 2.6. Herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Parameters and levels employed in the final ensemble assessment grid.**

Parameter	Levels	Progressive number of runs	Values
Natural mortality ( <i>M</i> )	3	3	M1_010
			M1_020
			M_lim10

A schematic graphical representation of the assessment workflow is provided in Figure 2.23. Its inclusion is designed to provide a guideline via which the process of ensemble model grid construction can be followed as well as the steps taken prior to its implementation.



**Figure 3.23. Herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Schematic graphical representation of the assessment workflow.**



### 2.4.3.3 Model weighting

The need to weigh models based on information in the available data is well recognised (Francis and Hilborn, 2011), but it is often difficult to do so within the context of fisheries stock assessment models, as their complexity prevents strict adherence to statistical rigor. In this context, the selected four runs are considered to represent the alternative ‘states of nature’ of the stock and must be weighted in the final ensemble model. This is a necessary step because assigning the same weight (reliability) to all hypotheses could introduce biases into the management advice if some hypotheses are, in fact, highly unlikely. To assign weights to the various models and hypotheses, it is preferable to establish a system of discrete weight categories. In this benchmark assessment, we decided to use diagnostic scores ( $W(\text{Diagnostics})$ ) as weighting metrics (Maunder *et al.*, 2020) to judge the plausibility of each candidate model based on each model’s fit. In fact, when all diagnostic tests are considered together, the power to detect model misspecification improves without a substantial increase in the probability of incorrectly rejecting a correctly specified model (Carvalho *et al.*, 2017). In this context, the  $W(\text{Diagnostics})$  component is calculated based on a series of interconnected diagnostic tests as discussed by Carvalho *et al.* (2021) and previously presented and explained in Section 2.4.2.4 for the reference run:

$$W(\text{Diagnostics}): \frac{W(\text{Diags 1}) + W(\text{Diags 2}) + W(\text{Diags 3}) \dots + W(\text{Diags N})}{\text{Num of } W(\text{Diags})}$$

where each  $W$  component is assigned a value of 1 when the run passes the diagnostic test and a 0 when it fails. A summary of all main diagnostics for the four model runs is provided in Table 2.7. Based on these results, different weights were used to stitch together the different runs in the final ensemble model.

The  $W(\text{Diagnostics})$  values are used as a scaling factor for the number of simulations used by the delta-MVLN estimator (i.e. 5000 simulations when the  $W(\text{Diagnostics})$  value is 100% and less according to the assigned weight such that a  $W(\text{Diagnostics})$  value of 50% would have 2500 simulations) when estimating the posterior distributions of the derived quantities.

### 2.4.3.4 Delta-MVLN estimator

To address structural uncertainties, the delta-Multivariate log-Normal (delta-MVLN) estimator (Walter and Winker, 2019; Winker *et al.*, 2019) was used to generate and stitch together the joint posterior distributions of the target-derived quantities (e.g. SSB/SSBtarget and F/Ftarget) from plausible models. It infers within-model uncertainty from maximum likelihood estimates (MLEs), standard errors (SEs) and the correlation of the untransformed quantities, and it has been demonstrated to be able to mimic the Markov Chain Monte Carlo (MCMC) closely (Winker *et al.*, 2019). These quantities are derived by using the delta-method to calculate asymptotic variance estimates from the inverted Hessian matrix of the Stock Synthesis model.

**Table 2.7. Herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Summary table of the diagnostics used in the weighting procedure. Green refers to a “Passed” score.**

Run name	Positive		Goodness of the fit					Consistency				Prediction skills				Weight	
	Hessian	Convergence	Run test and RMSE					Retrospective analysis				Hindcasting (MASE)					
			Survey	Age1	Age2	Survey	Ages	Retro_SSB	Forecast_SSB	Retro_F	Forecast_F	Survey	Age1	Age2	Joint		
Run1	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	1.00
Run2	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	1.00
Run3	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	1.00

### 2.4.3.5 Model results

To recap, to capture structural uncertainties, a range of alternative models were selected through diagnostics (interconnected diagnostic tests; Carvalho *et al.*, 2021; Maunder *et al.*, 2020; Kell *et al.*, 2021) and were stitched together in an ensemble using the delta-Multivariate log-Normal

estimator (delta-MVLN; Walter and Winker, 2019; Winker *et al.*, 2019). The run specifications and final weighting factors used in the ensemble procedure are reported below. The final outputs from the ensemble model are based on the weighted-median value of the three runs.

Name	Natural Mortality	Weighting
run1*	M1_010	1.00
run2	M1_020	1.00
run3	M_lim10	1.00

\*Reference run of the Stock Synthesis assessment model for Herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea) is described in detail in Section (2.4.2.4)

The ensemble model, based on the three runs proposed during the benchmark, has been proposed as the final model for providing scientific advice. Figures 2.25 present the main outputs from the final ensemble model, whereby Figure 2.24 provides a comparison between the three model runs. The main trends from the ensemble are summarized as:

- **State of the adult biomass (SBB):** Total spawning biomass of Central Baltic herring has declined from the beginning of the 1960s to a minimum in the beginning 2000s, thereafter it has slightly recovered but it declined again in the latest years.
- **State of exploitation (F):** Fishing mortality is defined as the average F of age classes 3 to 6. Since the beginning of 1960s F has generally increased with a peak in 2000 and 2018, remaining at high levels thereafter.
- **State of the juveniles (Recr):** Large year classes were observed in the 1980s. With the exception of the 2015-year class, recruitment has been low in the last decade.

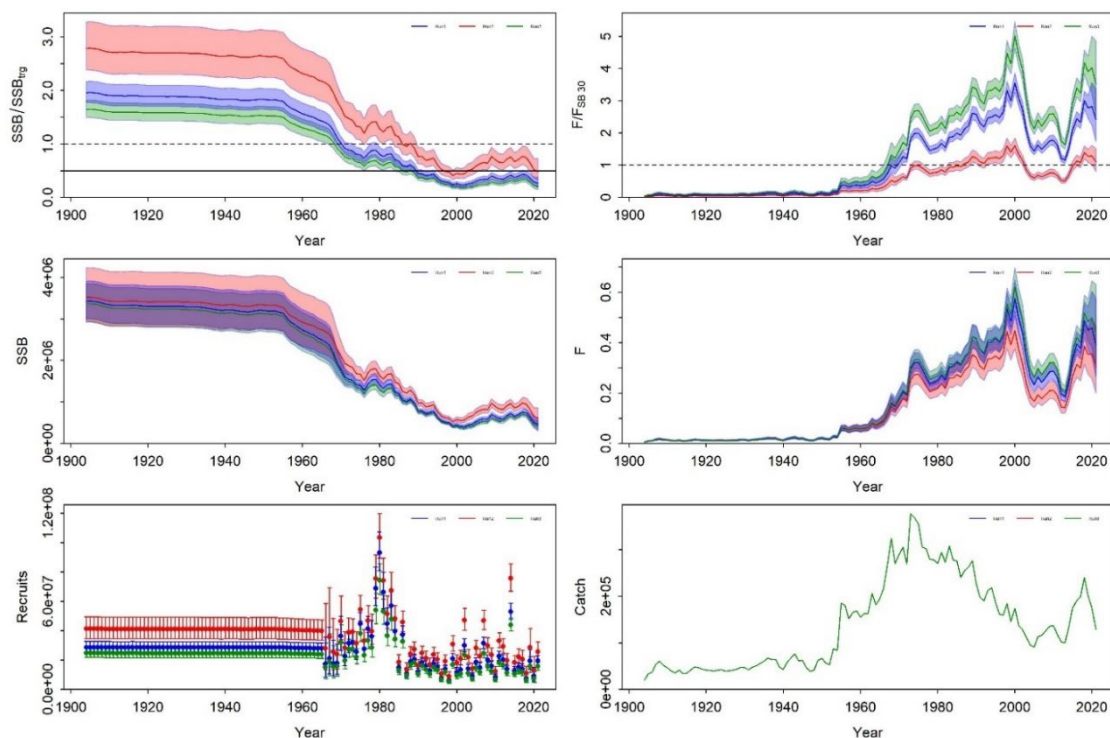
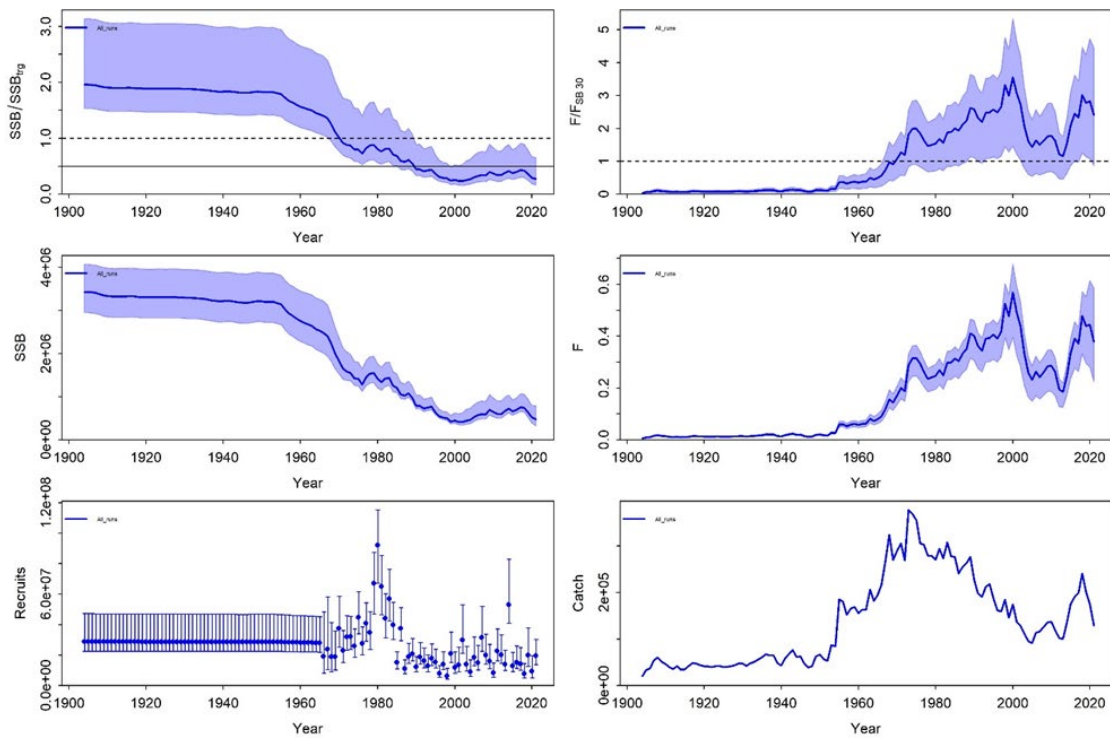


Figure 2.24. Herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Comparison of stock assessment results, SSB, F and Recr with 95% confidence intervals, across the 3 runs included in the ensemble. Trajectory of the stock and fishing mortality is compared to the reference points  $B_{30\%}$  and  $B_{lim}$  which is set as 15% of  $B_0$  (top figures). Dashed line is the  $B_{trg}$  and continuous line is  $B_{lim}$ .



**Figure 2.25. Herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Stock assessment results of the final ensemble. Weighted-median value of SSB, F and Recr with 95% confidence intervals from delta-MVLN. Trajectory of the stock and fishing mortality is compared to the reference points B<sub>30%</sub> and B<sub>lim</sub> which is set as 15% of B<sub>0</sub> (top figures). Dashed line is the B<sub>trg</sub> and continuous line is B<sub>lim</sub>.**

A Kobe plot for the ensemble model is presented in Figure 2.26. The Kobe plot considers the time-series of pressure ( $F/F_{target}$ ) on the y-axis and the state of the stock's biomass ( $SSB/SSB_{target}$ ) on the x-axis. The reference point is set to B<sub>30%</sub> only for representation and reference points will be estimated through Management Strategy Evaluation (MSE). The orange area indicates healthy stock sizes that are about to be depleted by overfishing. The red area indicates ongoing overfishing and that the stock is too small to produce maximum sustainable yields. The yellow area indicates that the biomass is too small/still recovering and that a reduction in fishing pressure is needed. The green area is the target area for management, indicating sustainable fishing pressure and a healthy stock size capable of producing high yields close to the chosen reference points (MSY or proxies).

The stock trajectory began in 1903 in the downright quadrant (i.e. green quadrant of the Kobe plot), when the biomass was higher compared to the reference points. In the period 1960–2000, the F level increased which resulted in a progressive erosion of the stock size, moving the stock trajectory towards the up-left quadrant (i.e. red quadrant of the Kobe plot). Following this, F remained above the F reference point and below the SSB reference point thereafter. For this reason, and over the last 40 years, the stock has been in the red quadrant of the plot. In 2021 there was an approx. 90% probability that the stock is in the red quadrant of the Kobe plot (i.e.  $SSB > SSB_{30}$  and  $F < F_{30}$ ) with a lower probability (approx. 10%) of being in the yellow (i.e.  $SSB < SSB_{30}$  and  $F < F_{30}$ ) and 0% probability of being in the green ( $SSB < SSB_{30}$  and  $F > F_{30}$ ).

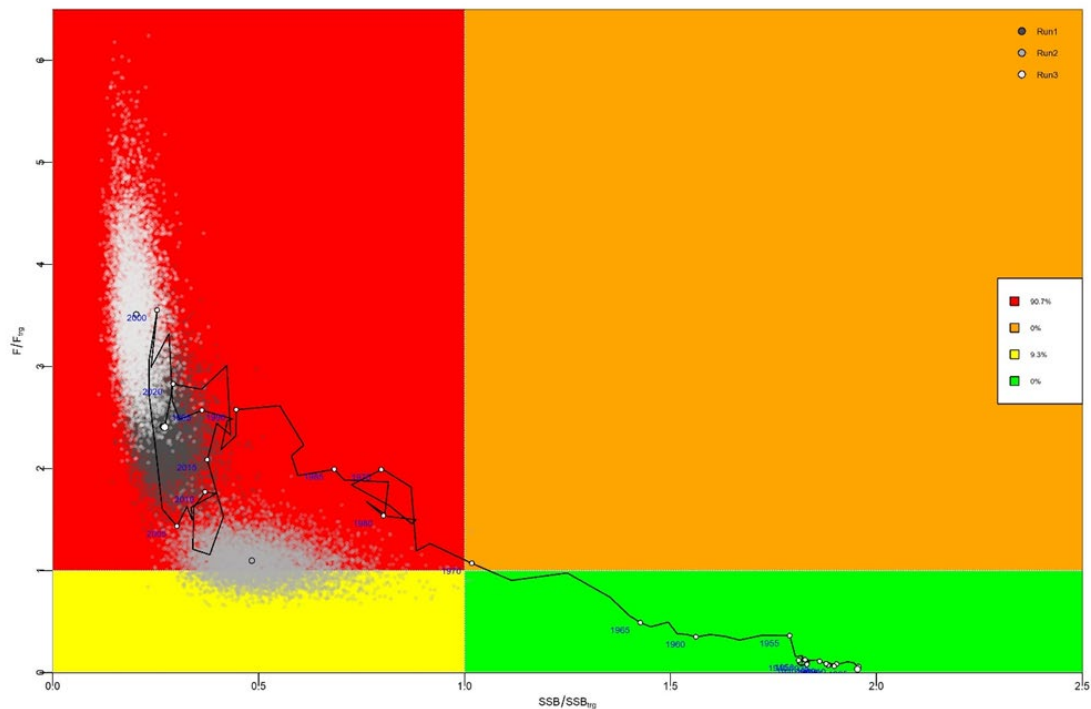


Figure 2.26. Herring in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Kobe plot showing the trajectory of relative stock size (SSB/SSB30) over relative exploitation (F/F30) based on the final ensemble model (white dot: the weighted-median value of 3 runs). The points represent 5000 iterations from delta-MVNL of the final assessment year (2021).

## 2.5 Short-term projections

Settings for short-term projections, used to provide catch advice, were discussed at the meeting. In SS3 it is possible to forward project the population under a range of catch and F scenarios. The agreed settings are indicated in the Stock Annex.

## 2.6 MSY reference points

MSE were used to determine the target and trigger reference points to be used to provide advice for Central Baltic herring (*Clupea harengus*; CBH; her.27.25-2932) in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). The same methodology has been recently used to determine reference points for Northern shrimp in division 3.a and 4.a East (ICES 2022a). To this aim, we used the simulation-testing framework available in the Fisheries Library for R (FLR; Kell *et al.*, 2007; <https://flr-project.org/>). The simulation framework was implemented in the FLR library `mse` (<https://github.com/flr/mse>) with `FLasher` (<https://github.com/flr/FLasher>) being used to carry out the forward projections. Reference points at equilibrium were calculated with `FLBRP` (<https://github.com/flr/FLSRTMB>). To facilitate customized reference point estimation and visualisation of  $F_{MSY}$  proxy (along to the true  $F_{MSY}$  estimated in a stochastic framework) (hereafter defined as  $F_{brp}$ , which in this case was expressed as the F that brings the stock at a given fraction of  $B_0$ , i.e.  $F_{B\%}$ ),  $B_{lim}$ ,  $F_{P.05}$ ,  $B_{trg}$ ,  $F_{trg}$ , we developed the FLR package `FLRef` (<https://github.com/henning-winker/FLRef>). `FLRef` makes use of the new fast forward projection `ffwd()` in `FLasher` together with the bisection function `bisect()` in `mse` to efficiently derive precise values of  $F_{P0.5}$  based stochastic simulations. R code used in this analysis will be made

available in the GitHub repository of `FLRef` and in the ICES SharePoint of WKBBALPELA.  $F_{MSY}$  is also calculated and compared to the other proxies.

The simulations were run for the 3 models included in the ensemble (hereafter defined as Run1, Run2 and Run3). The operating models were implemented as single sex and single fleet models with an annual time step. Future projections were run over 60 years (i.e. 2021-2080) with 250 iterations and based on the 10-years average of the most recent data years for weight-at-age, maturity-at-age, natural mortality-at-age and the F pattern determining the selectivity-at-age. This choice was made to account for non-stationary processes in these quantities and to align the biology with the calculation of  $B_0$  which is estimated as unexploited biomass at current conditions. The performance evaluations were based on the last 10 years of the 60-year projection horizon (i.e. 2071-208;  $10 \times 250 = 2500$  observations). For the simulation testing, stock and recruitment, steepness,  $\sigma R$  and autocorrelation were set as equal to the one derived for each model of the ensemble. The recruitment deviation is assumed to be associated with a first-order autocorrelation (AR1) process and a function of recruitment standard deviation  $\sigma$  and the AR1 coefficient  $\rho$  (Johnson *et al.*, 2016) which are both estimated from each model of the ensemble. Simulations included implementation error as estimated using the last 3 years (average +16.5%,  $sd = 0.149$ ; ICES, 2022b) or no implementation error (average 0%,  $sd = 0.149$ ). We used a bias instead of a random noise implementation error as:

1. ICES advice and EU TACs have decreased over the last five years, while the Russian quota has been relatively constant. This means that there is a high likelihood of directional implementation error also in the future as Russia is not part of TAC setting system and the stock is below  $B_{lim}$ .
2. Irrespective of the Russian quota having been rather constant around 25 000 tonnes per year over the last 3 years, EU has set a TAC resulting in a total catch (EU + Russian quota) exceeding ICES advice. The directional implementation error was estimated around +16% as the average of the last 3 years, which was considered substantial, especially as the stock is below  $B_{lim}$ .
3. The reference points were calculated as part of an MSE, so that it is natural to use directional implementation error in this case.

To simulation-test reference points with and without implementation error is equivalent to simulation-testing reference points under any other plausible scenario. The selected harvest control rule was not chosen in an ad hoc manner, it was shown to perform best under the most plausible scenario, which was considered by the WG to be the one with implementation error and therefore reference points were selected accordingly (as the values of F and  $B_{trigger}$  with the best catch outcome given the constraint of meeting the conservation requirements). It is also important to note that the catch overshoot was accounted for in the stock assessment (i.e. if the actual catches are included in the assessment) so there was not an *ad hoc* adjustment to the harvest control rule to try to anticipate overshooting.

Harvest control rules (HCRs) are kept generic and in the same form of the conventional ICES Advice Rule (ICES, 2021a), where the advice decreases from  $F_{trg}$  to zero and from  $B_{trigger}$  to zero SSB. Variations of the tested HCRs are therefore determined by the parameters  $F_{trg}$  and  $B_{trigger}$ . The HCRs were implemented using a simulated feedback control loop between the implementation system and the operating model, where the implementation system translates the assessment outcome via the HRC into the Total Allowable Catch (TAC) advice (Figure 2.27). The key difference to a simple stochastic risk simulation, such as EQsim, is the simulated feedback control loop between the implementation system and the operating model allows accounting for the lag between the last year of data used in the assessment and the implementation year of TAC advice. In ICES, the implementation system of the harvest control rule assumes that advice is given for year  $y+1$  based on an assessment completed in year  $y$ , which is typically fitted to data up until

year  $y-1$  (ICES, 2020b). Therefore, implementation of the TAC derived through HCR requires projection of the stock dynamics by way of a short-term forecast (Mildenberger *et al.*, 2021). In contrast to a full Management Strategy Evaluation (MSE) simulation design (Punt *et al.*, 2014), this MSE ‘short-cut’ approach (e.g. ICES, 2020b), omits the step of the annual updating of the estimation model (assessment) in the feedback control (Figure 2.27.). Instead, it passes the ‘true’ age-structured dynamics from the OM (or with assumed some error) to the HCR implementation. The merits of a short-cut MSE approach include the incorporation of the lag effect between data, assessment, and management implementation. The limitations of the MSE short-cut approach are that it cannot fully account for uncertainties resulting from imperfect sampling of the full age-structure (e.g. poorly sampled recruits), model estimation error and observation error. On the hand, the MSE is done on the model ensemble and thus consider model structural uncertainty, which differs from when MSE (either short-cut or full MSE) is run on the “best case scenario” only.

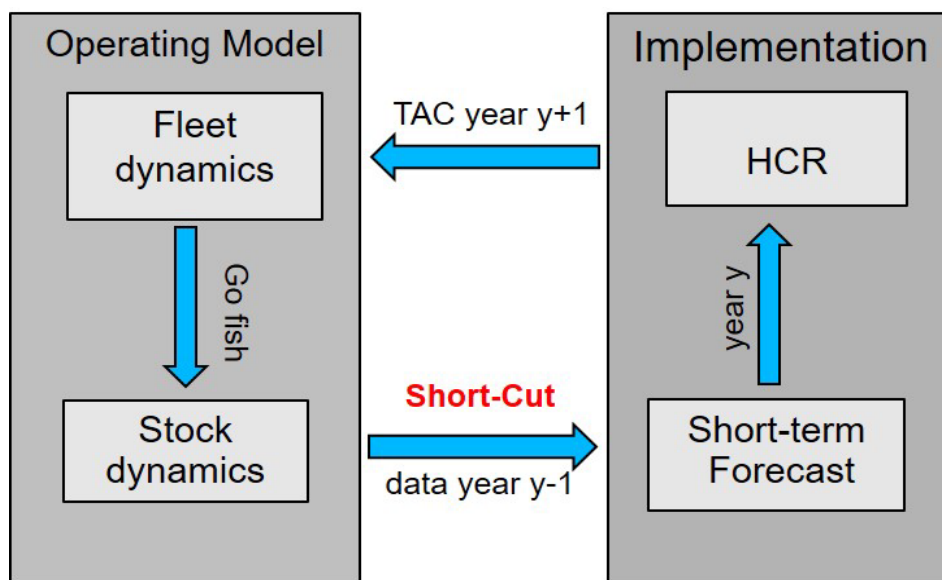


Figure 2.27. Schematic illustrating the key processes of the short-cut approach to MSE, showing the Operating model that simulates the fishery and stock dynamics on the left and Implementation System including the short-term forecast on the right. The short-cut denotes the omission of the estimation (stock assessment) model that updates to new observations (with estimation error) in a conventional MSE implementations with full feedback control loop.

### 2.6.1 Performance Evaluation Criteria

The consistency tests were designed to identify the generic rules for specifying  $F_{brp}$ ,  $B_{trg}$  and  $B_{trigger}$  according to stock-specific productivity that provide the optimal trade-offs among the following two main objectives: (1) to not exceed a 5% probability of SSB falling below  $B_{lim}$  in any single year and (2) to achieve high long-term yields that correspond to at least 95% of the median long-term yield attained by fishing at the deterministic  $F_{MSY}$  (MSY). An additional Objective, (3) to attain  $B_{MSY}$ , is included although (3) is not a conditional objective for the selection of the reference points, but it is used as an additional criterion when two or more candidates set of reference points have equal performances when considering criteria (1) and (2). Consistent with the objectives of ICES advice framework (ICES, 2020), the two objectives are interpreted hierarchically in that objective (1) is the overriding criteria of maintaining stock size above  $B_{lim}$  with at least 95% probability to be compliant with the ICES Precautionary Approach (PA). Conditional on objective (1), objective (2) is based on the ICES definition for using plausible values around  $F_{MSY}$  in the advice rule, which are derived so that they lead to no more than a 5% reduction of MSY obtained by fishing at  $F_{MSY}$  in the long term.

In the previous assessments (e.g. ICES, 2020b),  $B_{lim}$  was set as the lowest SSB that has resulted in above-average recruitment, i.e., the year 2002, which also happens to correspond to  $B_{loss}$ . In ensemble models,  $B_{loss}$  will be inherently different for the different model configurations and therefore fractions of  $B_{MSY}$  or  $B_0$  are used (ICES, 2022a). Here we have chosen to define  $B_{lim}$  as a fraction of  $B_0$ , which compared to  $B_{MSY}$  has the advantage to be independent to selectivity. The decision was based on several criteria. When expressed as a fraction of  $B_0$ , those generally ranges from 0.1 to 0.2  $B_0$  (ICES, 2022c). As for the Northern shrimp, the other stock that uses the same methodology (ICES 2022a),  $B_{lim}$  was set at 15% of  $B_0$  ( $B_{15\%}$ ). As shown by WKREF1, setting  $B_{lim}$  well under 10% of  $B_0$  renders  $F_{P.05}$  ineffective for most ICES stocks with or without the use of  $B_{trigger}$  (ICES, 2022c). This is also particularly important in the presence of the Allee effect (i.e. depensation) in exploited fish, which generally was identified to occur when the stock is below 15-25% of  $B_0$  (Perälä and Kuparinen, 2017; Perälä *et al.*, 2021). Following the analysis conducted by Tommi Perala on the output of the run1 of the ensemble, the inflection point when depensation occur with more than 95% probability was estimated around 10% of  $B_0$ . Further, additionally we estimated for comparison  $B_{lim}$  as 40% and 50% of  $B_{MSY}$ , which are used respectively in Canada and USA as proxy of  $B_{lim}$  (ICES 2022c). The 40% and 50% of  $B_{MSY}$  were in average 78% and 98% of  $B_{15\%}$ . Therefore, it was decided to use  $B_{15\%}$  as  $B_{lim}$  in the MSE as it was considered to be in line with both 50%  $B_{MSY}$ , the value used by Northern shrimp and as it is on the right side of the point where depensation occur with more than 95% probability. It is important to note that  $B_0$  is not the virgin biomass at the start of the time-series (i.e. 1903) but the unexploited SSB at current conditions (Bessell-Brown *et al.*, 2022), which are calculated using the biology averaged over the last 10 years. As said above, selectivity does not impact the calculation of  $B_0$ .

## 2.6.2 Results

17 scenarios (i.e.,  $5 \times F_{\%B_0}$  time  $4 \times B_{trigger}$  combinations and the deterministic  $F_{MSY}$ ) were tested for the three models of the ensemble. The SR relationship for the three different models of the ensemble is shown in Figure 2.28 while the reference points for Run1 is shown as an example in Figure 2.29. As an example of the realised simulations, trends in SSB, F, landings, and R for the different combinations of  $F_{target}$  and  $B_{trigger}$  as compared to the deterministic  $F_{MSY}$  are shown for Run1 (Figure 2.30).

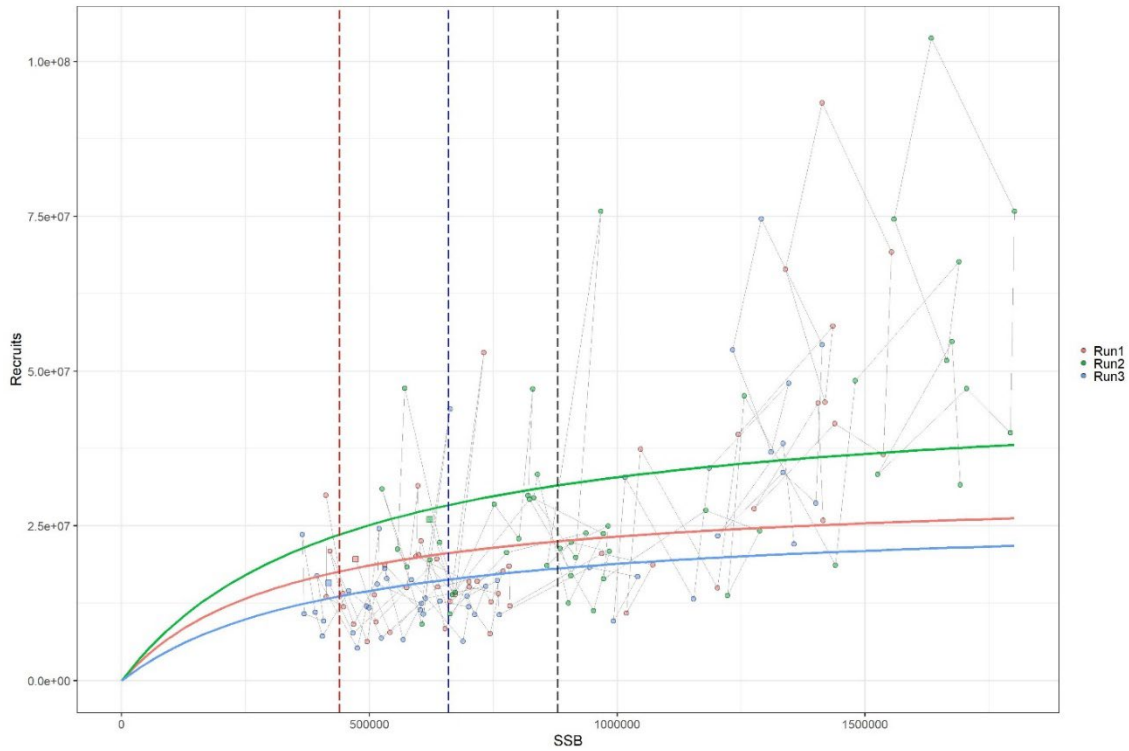


Figure 2.28. Central Baltic herring (*Clupea harengus*; CBH; her.27.25-2932) in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Stock-Recruitment relationship (i.e. Beverton and Holt) for the three models of the ensemble. Red, blue and black line are  $B_{10\%}$ ,  $B_{15\%}$  and  $B_{20\%}$ . Years before 1974 were excluded.

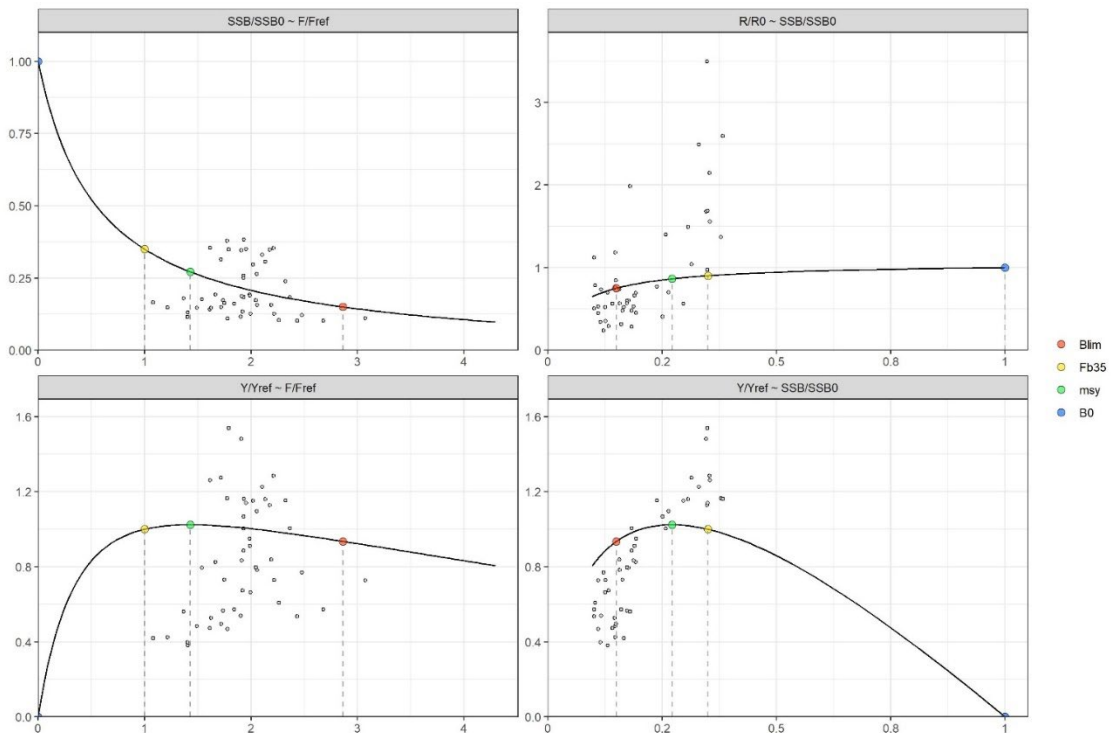


Figure 2.29. Central Baltic herring (*Clupea harengus*; CBH; her.27.25-2932) in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Estimated deterministic  $F_{MSY}$  reference point for Run1 expressed as relative.  $B_{lim}$  is set as 15% of  $B_0$ .



Fishing at  $F_{MSY}$  is precautionary as the probability of SSB to fall below  $B_{lim}$  is less than 5% and achieve catches equal to MSY. However, it implies a median  $F$  larger than  $F_{MSY}$  because of the implementation error and stochasticity. On the other hand, the results of the MSE showed that  $F_{B35\%}$  with  $B_{trigger}$  set at 0.6 of  $B_{35\%}$  achieve high long-term yields and has a median probability of SSB to fall below  $B_{lim}$  which is less than 5% (Figure 2.31 and Table 2.8). The difference in long term yield between  $F_{B35\%}$  with  $B_{trigger}$  set at 0.6 and fishing at the determinist  $F_{MSY}$  is about 1% with a long term SSB that is on average 22% larger than fishing at  $F_{MSY}$ .

It is important to note that several HCRs (indicated in bold green in the tables) do meet the specified criteria defined above. Amongst these, for example, a  $B_{trigger}$  of 60% x  $B_{40\%}$  and  $F$  target  $F_{B40\%}$  achieves long term catches equal to those under the deterministic MSY and results in a median  $B$  that is 20% larger than  $B_{MSY}$  despite a lower  $F$ . This scenario would also fulfil the criteria of the Common Fisheries Policy (CFP) but doesn't necessarily match with the ICES interpretation of it, for which the scenario with the highest  $F$  (that fulfils the above described evaluation criteria) is selected. In reality, the CFP refers to maximum sustainable yield and not to maximum sustainable fishing mortality. Thus, the general ICES interpretation of the CFP might not be fully appropriate and should be revisited by future working groups.

**Table 2.8. Central Baltic herring (*Clupea harengus*; CBH; her.27.25-2932) in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Results of the MSE for the estimation of the reference points with implementation error set to 0.165 with standard deviation equal to 0.149 (a) and implementation error set to 0 with standard deviation equal to 0.149 (b). HCR are the different scenarios tested. AAV is the Average annual variation in catches,  $B/B_{MSY}$  and  $F/F_{MSY}$  are median values of SSB and  $F$  over deterministic  $B_{MSY}$  and  $F_{MSY}$ , Catch/MSY is median catches over catches at the deterministic  $F_{MSY}$ ,  $P3(B < B_{lim})$  is the probability of SSB falling below  $B_{lim}$  in any single year.  $B_{lim}$  is set as 15% of  $B_0$ . See text for further details. In bold are candidates set of reference points. Scenarios in bold red are those that do not pass  $p < 0.05$  the probability of SSB falling below  $B_{lim}$  in any single year, scenarios in light red are those that do not pass  $F < F_{MSY}$ , scenarios in orange are those that do not pass  $B > B_{MSY}$  and finally, scenarios in bold green those that pass all of the above criteria.**

(a)

Variable	HCR	Median	Variable	HCR	Median	Variable	HCR	Median	Variable	HCR	Median	Variable	HCR	Median
AAV	true.Fmsy	0.15	$B/B_{MSY}$	true.Fmsy	0.81	Catch/MSY	true.Fmsy	1.00	$F/F_{msy}$	true.Fmsy	1.183	$P3(B < B_{lim})$	true.Fmsy	0.0680
AAV	fmsy.bt06	0.15	$B/B_{MSY}$	fmsy.bt06	0.82	Catch/MSY	fmsy.bt06	1.00	$F/F_{msy}$	fmsy.bt06	1.169	$P3(B < B_{lim})$	fmsy.bt06	0.0600
AAV	fmsy.bt08	0.15	$B/B_{MSY}$	fmsy.bt08	0.85	Catch/MSY	fmsy.bt08	1.00	$F/F_{msy}$	fmsy.bt08	1.123	$P3(B < B_{lim})$	fmsy.bt08	0.0320
AAV	fmsy.bt1	0.16	$B/B_{MSY}$	fmsy.bt1	0.90	Catch/MSY	fmsy.bt1	1.00	$F/F_{msy}$	fmsy.bt1	1.044	$P3(B < B_{lim})$	fmsy.bt1	0.0120
AAV	fb20.bt06	0.16	$B/B_{MSY}$	fb20.bt06	0.56	Catch/MSY	fb20.bt06	0.92	$F/F_{msy}$	fb20.bt06	1.751	$P3(B < B_{lim})$	fb20.bt06	0.4960
AAV	fb20.bt08	0.16	$B/B_{MSY}$	fb20.bt08	0.59	Catch/MSY	fb20.bt08	0.94	$F/F_{msy}$	fb20.bt08	1.676	$P3(B < B_{lim})$	fb20.bt08	0.3840
AAV	fb20.bt1	0.17	$B/B_{MSY}$	fb20.bt1	0.64	Catch/MSY	fb20.bt1	0.95	$F/F_{msy}$	fb20.bt1	1.558	$P3(B < B_{lim})$	fb20.bt1	0.2960
AAV	fb25.bt06	0.16	$B/B_{MSY}$	fb25.bt06	0.71	Catch/MSY	fb25.bt06	0.98	$F/F_{msy}$	fb25.bt06	1.369	$P3(B < B_{lim})$	fb25.bt06	0.1480
AAV	fb25.bt08	0.16	$B/B_{MSY}$	fb25.bt08	0.74	Catch/MSY	fb25.bt08	0.98	$F/F_{msy}$	fb25.bt08	1.306	$P3(B < B_{lim})$	fb25.bt08	0.0920
AAV	fb25.bt1	0.16	$B/B_{MSY}$	fb25.bt1	0.79	Catch/MSY	fb25.bt1	0.99	$F/F_{msy}$	fb25.bt1	1.219	$P3(B < B_{lim})$	fb25.bt1	0.0400
AAV	fb30.bt06	0.15	$B/B_{MSY}$	fb30.bt06	0.86	Catch/MSY	fb30.bt06	1.01	$F/F_{msy}$	fb30.bt06	1.085	$P3(B < B_{lim})$	fb30.bt06	0.0200
AAV	fb30.bt08	0.15	$B/B_{MSY}$	fb30.bt08	0.90	Catch/MSY	fb30.bt08	1.01	$F/F_{msy}$	fb30.bt08	1.040	$P3(B < B_{lim})$	fb30.bt08	0.0120
AAV	fb30.bt1	0.16	$B/B_{MSY}$	fb30.bt1	0.95	Catch/MSY	fb30.bt1	1.01	$F/F_{msy}$	fb30.bt1	0.963	$P3(B < B_{lim})$	fb30.bt1	0.0040
AAV	fb35.bt06	0.15	$B/B_{MSY}$	fb35.bt06	1.03	Catch/MSY	fb35.bt06	1.01	$F/F_{msy}$	fb35.bt06	0.876	$P3(B < B_{lim})$	fb35.bt06	0.0040
AAV	fb35.bt08	0.16	$B/B_{MSY}$	fb35.bt08	1.06	Catch/MSY	fb35.bt08	1.01	$F/F_{msy}$	fb35.bt08	0.848	$P3(B < B_{lim})$	fb35.bt08	0.0000
AAV	fb35.bt1	0.16	$B/B_{MSY}$	fb35.bt1	1.11	Catch/MSY	fb35.bt1	1.00	$F/F_{msy}$	fb35.bt1	0.792	$P3(B < B_{lim})$	fb35.bt1	0.0000
AAV	fb40.bt06	0.15	$B/B_{MSY}$	fb40.bt06	1.20	Catch/MSY	fb40.bt06	1.00	$F/F_{msy}$	fb40.bt06	0.715	$P3(B < B_{lim})$	fb40.bt06	0.0000
AAV	fb40.bt08	0.16	$B/B_{MSY}$	fb40.bt08	1.22	Catch/MSY	fb40.bt08	0.99	$F/F_{msy}$	fb40.bt08	0.696	$P3(B < B_{lim})$	fb40.bt08	0.0000
AAV	fb40.bt09	0.15	$B/B_{MSY}$	fb40.bt09	1.24	Catch/MSY	fb40.bt09	0.98	$F/F_{msy}$	fb40.bt09	0.672	$P3(B < B_{lim})$	fb40.bt09	0.0000
AAV	fb40.bt1	0.16	$B/B_{MSY}$	fb40.bt1	1.28	Catch/MSY	fb40.bt1	0.98	$F/F_{msy}$	fb40.bt1	0.646	$P3(B < B_{lim})$	fb40.bt1	0.0000
AAV	fb45.bt06	0.15	$B/B_{MSY}$	fb45.bt06	1.38	Catch/MSY	fb45.bt06	0.97	$F/F_{msy}$	fb45.bt06	0.588	$P3(B < B_{lim})$	fb45.bt06	0.0000
AAV	fb45.bt08	0.15	$B/B_{MSY}$	fb45.bt08	1.40	Catch/MSY	fb45.bt08	0.96	$F/F_{msy}$	fb45.bt08	0.570	$P3(B < B_{lim})$	fb45.bt08	0.0000
AAV	fb45.bt09	0.15	$B/B_{MSY}$	fb45.bt09	1.43	Catch/MSY	fb45.bt09	0.96	$F/F_{msy}$	fb45.bt09	0.554	$P3(B < B_{lim})$	fb45.bt09	0.0000
AAV	fb45.bt1	0.15	$B/B_{MSY}$	fb45.bt1	1.46	Catch/MSY	fb45.bt1	0.95	$F/F_{msy}$	fb45.bt1	0.533	$P3(B < B_{lim})$	fb45.bt1	0.0000

(b)

Variable	HCR	Median	Variable	HCR	Median	Variable	HCR	Median	Variable	HCR	Median	Variable	HCR	Median
AAV	true.Fmsy	0.15	B/B <sub>MSY</sub>	true.Fmsy	0.94	Catch/MSY	true.Fmsy	1.00	F/F <sub>msy</sub>	true.Fmsy	1.00	P3(B<Blim)	true.Fmsy	0.020
AAV	fmsy.bt06	0.15	B/B <sub>MSY</sub>	fmsy.bt06	0.95	Catch/MSY	fmsy.bt06	1.00	F/F <sub>msy</sub>	fmsy.bt06	1.00	P3(B<Blim)	fmsy.bt06	0.020
AAV	fmsy.bt08	0.15	B/B <sub>MSY</sub>	fmsy.bt08	0.97	Catch/MSY	fmsy.bt08	1.01	F/F <sub>msy</sub>	fmsy.bt08	0.98	P3(B<Blim)	fmsy.bt08	0.016
AAV	fmsy.bt1	0.16	B/B <sub>MSY</sub>	fmsy.bt1	1.00	Catch/MSY	fmsy.bt1	1.01	F/F <sub>msy</sub>	fmsy.bt1	0.94	P3(B<Blim)	fmsy.bt1	0.004
AAV	fb25.bt06	0.15	B/B <sub>MSY</sub>	fb25.bt06	0.83	Catch/MSY	fb25.bt06	0.99	F/F <sub>msy</sub>	fb25.bt06	1.16	P3(B<Blim)	fb25.bt06	0.060
AAV	fb25.bt08	0.16	B/B <sub>MSY</sub>	fb25.bt08	0.84	Catch/MSY	fb25.bt08	0.99	F/F <sub>msy</sub>	fb25.bt08	1.14	P3(B<Blim)	fb25.bt08	0.032
AAV	fb25.bt1	0.16	B/B <sub>MSY</sub>	fb25.bt1	0.88	Catch/MSY	fb25.bt1	1.00	F/F <sub>msy</sub>	fb25.bt1	1.08	P3(B<Blim)	fb25.bt1	0.024
AAV	fb30.bt06	0.15	B/B <sub>MSY</sub>	fb30.bt06	1.00	Catch/MSY	fb30.bt06	1.00	F/F <sub>msy</sub>	fb30.bt06	0.92	P3(B<Blim)	fb30.bt06	0.008
AAV	fb30.bt08	0.15	B/B <sub>MSY</sub>	fb30.bt08	1.01	Catch/MSY	fb30.bt08	1.00	F/F <sub>msy</sub>	fb30.bt08	0.90	P3(B<Blim)	fb30.bt08	0.004
AAV	fb30.bt1	0.16	B/B <sub>MSY</sub>	fb30.bt1	1.05	Catch/MSY	fb30.bt1	1.01	F/F <sub>msy</sub>	fb30.bt1	0.86	P3(B<Blim)	fb30.bt1	0.004
AAV	fb35.bt06	0.15	B/B <sub>MSY</sub>	fb35.bt06	1.17	Catch/MSY	fb35.bt06	0.99	F/F <sub>msy</sub>	fb35.bt06	0.74	P3(B<Blim)	fb35.bt06	0.004
AAV	fb35.bt08	0.15	B/B <sub>MSY</sub>	fb35.bt08	1.19	Catch/MSY	fb35.bt08	0.99	F/F <sub>msy</sub>	fb35.bt08	0.73	P3(B<Blim)	fb35.bt08	0.000
AAV	fb35.bt1	0.15	B/B <sub>MSY</sub>	fb35.bt1	1.22	Catch/MSY	fb35.bt1	0.99	F/F <sub>msy</sub>	fb35.bt1	0.70	P3(B<Blim)	fb35.bt1	0.000
AAV	fb40.bt06	0.15	B/B <sub>MSY</sub>	fb40.bt06	1.35	Catch/MSY	fb40.bt06	0.96	F/F <sub>msy</sub>	fb40.bt06	0.61	P3(B<Blim)	fb40.bt06	0.000
AAV	fb40.bt08	0.15	B/B <sub>MSY</sub>	fb40.bt08	1.36	Catch/MSY	fb40.bt08	0.97	F/F <sub>msy</sub>	fb40.bt08	0.60	P3(B<Blim)	fb40.bt08	0.000
AAV	fb40.bt09	0.15	B/B <sub>MSY</sub>	fb40.bt09	1.38	Catch/MSY	fb40.bt09	0.96	F/F <sub>msy</sub>	fb40.bt09	0.59	P3(B<Blim)	fb40.bt09	0.000
AAV	fb40.bt1	0.15	B/B <sub>MSY</sub>	fb40.bt1	1.40	Catch/MSY	fb40.bt1	0.95	F/F <sub>msy</sub>	fb40.bt1	0.57	P3(B<Blim)	fb40.bt1	0.000

For completeness, we run a bisect analysis for model Run1 to estimate  $F_{P.05}$ , which is equivalent to EQsim. The results (Figure 2.32) show the trajectories of SSB, R and catches with 1000 simulations. The associated probability of falling below  $B_{lim}$  when fishing at  $F_{B35\%}$  was 0.04% (Table 2.8.).  $F_{P.05}$  for Run 1 is estimated to be 0.345, which is larger than  $F_{B35\%} = 0.204$ . Therefore,  $F_{B35\%}$  with  $B_{trigger}$  set at 0.6 of  $B_{35\%}$  is a suitable set of candidate reference points for the Central Baltic herring stock. Catches 5% greater than MSY are not precautionary, thus limiting  $F_{B35\%upper}$  to  $F_{B35\%}$ .  $F_{B35\%lower}$ , i.e. when  $Catch/MSY = 0.95$ , is equal to  $0.70 * F_{B35\%}$  as estimated by the MSE. The results of the MSE with the implementation error set to 0 with standard deviation equal to 0.149, are included in the report for comparative reasons (Table 2.8 a). Without implementation error, a  $F_{B30\%}$  with  $B_{trigger}$  set at  $B_{30\%}$  would be considered a suitable set of candidate reference points for the Central Baltic herring stock (Table 2.8 b). As the benchmark decided to set the reference points from the results of the MSE with implementation error however, the results from the MSE without implementation error will not be discussed further.

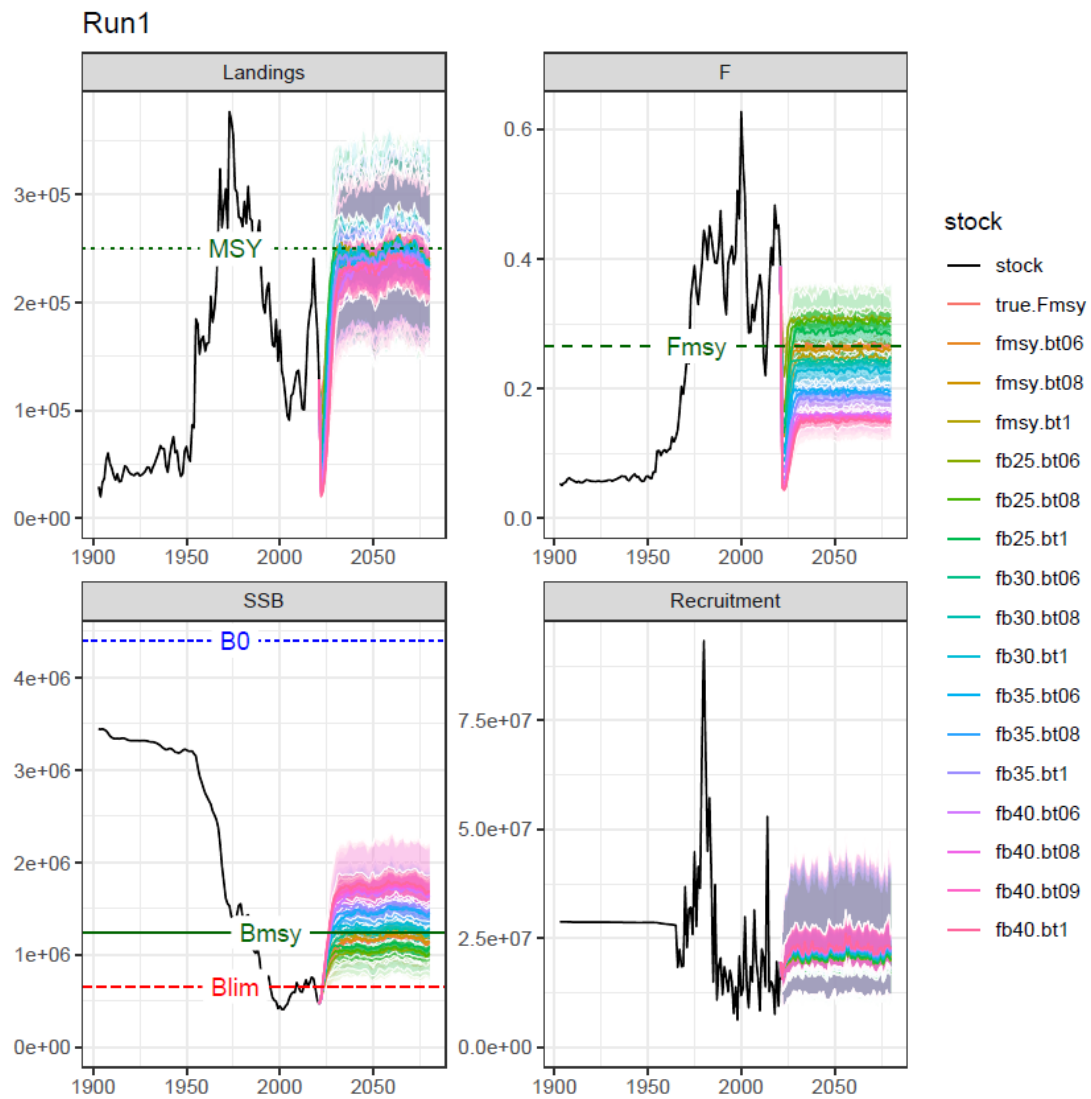
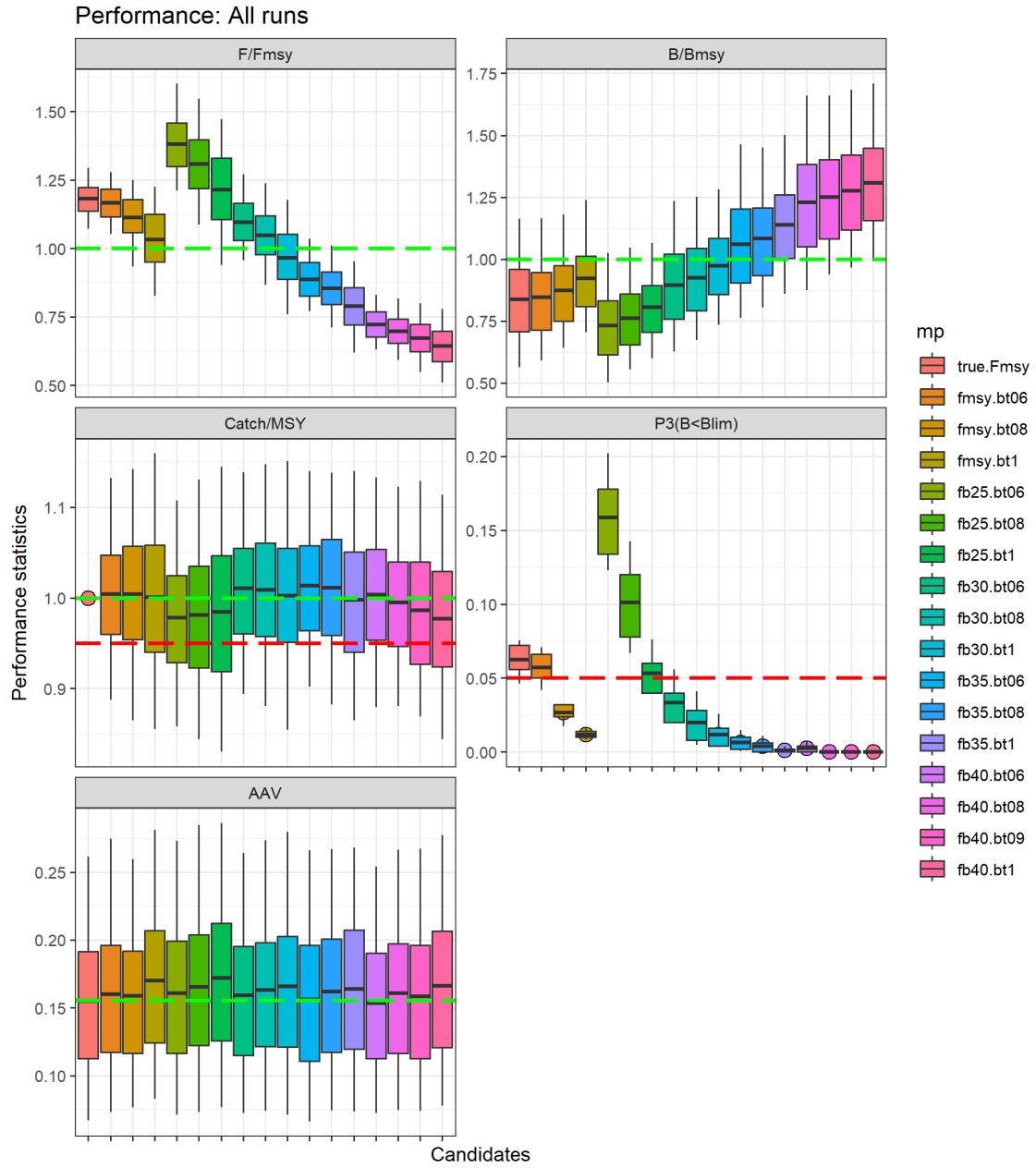


Figure 2.30. Central Baltic herring (*Clupea harengus*; CBH; her.27.25-2932) in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Long term simulations for Run1 with average implementation error set to 0 with standard deviation equal to 0.119. Trends in SSB, F, landings, and R for different combinations of  $F_{target}$  and  $B_{trigger}$  and compared to the deterministic  $F_{MSY}$ .  $B_{lim}$  is set as 15% of  $B_0$ .

(a)



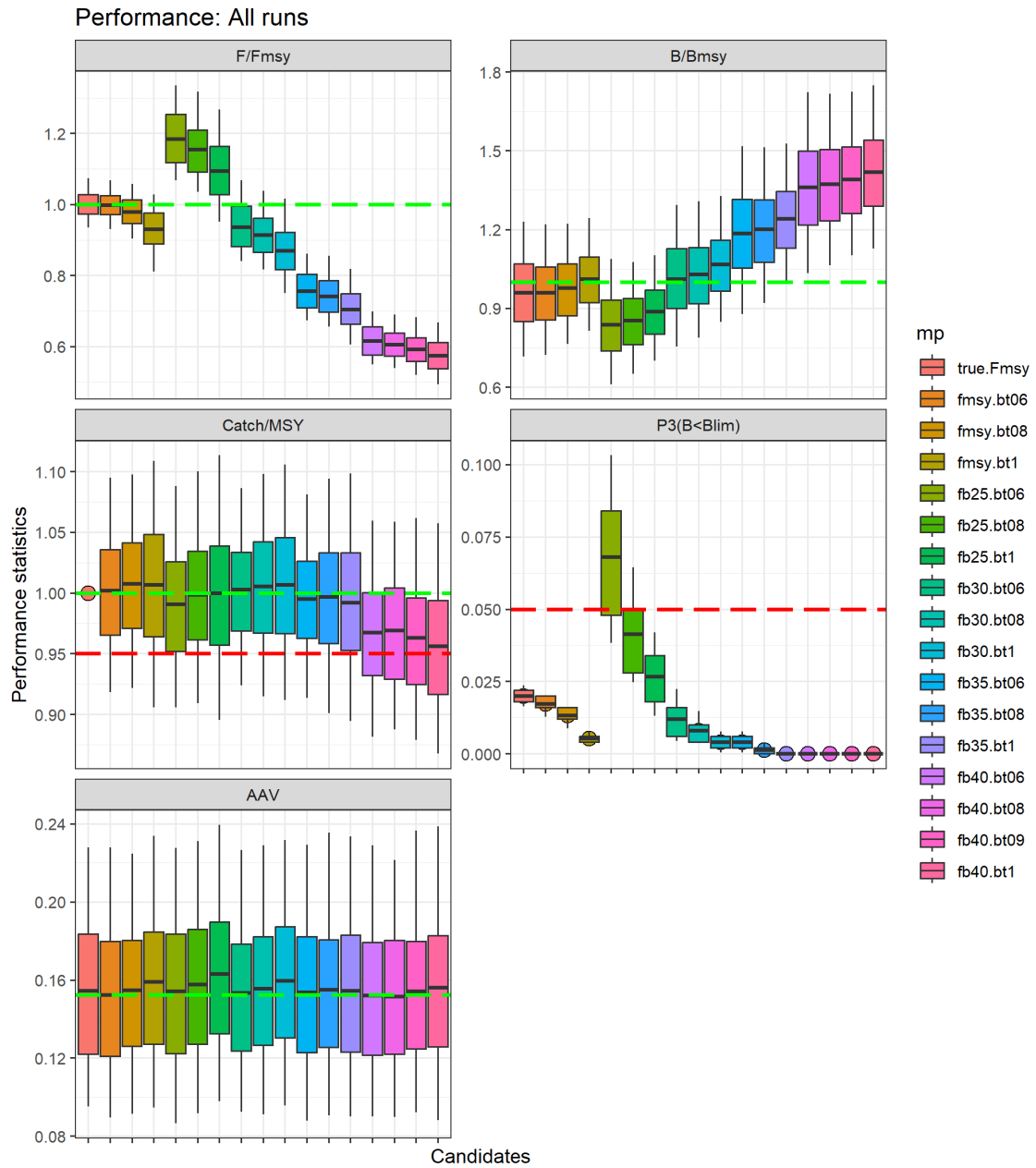


Figure 2.31. Central Baltic herring (*Clupea harengus*; CBH; her.27.25-2932) in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Results of MSE used of evaluate reference point systems, showing the type 3 risk probabilities (P3) of SSB falling below  $B_{lim}$ , the median long-term yield relative the median long term obtained at the true  $F_{MSY}$  (MSY), the median long-term F and SSB relative to the true  $F_{MSY}$  and  $B_{MSY}$  and the median long term interannual variation in catches. Green and red dashed lines denoting the target and limit thresholds, respectively. Candidates based on  $F_{B\%}$  and  $B_{trigger}$  as fraction of  $B_{\%}$ . Plots refers to MSE with implementation error set to 0.165 with standard deviation equal to 0.149 (a) and implementation error set to 0 with standard deviation equal to 0.149 (b)

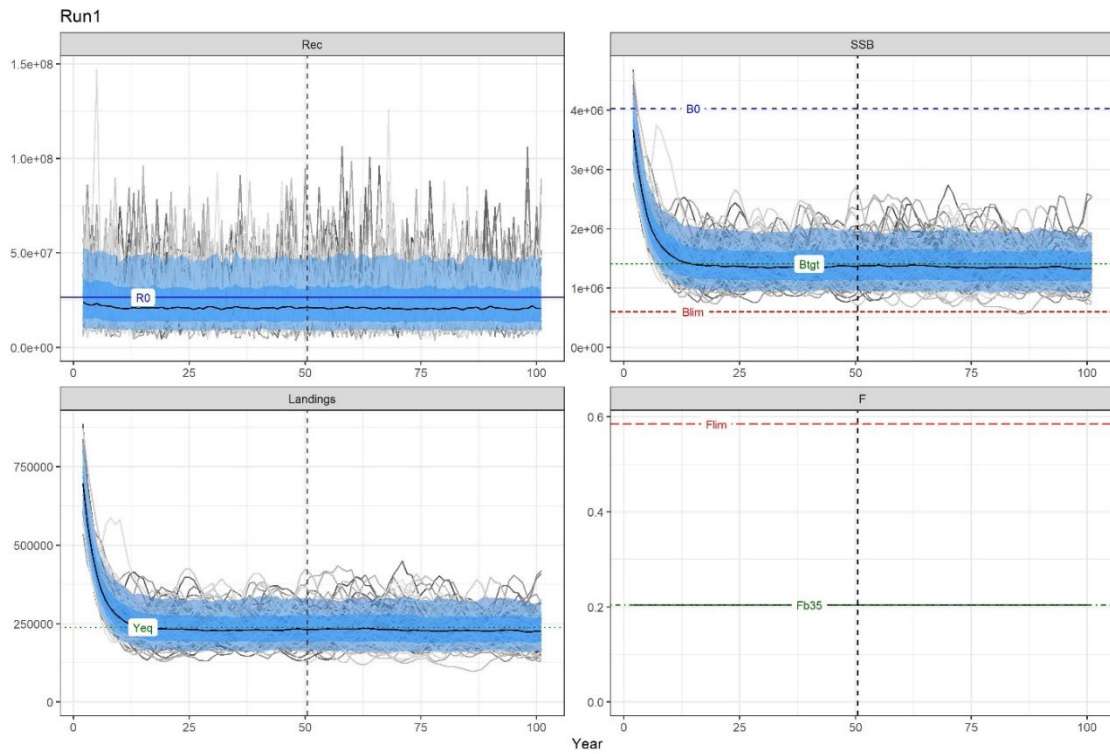


Figure 2.32. Central Baltic herring (*Clupea harengus*; CBH; her.27.25-2932) in subdivisions (SDs) 25-29, 32 excluding the Gulf of Riga (Central Baltic Sea). Results of the bisect analysis for model Run1 to estimate  $F_{p,0.5}$ , which is equivalent to EQsim.

## 2.7 Roadmap for the future work on the stock assessment model for Central Baltic Herring.

Goal: To develop a spatially structured length- and age-based stock assessment model for Central Baltic Herring.

This roadmap was drafted during the BWKBALPEL meeting and needs to be approved also by those countries not attending this meeting before it can be settled. The roadmap will be anchored with these countries at the first in a series of meetings with data providers outlined below.

Plan for 2023 – 20XX (year to be decided)

1. Identify available data
  - a. Fishery-dependent data per country, year, SD, fleet segment
    - i. Catch, including age 0
    - ii. Lengths
    - iii. ALKs
    - iv. Maturities
    - v. WAAs
  - b. Survey data per SD
    - i. Index of abundance
    - ii. Lengths
    - iii. ALKs
  - c. Decide on age+ group
  - d. Explore age groups for  $F_{bar}$
  - e. Tagging data
  - f. Species misreporting
  - g. Potential productivity changes due to regime shifts. How SR relationship would be affected.

*How: Set up a series of online meetings between assessors and data providers to decide on a time plan discuss data types, data availability, preparation and data uploading of landings and biological data to RDBES (or other data format), preparation of national scripts capable of estimating biological parameters (length, age, weight, maturity) from RDBES data. The progress on catch data corrected for species misreporting should also be discussed as well as aspects that may impact data quality such as sampling design and data storage practiced (see e.g. WD3 in Annex 5). During the first meeting core groups of people for further meetings, specific analyses and communication should be identified.*

*The assessment WG should further communicate with the RCG on future data needs.*

*Who: The stock assessor and coordinator of central Baltic herring will take the responsibility to set up these meetings (Mikaela Bergenius Nord, Szymon Smoliński). The meetings will be attended by the core group identified during the first meeting.*

2. Preparatory work individual countries

*How: Each country needs to anchor data needs and time plan decided during the online meetings with home institutes and plan for delivery of data in time for the data call.*

*Who: Responsible people will be appointed by each country.*

3. Preparatory work WGBIFS and others

h. Modeling of abundance indices

Abundance indices from scientific surveys are key stock assessment inputs, but when the availability of fish varies in space and time (Monnahan *et al.*, 2021), or biological samples and acoustic information arise from spatially unbalanced sampling (Thorson *et al.*, 2020), the estimated indices and associated uncertainties do not accurately reflect changes in population abundance (Monnahan *et al.*, 2021). These issues complicate also the creation of spatial maps in unsampled areas. For these reasons, spatio-temporal modeling, such as the vector autoregressive state space modeling platform (VAST; (Thorson and Barnett, 2017; Thorson, 2019), can be applied to “correct” the indices for the effects of the viable horizontal and vertical distribution of fish, possible environmental effects, and sampling coverage of acoustic surveys.

*How: WKBPELA recommends a workshop on the improved standardization of survey data during 2023/2024?*

*Who: The stock assessor and coordinator of central Baltic herring will take the responsibility that this will be communicated with the chair of EOSG*

i. Prepare index, numbers and lengths and ALKs for each SD

*How: The assessment WG should communicate with WGBIFS on data needs and time plan.*

*Who: The stock assessor and coordinator of central Baltic herring will take the responsibility that this will be communicated with the chair of the WGBIFS (Mikaela Bergenius Nord, Szymon Smoliński)*

4. Data call

j. Specifics of the data needed



- k. Upload to RDBES?
- l. Documentation of available data, sampling design, methodologies for whole filling and the development of ALKs...

*How: decide on the specifics of the data call before the send out together with the data providers during the online meetings (issue 1 above).*

*Who: ICES secretariat in consultation with core data group identified during the first meeting*

#### 5. Mixing of stocks and substocks

Analyses presented at the previous meetings suggest that a better understanding of the central Baltic herring stock structure is needed (ICES, 2018). The central Baltic herring is known to be dominated by a northern and a southern component but the latter shares numerous characteristics with the adjacent western Baltic herring stock. Its growth and otolith shape are more similar to those of herring of western origin than to fish from the northern component. Based on only growth, a high proportion of fast-growing herring is found in SD25 and especially in the westernmost rectangles but it remains unclear if those fish are part of the southern component of the central Baltic or if they are the results of extensive mixing with the western Baltic herring. Preliminary analyses suggest a progressive genetic differentiation along the entire southern Baltic coasts from SD24 to SD26 rather than a clear-cut division between different assessment units.

There is a need for a separate research project with the aim to clarify the stock structure of the central Baltic herring, validate herring assessment units, and look for operational methods to separate different components in mixed catches. A general concept of the project and sampling design were presented in ICES (2018).

*How: Attend WKSIDAC2 (2023) to discuss the central Baltic herring case and if needed recommend a workshop for 2024 on the stock structure of central Baltic herring using different stock identification techniques.*

*Who: SLU Aqua, Sweden and NMFRI, Poland.*

#### 6. Accuracy in age estimates

After developing a reference radiocarbon ( $^{14}\text{C}$ ) database, it will be possible to use methods based on the analysis of  $^{14}\text{C}$  carbon isotopes in the otoliths for validating the age readings of fish species in the Baltic Sea. This method has been continuously improved over the last 30 years (Andrews *et al.*, 2019; Lackmann *et al.*, 2019). These technological improvements, together with new knowledge about the propagation of the  $^{14}\text{C}$  signal associated with nuclear tests in the 1950s and

1960s in marine ecosystems, allow validation of otolith-based fish age readings (Andrews *et al.*, 2020). Once the reference system and validation of the expert readings are established, it is possible to estimate the accuracy of the readings (as opposed to the precision measured during the 2022 otolith exchanges) and possibly improve the reading methods. The age error matrices can be further integrated into the SS3 models.

*How: Await the results of an ongoing Swedish project on using radiocarbon to determine the age of herring in the Baltic 2023/2024.*

*Who: Francesca Vitale*

#### 7. Data compilation workshop – Dates to be decided

Final compilation of commercial data at international level. Decisions on imputations. Decisions on survey data, etc. Some stock level analysis ( e.g. maturity) may need to be repeated taking into account the newly available data in RDBES format and the information that has since been collected on sampling methodologies employed by the different countries over the years.

#### 8. Benchmark

**Table 2.9. Summary of steps to take/issues to solve as part of the road map and proposed timelines. The timelines need to be decided in consultation with the data providers and are thus indicative only.**

	2023	2024	2025?
1. Identify available data			
2. Preparatory work individual countries			
3. Preparatory work WGBIFS and others			
4. Data call			
5. Mixing of stocks and substocks			
6. Accuracy in age estimates			
7. Data compilation			
8. Benchmark			

## 2.8 Progress Feco

The progress on the work to calculate Feco for the stocks benchmarked was briefly presented during the benchmark. The work has been partly postponed, but will continue during spring 2023.

## 2.9 References

- Aro, E. 1989. A review of fish migration patterns in the Baltic. *Rapports et Procès-Verbaux des Réunions du Conseil International pour l'Exploration de la Mer* 190: 72-96. Aro, E. 2002. Fish migration studies in the Baltic Sea—a historical review. In *ICES Marine Science Symposia* 215, 361-370.
- Anderson S.C., Cooper A.B., Jensen O.P., Minto C., Thorson J.T., Walsh J.C., Afflerbach J., Dickey-Collas M., Kleisner K. M., Longo C., Osio G.C., Ovando D., Mosqueira I., Rosenberg A.A., Selig E.R, 2017. Improving estimates of population status and trend with superensemble models. *Fish Fisheries*,18: 732–741. Doi: <https://doi.org/10.1111/faf.12200>.
- Bekkevold, D., Berg, F., Polte, P., Bartolino, V., Ojaveer, H., Mosegaard, H., Farrell, E. D., 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: 173–184.
- Bessell-Brown, P., Punta, A. E., Tucka, G. N., Day J., Klaer, N., & Penney, A., 2022. The effects of implementing a 'dynamic B0' harvest control rule in Australia's Southern and Eastern Scalefish and Shark Fishery. *Fisheries Research* 252. <https://doi.org/10.1016/j.fishres.2022.106306>.
- Carvalho, F., Punt, A.E., Chang, Y.-J., Maunder, M.N., Piner, K.R., 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? *Fish. Res.* 192, 28–40. <https://doi.org/10.1016/j.fishres.2016.09.018>.
- Carvalho, F., Winker, H., Courtney, D., Kell, L., Kapur, M., Cardinale, M., Schirripa, M., Kitakado, T., Ghebrehiwet, D.Y., Piner, K.R., Maunder, M.N., Methot, R., 2021. A Cookbook for Using Model Diagnostics in Integrated Stock Assessments. *Fisheries Research*, <https://doi.org/10.1016/j.fishres.2021.105959>.
- Dietterich, T.G. 2000. Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1–15). Berlin, Heidelberg: Springer.
- Eero, M., Köster, F.K., MacKenzie, B.R., 2008. Reconstructing historical stock development of Atlantic cod (*Gadus morhua*) in the eastern Baltic Sea before the beginning of intensive exploitation. *Can. J. Fish. Aquat. Sci.* 65: 2728–2741 (2008).
- FAO-GFCM. 2021. Report of the Working Group on Stock Assessment of Demersal Species (WGSAD) – Benchmark session for the assessment of common sole in GSA 17, Scientific Advisory Committee on Fisheries (SAC). Online via Microsoft Teams, 12–16 April 2021.
- Francis, R.C. and Hilborn, R. 2011. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences* 68(6): 1124–1138. doi:10.1139/f2011-025.
- Fromentin, J.-M., Bonhommeau, S., Arrizabalaga, H., & L. Kell, L.T. 2014. The spectre of uncertainty in management of exploited fish stocks: the illustrative case of Atlantic Bluefin tuna. *Marine Policy*, 47, 8–14.
- Gibbons, J.D. & Chakraborti, S. (1992). *Nonparametric Statistical Inference*. New York: Marcel Dekker.
- Gunderson, D.R. 1993. *Surveys of fisheries resources*. Wiley. New York. 248 pp.
- Grohslar T, Oeberst R, Schaber M, Larson N, Kornilovs G (2013). Discrimination of western Baltic spring-spawning and central Baltic herring (*Clupea harengus* L.) based on growth vs. natural tag information. *Ices Journal of Marine Science*. 2013;70(6):1108-17.
- Han F, Jamsandekar M, Pettersson ME, Su L, Fuentes-Pardo A, Davis B et al (2020). Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. *eLife*. DOI: 10.7554/eLife.61076
- Hilborn, R. (2016). Correlation and causation in fisheries and watershed management. *Fisheries*, 41(1), 18–25.
- Hilborn, R., and Walters, C. J. 1992. *Quantitative Fish Stock Assessment. Choice, Dynamics and Uncertainty*. New York: Chapman and Hall, 570.
- Hyndman, R.J. and Koehler, A.B., 2006. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), pp.679-688.

- Hyndman, R.J. and Athanasopoulos, G. 2013. Forecasting: principles and practice, an online text book. Retrieved September 16, 2012, from <http://otexts.com/fpp/>.
- Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *Ices J. Mar. Sci.* 72, 99–110. <https://doi.org/10.1093/icesjms/fsu198>.
- ICES. 2002. Report of the study group of herring assessment units of the Baltic.
- ICES. 2013. Report of the Benchmark Workshop on Pelagic Stocks (WKPELA), 4-8, February 2013, Copenhagen, Denmark.
- ICES. 2018. Report of the workshop on mixing of western and central Baltic herring stocks (WKMixHER). 11-13 September, Gdynia, Poland. 43 pp.
- ICES. 2020a. Inter-Benchmark Process on Baltic Sprat (*Sprattus sprattus*) and Herring (*Clupea harengus*) (IBPBash). ICES Scientific Reports, 2:34. 44 pp. <https://doi.org/10.17895/ices.pub.5971>.
- ICES. 2020b. The third Workshop on Guidelines for Management Strategy Evaluations (WKG MSE3). ICES Scientific Reports. 2:116. 112 pp. <http://doi.org/10.17895/ices.pub.7627>.
- ICES. 2021a. Benchmark Workshop on the development of MSY advice for category 3 stocks using Surplus Production Model in Continuous Time; SPiCT (WKMSYSPiCT). ICES Scientific Reports, 3: 1–317.
- ICES. 2021b. Joint NAFO\ICES Pandalus Assessment Working Group (NIPAG). ICES Scientific Reports. 3:22. 25 pp. <https://doi.org/10.17895/ices.pub.5990>.
- ICES. 2022. Baltic Fisheries Assessment Working Group (WGBFAS). ICES Scientific Reports. Report. <https://doi.org/10.17895/ices.pub.19793014.v2>
- ICES. 2022a. Benchmark workshop on Pandalus stocks (WKPRAWN). ICES Scientific Reports. 4:20. 249 pp. <http://doi.org/10.17895/ices.pub.19714204>
- ICES. 2022b. Herring (*Clupea harengus*) in subdivisions 25–29 and 32, excluding the Gulf of Riga (central Baltic Sea). In Report of the ICES Advisory Committee, 2022. ICES Advice 2021, her.27.25–2932. <https://doi.org/10.17895/ices.advice.19447970>.
- ICES. 2022c. Workshop on ICES reference points (WKREF1). ICES Scientific Reports. 4:2. 70 pp. <http://doi.org/10.17895/ices.pub.9749>.
- ICES. 2023. Working Group on Baltic International Fish Survey (WGBIFS; outputs from 2022 meeting). ICES Scientific Reports. 05:10. 529pp. <https://doi.org/10.17895/ices.pub.22068821>
- ICES. 2022a. Benchmark workshop on Pandalus stocks (WKPRAWN). ICES Scientific Reports. 4:20. 249 pp. <http://doi.org/10.17895/ices.pub.19714204>.
- ICES. 2022b. Baltic Fisheries Assessment Working Group (WGBFAS). ICES Scientific Reports. 4:44. 659 pp. <http://doi.org/10.17895/ices.pub.19793014>.
- ISSF 2011. Report of the 2011 ISSF Stock Assessment Workshop Rome, Italy, March14-17, 2011. ISSF Technical Report 2011-02 International Seafood Sustainability Foundation, Washington, D.C., USA (2011).
- Johnson, K.F., Councill, E., Thorson, J.T., Brooks, E., Methot, R.D., Punt, A.E., 2016. Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts? *Fish. Res.* 183, 222–232.
- Kaljuste, O., Blass, M. och Söderberg, K. (2009). Research report from the Bothnian Sea spring survey (Vår-survey, 16–25.03.2009), Fiskeriverkets Kustlaboratorium, 19 pp.
- Kell, L.T., Kimoto, A., Kitakado, T., 2016. Evaluation of the prediction skill of stock assessment using hindcasting. *Fish. Res.* 183, 119–127. <https://doi.org/10.1016/j.fishres.2016.05.017>.
- Kell, K., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., Fu, D., 2021. Validation of stock assessment methods: is it me or my model talking? *ICES Journal of Marine Science*, <https://doi.org/10.1093/icesjms/fsab104>.

- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G.A. 2009. Challenges in combining projections from multiple climate models. *Journal of Climate*, 23, 2739–2758.
- Kolody, D. S., Eveson, J. P., Preece, A. L., Davies, C. R., & Hillary, R. M. (2019). Recruitment in tuna RFMO stock assessment and management: A review of current approaches and challenges. *Fisheries Research*, 217, 217-234.
- Mannini A, Pinto C, Konrad C, Vasilakopoulos P and Winker H. 2020. “The Elephant in the Room”: Exploring Natural Mortality Uncertainty in Statistical Catch at Age Models. *Front. Mar. Sci.* 7:585654. doi: 10.3389/fmars.2020.585654.
- Maunder, M.N., Piner, K.R. 2015. Contemporary fisheries stock assessment: many issues still remain ICES J. Mar. Maunder, M.N., Xu, H., Lennert-Cody, C.E., Valero, J.L., Aires-da-Silva, A., Mente Vera, C., 2020. Implementing Reference Point-based Fishery Harvest Control Rules Within a Probabilistic Framework That Considers Multiple Hypotheses (No. SAC-11- INF-F). Scientific Advisory Committee, Inter-American Tropical Tuna Commission, San Diego. *Sci.*, 72 (1) (2015), pp. 7-18.
- Merino, G., Urtizberea, A., Fu, D., Winker, H., Cardinale, M., Lauretta, M., Murua, H., Kitakado, T., Arrizabalaga, H., Scott, R., Pilling, G., Laborda, A., Erauskin-Extramiana, M., Santiago, J., 2022. Investigating trends in process error as a diagnostic for integrated fisheries' stock assessments. *Fisheries Research*, <https://doi.org/10.1016/j.fishres.2022.106478>.
- Methot, R.D., Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research* 142 (2013) 86–99.
- Methot, R.D., Wetzel, C.R., Taylor, I.G., Doering, K.L., and Johnson, K.F. 2022. Stock Synthesis User Manual Version 3.30.20. NOAA Fisheries Seattle, WA, September 22, 2022.
- Minte-Vera, C.V., Maunder, M.N., Aires-da-Silva, A.M., Satoh, K., Uosaki, K., 2017. Get the biology right, or use size-composition data at your own risk. *Fish. Res.* 192, 114–125. <https://doi.org/10.1016/j.fishres.2017.01.014>.
- Mildenberger, T. K., Berg, C. W., Kokkalis, A., Hordyk, A. R., Wetzel, C., Jacobsen, N. S., Punt, A. E., et al. 2021. Implementing the precautionary approach into fisheries management: Biomass reference points and uncertainty buffers. *Fish and Fisheries*: 1–20.
- Monnahan, C.C., Branch, T. A., Thorson, J. T., Stewart, I. J., Szuwalski, C. S., 2019. Overcoming long Bayesian run times in integrated fisheries stock assessments. *ICES Journal of Marine Science*, fsz059, <https://doi.org/10.1093/icesjms/fsz059>.
- Myers, R. A., Bowen, K. G., & Barrowman, N. J. (1999). Maximum reproductive rate of fish at low population sizes. *Canadian Journal of Fisheries and Aquatic Sciences*, 56(12), 2404-2419.
- Nilsson L, Thygesen U, Lundgren B, Nielsen B, Nielsen J, Beyer J (2003) Vertical migration and dispersion of sprat (*Sprattus sprattus*) and herring (*Clupea harengus*) schools at dusk in the Baltic Sea. *Aquatic Living Resources* 16:317–324
- Ojaveer, E. 1981. Marine pelagic fishes. In *The Baltic Sea*. Elsevier Oceanography Series No. 30, pp. 276-292. Ed. By A. Voipio. Elsevier Scientific Publishing Company, Amsterdam-Oxford-New York.
- Ojaveer E. Population-Structure of Pelagic Fishes in the Baltic. *Baltic Sea Fishery Resources* (1989). 190:17-21.
- Ojaveer, E. (2003). Baltic herring, *Clupea harengus membras* L. In: *Fishes of Estonia*, pp. 58-79. Editors Ojaveer, E., Pihu, E., Saat, T. Estonian Academy Pub.
- Pechlivanidis, IG, Gupta, H, & Bosshard, Thomas. 2018. An Information Theory Approach to Identifying a Representative Subset of Hydro-Climatic Simulations for Impact Modeling Studies. *Water resources research*, 54(8), 5422–5435.
- Perälä T, Kuparinen A. 2017. Detection of Allee effects in marine fishes: analytical biases generated by data availability and model selection. doi/full/10.1098/rspb.2017.1284
- Perälä T, Kuparinen A., Jeffrey Huthchings 2021. Allee Effects and the Allee-effect Zone in Atlantic Cod. *Biology Letter*, *Biol. Lett.* 18: 20210439, <https://doi.org/10.1098/rsbl.2021.0439>.in press

- Popiel J. Differentiation of the biological groups of herring in the Southern Baltic (1958). *Rapp P-v Reun Cons int Explor Mer.* 143(2):121.
- Polte, P., Kotterba, P., Moll, D., and Von Nordheim, L. (2017). Ontogenetic loops in habitat use highlight the importance of littoral habitats for early life-stages of oceanic fishes in temperate waters. *Sci. Rep.* 7:42709. doi: 10.1038/srep42709
- Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A., Haddon, M., 2014. Management strategy evaluation: best practices. *Fish and Fisheries*, <https://doi.org/10.1111/faf.12104>
- Raid T, Järv, Pönni J, Raitaniemi J, Kornilovs G. Central Baltic herring stock: What does the assessment of combined stock say about the status of its components? (2016). In: Guedes Soares C, Santos TA, editors. *Maritime Technology and Engineering*. London: Taylor & Francis Group; p. 961-6.
- Sampson, D.B., & Scott, R.D. (2012). An exploration of the shapes and stability of population–selection curves. *Fish and Fisheries*, 13, 89–104.
- Siekmann, I., Sneyd, J., & Crampin, E. J. (2012). MCMC can detect non identifiable models. *Biophysical journal*, 103(11), 2275-2286.
- Subbey, S. (2018). Parameter estimation in stock assessment modelling: caveats with gradient-based algorithms. *ICES Journal of Marine Science*, 75(5), 1553-1559.
- Tredennick, A. T., G. Hooker, S. P. Ellner, and P. B. Adler. 2021. A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology* 102(6), <https://doi.org/e03336>. 10.1002/ecy.3336.
- Trijoulet, V., Albertsen, C.M., Kristensen, K., Legault, C.M., Miller, T.J., Nielsen, A., 2022. Model validation for compositional data in stock assessment models: Calculating residuals with correct properties <https://doi.org/10.1016/j.fishres.2022.106487>.
- Walters, C. J., Hilborn, R., & Christensen, V. (2008). Surplus production dynamics in declining and recovering fish populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(11), 2536-2551.
- Walter, J., Winker, H., 2019. Projections to create Kobe 2 Strategy Matrices using the multivariate log normal approximation for Atlantic yellowfin tuna. *ICCAT-SCRS/2019/145* 1–12.
- Winker, H., Carvalho, F., and Kapur, M., 2018. JABBA: Just Another Bayesian Biomass Assessment. *Fisheries Res* Winker, H., Walter, J., Cardinale, M., Fu, D., 2019. A multivariate lognormal Monte-Carlo approach for estimating structural uncertainty about the stock status and future projections for Indian Ocean Yellowfin tuna. *IOTC-2019-WPM10-XX*.
- Wu, H., and Levinson D., 2021. The ensemble approach to forecasting: A review and synthesis. <https://doi.org/10.1016/j.trc.2021.103357>earch, Volume 204, August 2018, Pages 275-288.

## 2.10 Minority statement on the estimation of $B_{lim}$ for CBH stock

Jan Horbowy (National Marine Fisheries Research Institute, Gdynia, Poland)

Szymon Smoliński (National Marine Fisheries Research Institute, Gdynia, Poland)

Stefanie Haase (Thünen Institute of Baltic Sea Fisheries, Rostock, Germany)

Maria Pierce (Thünen Institute of Baltic Sea Fisheries, Rostock, Germany)

On average, the stock biomass estimated with the stock assessment model SS3 was by ca. 20% higher compared to the estimates calculated with XSA for the time-series since 1990 (Figure 2.33). Meanwhile, the value of  $B_{lim}$  proposed at WKBBALTPEL (561 kt) is 70% higher than the previous one (330 kt; Table 2.10). Such a high change in  $B_{lim}$  compared to the change in assessment should be well justified as it may have a large impact on the management of the stock and jeopardizes the trust in the reliability of ICES recommendations. The minority group did not find convincing

scientific reasons for such a high change in  $B_{lim}$ . They presented an alternative view on the selection of  $B_{lim}$  for the CBH stock, which was, however, not adopted by the WKBBALPEL.

$B_{lim}$  for CBH has been estimated by WKBBALPEL as 15% of  $B_0$ , defined as the stock biomass at equilibrium if the stock is unfisher. The derivation of  $B_{lim}$  as a fraction of  $B_0$  is gaining some popularity, particularly when assessing the stock using stock-production models. However, this approach has some potential issues when applied to CBH:

1. The estimation of  $B_0$  goes well beyond the observed series of the biomass and fishing mortality (series since 1974 which is the first year when detailed data on the stock parameters were available), so the derived  $B_0$  is an extrapolation that may be biased.
2. It is difficult to objectively set the fraction by which  $B_0$  is multiplied to get  $B_{lim}$ ; usually such a fraction is constrained between 0.1-0.2. The used fraction of  $B_0$  considerably affects  $B_{lim}$ , as 20% of  $B_0$  results in 100% higher  $B_{lim}$  than 10% of  $B_0$ .

The sigmoidal Beverton-Holt model was fitted to the new (estimated by SS3) S-R data on CBH at the WKBBALPEL. It was shown that the inflection in the S-R relationships (i.e. Allee effect) is probably observed for that stock. The inflection point (i.e. the upper border of the Allee effect "zone") was estimated at SSB of about 302 kt and its 95<sup>th</sup> percentile at 423 kt.

The 15% of  $B_0$  was selected somewhat arbitrarily, using an argument that the Allee effect may still be present up to a fraction of ~10% of  $B_0$ . Following this argument, the minority group questioned why e.g. 11% of  $B_0$  cannot be used.

According to ICES definitions,  $B_{lim}$  is "a deterministic biomass limit below which a stock is considered to have reduced reproductive capacity", so the minority group was of the opinion that  $B_{lim}$  should be estimated directly from the S-R relationship. The inflection point of the S-R relationship (302 kt) may be a good candidate for  $B_{lim}$  because depensation in the S-R relationship is observed when SSB drops below that point. But since there are no observations of S-R points in the vicinity of the estimate of the inflection point for CBH, the 95<sup>th</sup> percentile of the estimated inflection point (423 kt) could be taken as a proxy of  $B_{lim}$ .

Such an estimate of  $B_{lim}$  has a good biological basis (low probability of SSB below the level at which depensation in the S-R relationship may occur, which meets precautionary objectives of the ICES) and is more in line with a previous estimate of the reference point (28% increase in  $B_{lim}$ ) in light of a new assessment of CBH (20% increase in SSB). This value is also close to an estimated 40% of  $B_{MSY}$  (421 kt) – a proxy of  $B_{lim}$  used in Canada. Thus, the minority group suggests to use a **SSB of ca. 423 kt** as a new estimate of  $B_{lim}$ .

The current state of CBH is not good and recent advice has used the ICES advice rule to set TAC. However, the lack of improvement of the stock status in recent years is not due to a  $B_{lim}$  which was set too low, but due to the fact that the realized fishing mortality was almost twice as high as the advised  $F$  ( $F_{MSY}$  or reduced  $F$  following ICES advice rule).

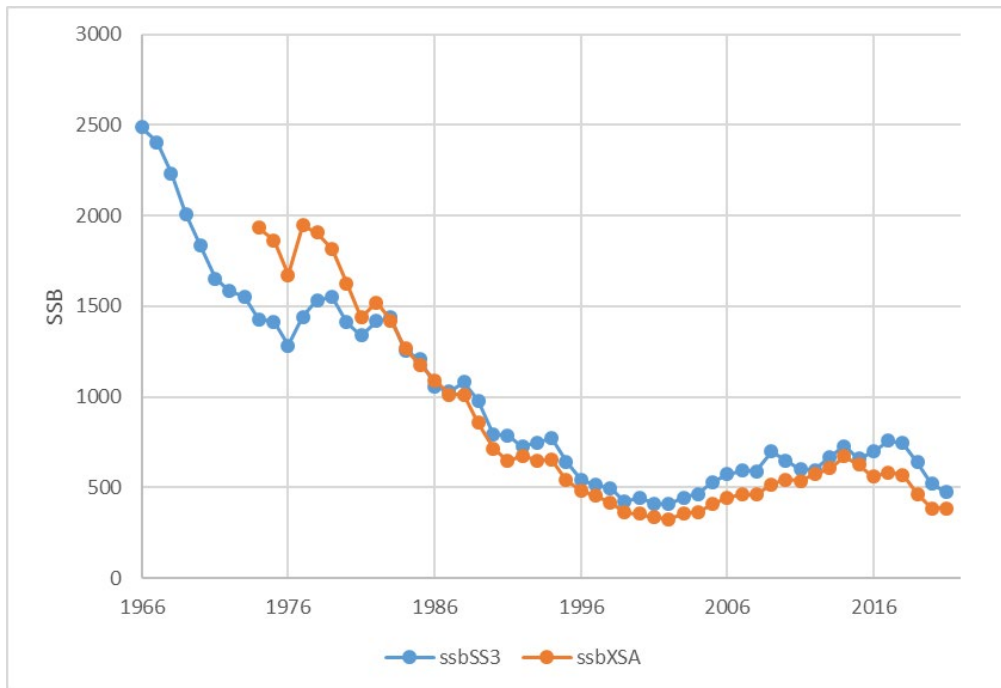


Figure 2.33. The SSB estimated in present assessment (SS3) and former XSA.

Table 2.10. Historical and potential reference points estimated at the WKBBALPEL.  $B_{lim}$  proposed by the minority group was marked with an asterisk (\*),  $B_{lim}$  adopted by the WKBBALPEL was marked with \*\*.

Blim estimate	Technical basis
330	Previous $B_{lim}$ (the lowest SSB that has resulted in above-average recruitment)
302	50 <sup>th</sup> percentile of the S-R inflection point
423	95 <sup>th</sup> percentile of the S-R inflection point*
561	$B_{15\%B_0}$ **
374	$B_{10\%B_0}$
421	$40\%B_{MSY}$



### 3 Sprat (*Sprattus sprattus*) in subdivisions 22–32 (Baltic Sea)

The following working documents supports the summarizing texts in this report.

- WD1\_MultiSpecies\_M for the central Baltic herring her.27.25-2932 and Baltic sprat spr.27.22-32
- WD8\_Working Document catch data of Baltic sprat and central Baltic herring\_all countries
- WD9\_reference points for Baltic Sprat\_22\_32

#### 3.1 Issue list

The issue list compiled for the meeting is detailed below in Table 3.1. An extra column ‘Conclusions and outcomes’ has been added to provide concluding remarks or outcomes for each issue.

**Table 3.1. Sprat (*Sprattus sprattus*) in subdivisions 22–32 issue list for the WKBBALPEL meeting.**

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed /are these available / where should these come from?	Research/WG input needed	Main Person	Conclusions and outcomes
Tuning series	BIAS data. Do we have new BIAS data from SD 32 that could be used in the assessment?	Produce index	Index produced by WGBIFS members	WGBIFS	Olavi Kaljuste (WGBIFS)	An index included SD 32 was provided. This index starts in 2000 as opposed to the old index that started in 1991.  It was decided not to use the index calculated by StoX.
		Investigate performance of new index			<a href="#">Jan Horbovy</a>	New BIAS index was finally decided to be used in parallel with the old BIAS index, while from the old index were used only these years that are lacking in the new one.
Biological Parameters	Mean weight in the stock. Equals currently mean weight in the catch!	Compare WECA with WEST from the BIAS survey (base)	Mean weights at age from commercial and BIAS data		<a href="#">Stefanie Haase</a> , <a href="#">Szymon Smolinski</a>	
	Natural Mortality (base) currently constant over	New key from SMS	WGSAM to produce a new key run	WGSAM	Morten Winter (WGSAM)	

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed /are these available / where should these come from?	Research/WG input needed	Main Person	Conclusions and outcomes
	time and ages (SMS provides predation mortality)	(predation mortality)				
Misreporting of herring and sprat.	Misreporting of herring and sprat in the mixed catches.	Misreporting estimates from relevant countries	Logbooks data, VMS data reports from landing controls	Project by RCG	ISSG	Only Denmark presented new sprat landings based on the corrections because of the historic misreporting of herring and sprat. Historic sprat catches were corrected based on the new Danish catch data.
Assessment method	A change to the SAM model instead of the currently used XSA.	Configuration and subsequent testing of the SAM model.	CANUM, WECA, mortality, etc	Model developer	<a href="#">Jan Horbovy</a> , <a href="#">Stefanie Haase</a>	

## 3.2 Ecosystem drivers

The main identified ecosystem driver for changes in sprat biomass is predation of cod. Predation mortality of cod has been included in assessments for years and during the current benchmark assessment new natural mortalities, derived from the most recent multispecies assessment model fit (SMS) were applied and tested.

The other driver used in previous years (last in 2007) was the effect of temperature on sprat recruitment (McKenzie *et. al.*, 2008). The temperature explained somewhat less than 30% of the recruitment variance and in some years, it was used to predict recruitment in the prediction year. However, as the recruitment in the prediction year constitutes only a minor part of the catches and SSB, and the temperature explains only a small part of recruitment variance, the effect (on catches and SSB) of using such predicted recruitment was only slightly different from the standard approach with geometric mean recruitment. Temperature effects can however affect long-term reference points such as  $F_{MSY}$  if differences in the temperature regime are expected.

## 3.3 Stock Assessment

### 3.3.1 Data

Below the fishery-dependent, biological, and fishery-independent data that are used as input data in the assessment model are outlined.

### 3.3.1.1 Landings and discards

The total reported landings by country used for this benchmark were kept the same as in the latest 2022 assessment (Table 7.1 in ICES 2022) with two exceptions:

1. Landings for 1974-1976, that are missing in the above-mentioned table, were taken from the Report of the Working Group on Assessment of Pelagic Stocks in the Baltic (ICES 1986).
2. Denmark provided updated national sprat and herring landing figures (see section 3.3.1.1.1), which were used to correct the historic landing data series.

The new CATON input file values are presented in Table 3.2.

Discarding at sea continued to be regarded as negligible as in the previous assessments.

#### 3.3.1.1.1 Updated Danish landings

Data provided by Denmark to the benchmark workshop represent old and corrected Danish catches from 1987 onwards (Figure 3.1). Old XSA input file (CATON from 2022 WGBFAS assessment) was used to recalculate CATON. Old CATONs were corrected using the difference between the old and corrected Danish catch time-series. For most years, the correction of old CATON (old XSA input) was within the range  $\pm 5\%$  (Figure 3.2). In the years 1990, 1991 and 1993 the CATON values increased up to 13% compared to the initial values.

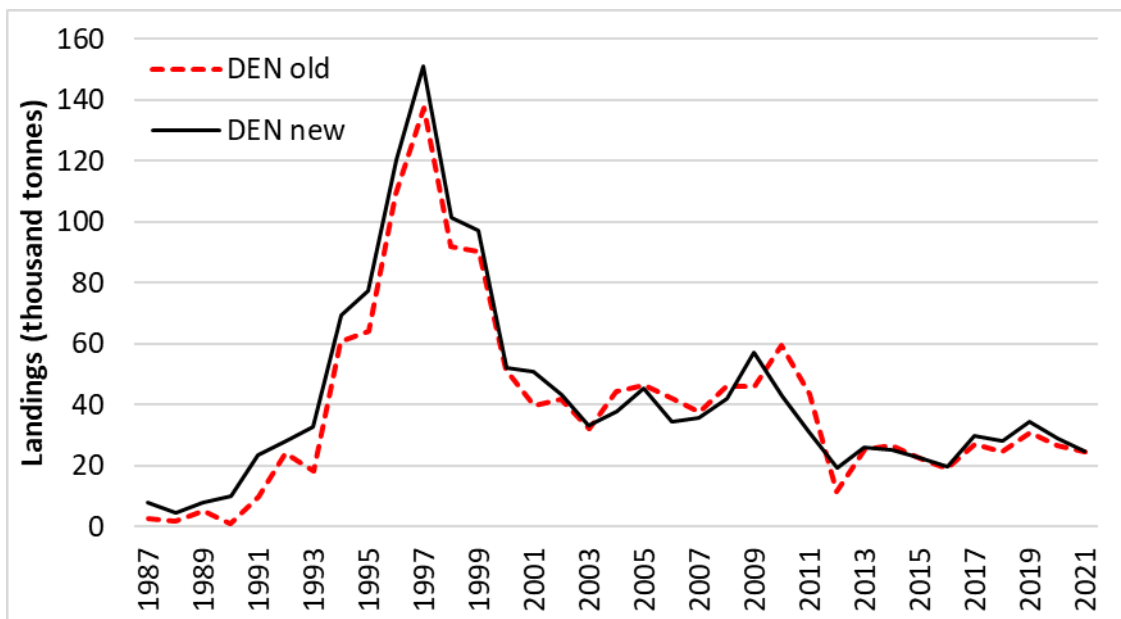


Figure 3.1. Danish landings of the sprat with indicated old (red, dashed line) and corrected (black, solid line) values.

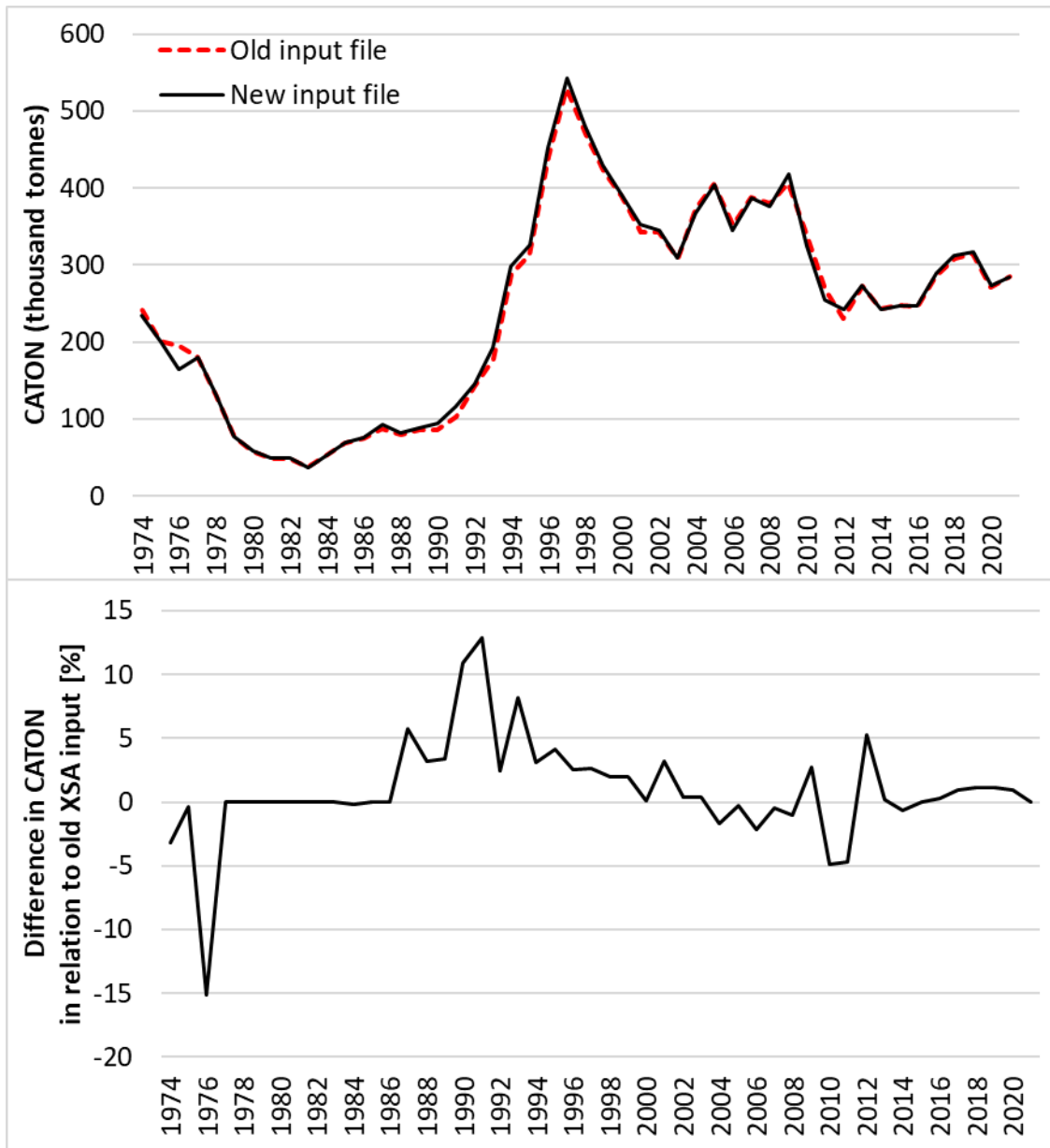


Figure 3.2. Total catch (CATON) of the sprat (upper panel) with indicated old (red, dashed line) and corrected (black, solid line) values. Difference between old and recalculated CATON (lower panel).

### 3.3.1.1.2 Misreporting

No new information about misreporting was available for the workshop with exception of Danish data (see section 3.3.1.1.1 about the updated Danish Landings). It is expected, however, that misreporting of catches might occur to some extent even in other countries, as the estimates of species composition of the clupeid catches are imprecise in some mixed pelagic fisheries.

In WD8 in Annex 5, a few assumed levels of misreporting were considered (e.g. 5, 10, 20%); option with misreporting varying from year to year was also simulated. Main outcomes of the analysis were that:

1. Overreporting of one stock by x% leads to an overestimation of its biomass by approximately the same percentage while the estimates of average fishing mortality are only slightly affected.
2. If misreporting „fluctuates“ (catches in some years are underreported while overreported in others) then

- a) biomass fluctuates similarly (compared to basic run),
- b) changes in estimated fishing mortality are relatively low.

### 3.3.1.2 Biological Information

#### 3.3.1.2.1 Catch in numbers

Old XSA catch in numbers (CANUM) input file from the latest 2022 assessment (Table 4.2.5 in ICES 2022) was used to recalculate CANUMs in accordance with the updates of the total landings in 1974-76 and updates in Danish landings 1987-2021. The ratio between the old and corrected CATON was used to up-scale or down-scale the CANUMs proportionally.

With respect to the interannual variation of the time-series, the corrected CANUMs were similar to the initial time-series and differences were visible only in a few years and in particular age groups (Figure 3.3). Apart from rescaling the original CANUM, the data were not scrutinized or discussed in detail at this benchmark. Corrected CANUM input file values are presented in table 3.3.

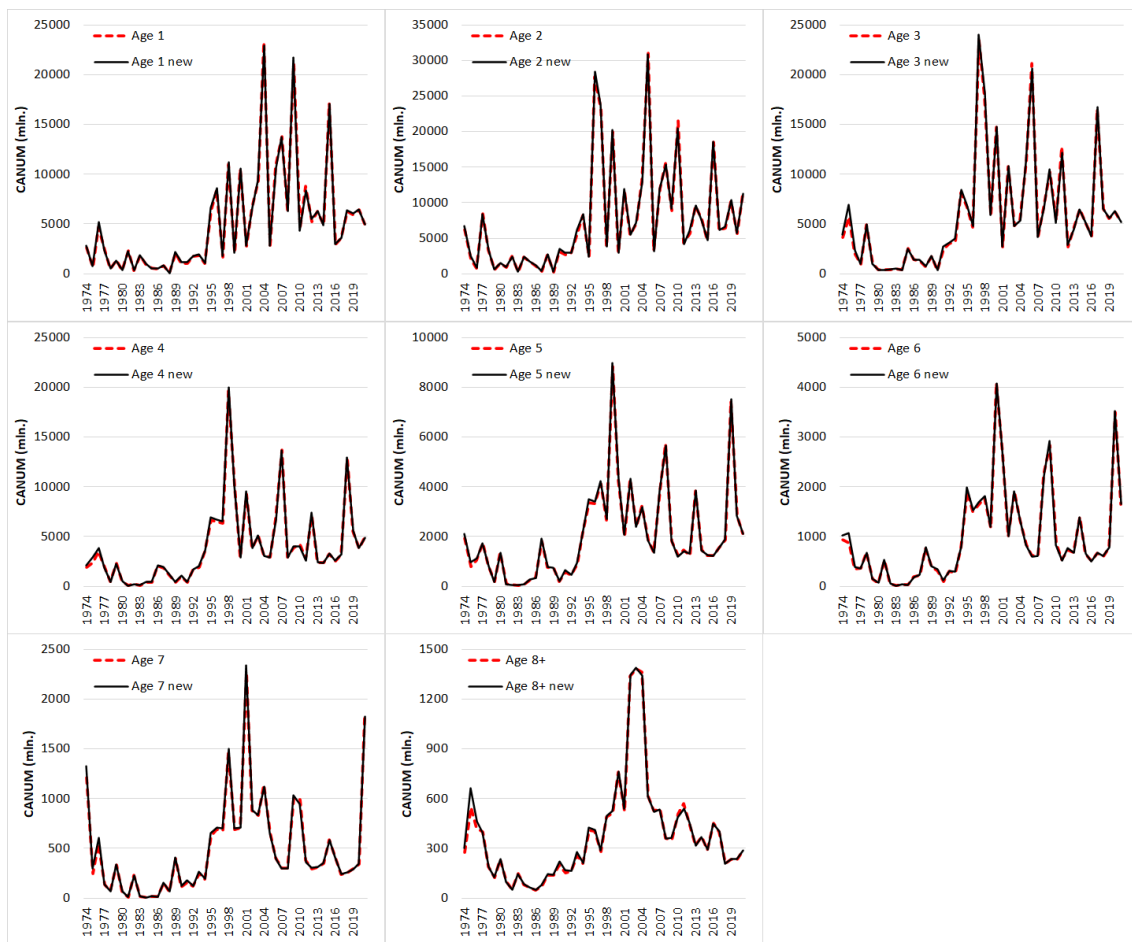


Figure 3.3. Total catch in numbers (CANUM) of the sprat by age group with indicated old (red, dashed line) and corrected (black, solid line) values. Note that the lines are strongly overlapping.

#### 3.3.1.2.2 Weight-at-age

The average weight-at-age in the catch (WECA) and the average weight-at-age in the stock (WEST) are used to estimate the SSB and CATON (catch in tonnes). WECA is annually estimated from the commercial catch data covering all seasons. At the moment, WEST=WECA is assumed in the stock assessment models as survey data are only available from the annual Baltic International Acoustic Survey (BIAS) in quarter 4. The mean weights in the catch from the first quarter

could be an alternative candidate to be taken as mean weight in the stock. However, it was not possible to compile these data during the benchmark due to a lack of corresponding input data for the years before 2008.

Another alternative candidate could be to use BIAS data from quarter 4 to represent the mean weight in the stock for the following year. In order to test if the commercial fishery is selecting for individuals with higher mean weight at age compared to the weight-at-age in the stock and thus overestimates WEST by using WECA, we compare weight data in quarter 4 from commercial samples with weight data from the BIAS survey. Survey catches cover a larger part of the distribution area of the stocks and are taken using a smaller mesh size in the codend.

WECA (weight-at-age) files per ICES SD were extracted from Intercatch (IC) and selected for quarter 4 to ensure comparability between WECA and WEST (<https://intercatch.ices.dk/login.aspx>).

WEST was calculated based on the survey catches of the BIAS which were uploaded to the Acoustic Trawl Database (<https://www.ices.dk/data/data-portals/Pages/acoustic.aspx>). Individual fish, which are sampled length-stratified in the BIAS, were used to calculate an age-length key (ALK) per SD and year. Further, a length-weight relationship (LWR) was estimated per SD and year. The number of fish per length class in the catches was standardized to the 30-minute hauls and total length distributions were calculated using standardized values. Lengths were transformed to weights by the LWR and an ALK was applied to get weight-at-ages per SD, year and species.

WECA and WEST were compared for ICES SDs 25-32 and the years 2015-2021 as these years were available in the Acoustic Trawl Database.

Weight-at-age in the stock (WEST) and the catch (WECA) shows a high correlation (figures 3.4 and 3.5), but the correlation is slightly lower compared to herring. Differences in WEST and WECA are larger for younger age groups. Figure 3.6 shows the difference between WECA and WEST per age group and SDs over the years for sprat. Differences are particularly high for age groups 1 and 2. In most years, WEST is larger than WECA in SD 27-32 independent of the age class. WECA is larger than WEST in SD 25 and 26 in age class two and older. The weight-at-age shows a distinct variation between years in some age groups and SDs in both WECA and WEST (e.g. SD 30).

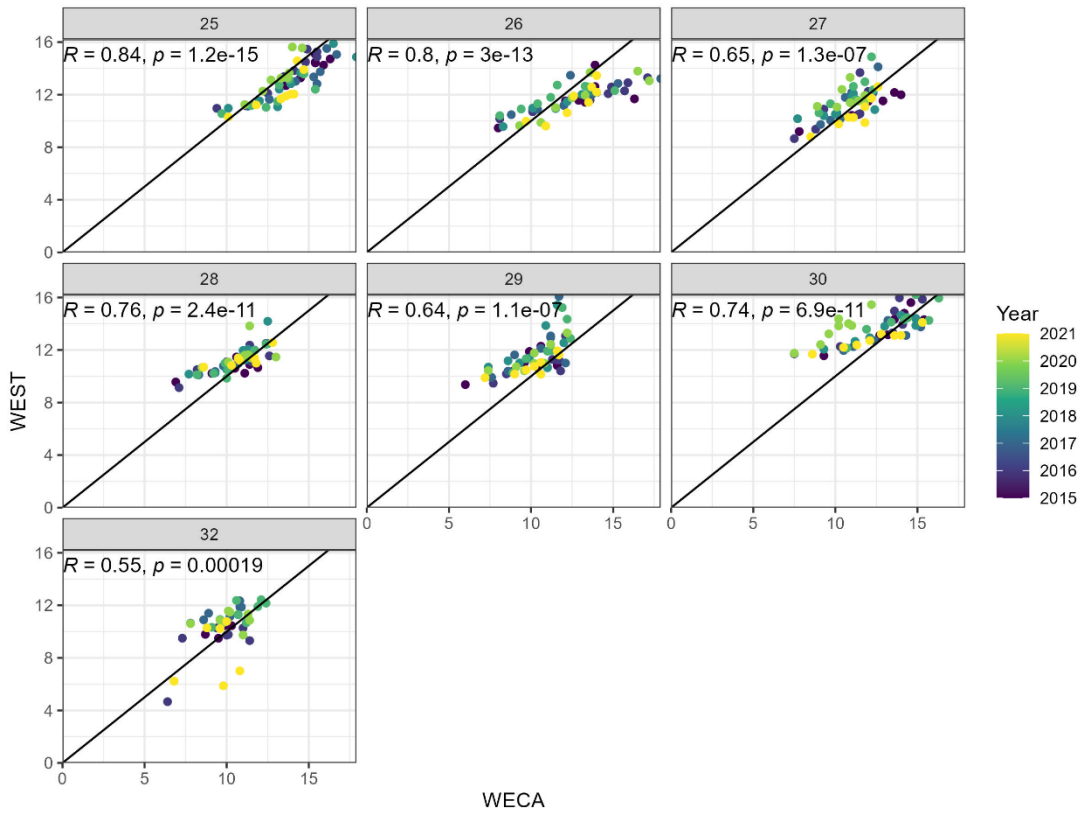


Figure 3.4. Comparison of weight-at-age in stock (WEST) and catch (WECA) of sprat per ICES subdivision. The black line refers to the line through the origin.

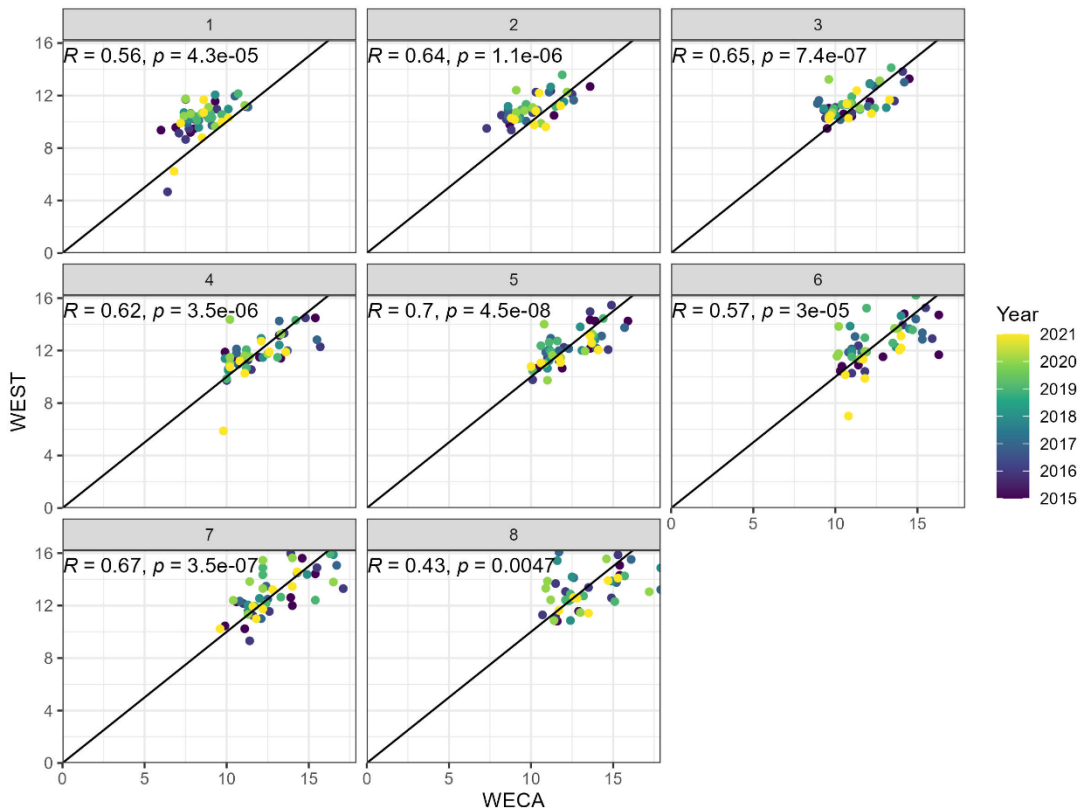
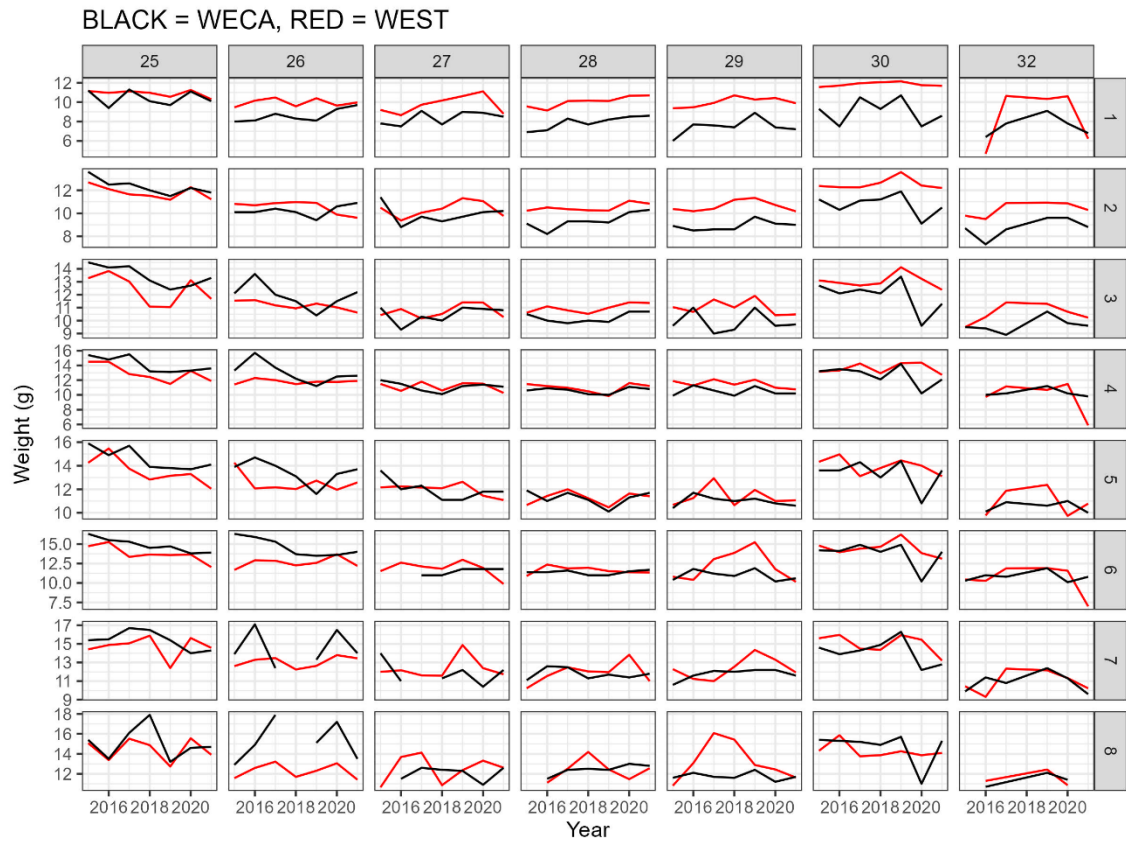


Figure 3.5. Comparison of weight-at-age in stock (WEST) and catch (WECA) of sprat per age group. The black line refers to the line through the origin.



**Figure 3.6.** Comparison of weight-at-age in catch (WECA; black) and weight-at-age in the survey (WEST; red) per ICES subdivision (columns), age (rows), and year for sprat.

Differences in weight between the survey and commercial catches might occur due to the different selectivity of the gears and unequal geographical coverage of fishing effort and surveys. One further explanation for the difference in weight could be that the BIAS samples are weighted and measured fresh while commercial samples are often frozen before they are measured and weighted. Clupeids shrink when they are frozen, and this effect can be higher for smaller fish than for larger specimens (e.g. Santos *et al.*, 2009).

A quarterly coverage of survey weights is currently unavailable. Therefore, we only compared the weight-at-age from quarter 4. As SSB is calculated at the beginning of the year, weights from the quarter 4 BIAS could be used as WEST estimates. There is, however, a large difference in weights-at-age between the different SDs in both WEST and WECA. As there is currently no spatially resolved stock assessment in place, one WECA value is used for all SDs. Further investigation is needed on how to raise weight-at-ages from the survey based on the spatial distribution of the stocks to implement weight-at-age from the survey as WEST.

It was decided to assume (as previously done) that the mean weight in the stock is equivalent to the mean weight in the catch. The mean weights in the catch used for this benchmark were kept the same as in the latest 2022 assessment (Table 7.7 in ICES 2022).

### 3.3.1.2.3 Natural mortality

Natural mortalities were provided to the group from the multispecies stochastic age-structured model (SMS) calculations made by WGSAM (ICES 2023b). The SMS model is updated every few years. In years for which SMS estimates will be missing,  $M$  will be assumed to be equal to the previous year's level or will be estimated from the regression of  $M$  against cod biomass. The same procedure was applied in previous assessments of sprat stock.



New natural mortality values are presented in Table 3.4.

### 3.3.1.3 Tuning indices

The data on sprat stock size in the Baltic Sea, estimated by hydroacoustic methods, are collected annually by two internationally coordinated surveys. The Baltic International Acoustic Survey (BIAS) is conducted in autumn (October) and the Baltic Acoustic Spring Survey (BASS) in May. The results from the individual national surveys are placed in the Access-databases. These database files include queries with the algorithms used to create the report tables and the calculation of the different tuning fleets.

WGBIFS (ICES 2023a) provides in their latest report (Table 4.12) an updated tuning index for the assessment of the Baltic sprat based on the BIAS sprat abundance estimates in the ICES Subdivisions 22-29 per age-group (1-8+) for the years 1991-2021. Additionally, also the recruitment index for Baltic sprat (age 0) is presented there (in Table 4.13). Compared to the previous tuning indices used in the assessment, even some historic corrections were made. Namely, Finland presented corrections for their 2016, 2018 and 2019 survey results, which were implemented in the BIAS database. As a result, the sprat abundance estimates changed very slightly for those years. WGBIFS (ICES 2023a) recommends that the updated and corrected BIAS index series can be used in the assessment of the Baltic sprat stock with the restriction that the years 1993, 1995 and 1997 are excluded from the index series.

Table 4.19 of the WGBIFS report (ICES 2023a) gives the sprat tuning index based on the BASS survey in ICES Subdivisions 24, 25, 26 and 28\_2. WGBIFS recommends that the BASS index series can be used in the assessment of sprat stock in the Baltic Sea with the restriction that the year 2016 is excluded from the index series.

During the WGBFAS 2022 meeting, stock assessors of Baltic sprat requested that WGBIFS would provide WKBBALTPEL with some alternative acoustic tuning indices (e.g. indices calculated with StoX and/or including data from the Gulf of Finland). These requests were discussed during the second meeting of WGBIFS in November 2022. Based on the results of the comparison exercises between StoX and the traditional BIAS calculation methods WGBIFS concluded that the StoX calculated acoustic time-series cannot be used for the stock assessment of Baltic sprat yet. WGBIFS (ICES 2023a) provided (in Table 5.3.) a new Baltic sprat tuning index, which also includes the survey data from the Gulf of Finland (SD 32). WGBIFS (ICES 2023a) recommends that the alternative BIAS index series (including data from SD 32) can be tested in the benchmark process of the Baltic sprat stock with the restriction that the years 1999, 2001-2005 and 2008 are excluded from the index series.

Four acoustic time-series (Table 4.5) were selected for the final assessment of Baltic sprat: BASS tuning fleet index for Baltic sprat in the SDs 24–26 and 28 for the years 2001–2021, BIAS tuning fleet index for Baltic sprat in the SDs 22–29 for the years 1991–2008, BIAS tuning fleet index for Baltic sprat in the SDs 22–29 and 32 for the years 2000–2021, and BIAS tuning fleet index for Baltic sprat recruitment (age 0) in the SD 22–29 and 32 for the years 2010–2021. Index values for the year 2016 were excluded from the BASS time-series. Index values for the years 1993, 1995 and 1997 were excluded from the BIAS time-series. The new BIAS index (Fleet 2) was decided to be used in parallel with the old BIAS index (Fleet 1), while the old index is used only for these years that are lacking in the new index (Table 4.5).

### 3.3.1.4 Summary of data input to the assessment model

Here below are presented the input data for the assessment model that has been updated/corrected compared to the latest assessment (ICES 2022).

**Table 3.2. CATON input data of the sprat used for the benchmark assessment (total catch in thousand tonnes).**

Year	CATON
1974	234.1
1975	200.7
1976	165.3
1977	180.8
1978	132.4
1979	77.1
1980	58.1
1981	49.3
1982	48.7
1983	37.3
1984	52.5
1985	69.5
1986	75.8
1987	93.4
1988	82.8
1989	88.7
1990	94.9
1991	116.5
1992	145.7
1993	192.6
1994	297.6
1995	326.0
1996	452.3
1997	543.2
1998	480.3
1999	429.7
2000	389.6
2001	353.2
2002	344.6

Year	CATON
2003	309.4
2004	367.3
2005	404.4
2006	344.6
2007	386.8
2008	376.6
2009	418.1
2010	324.7
2011	255.4
2012	243.0
2013	272.9
2014	242.2
2015	247.3
2016	247.1
2017	288.5
2018	312.2
2019	317.7
2020	274.1
2021	284.9

**Table 3.3. CANUM input data of the sprat used for the benchmark assessment (catch in numbers in mln.).**

Year	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8+
1974	2854.471	6737.206	3949.321	2117.657	2105.650	1018.440	1324.081	303.458
1975	764.470	2473.570	6911.876	2905.715	961.674	1068.797	300.675	664.650
1976	5158.494	901.249	2320.331	3867.218	1145.842	385.620	603.771	464.948
1977	2371.000	8399.000	997.000	1907.000	1739.000	364.000	140.000	399.000
1978	500.000	3325.000	4936.000	480.000	817.000	683.000	73.000	189.000
1979	1340.000	597.000	1037.000	2291.000	188.000	150.000	335.000	125.000
1980	369.000	1476.000	378.000	500.000	1357.000	72.000	67.000	235.000
1981	2303.000	920.000	405.000	94.000	88.000	527.000	13.000	99.000

Year	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8+
1982	363.000	2460.000	425.000	225.000	64.000	57.000	231.000	51.000
1983	1852.000	297.000	531.000	107.000	47.000	12.000	18.000	148.000
1984	1005.000	2393.000	388.000	447.000	77.000	38.000	9.000	83.000
1985	566.000	1703.000	2521.000	447.000	271.000	30.000	19.000	65.000
1986	495.000	1142.000	1425.000	2099.000	340.000	188.000	16.000	50.000
1987	824.719	417.123	1397.470	1940.577	1910.934	240.322	157.745	77.284
1988	80.472	2781.435	753.133	1185.411	786.147	784.084	67.060	145.468
1989	2172.410	299.714	1831.356	417.533	763.754	403.064	411.332	141.589
1990	1163.452	3516.976	383.751	1055.869	208.512	350.478	124.220	221.821
1991	1178.590	2990.502	2753.429	459.469	642.354	119.665	180.627	171.595
1992	1827.431	3013.928	3117.503	1684.887	455.320	318.929	124.085	167.156
1993	1981.094	6147.662	3508.006	2052.465	955.942	288.729	263.857	277.915
1994	1111.832	8417.569	8424.782	3632.260	2267.973	802.704	198.873	214.329
1995	6646.975	2441.639	6928.582	6921.281	3510.704	1983.767	653.955	426.583
1996	8603.566	28382.846	4824.315	6683.686	3407.993	1537.340	707.648	413.308
1997	1762.808	23786.621	24005.176	6508.435	4215.143	1694.061	700.814	286.277
1998	11239.592	3879.485	18043.738	20012.554	2712.477	1813.759	1497.524	498.835
1999	2116.903	20234.621	5929.768	10139.171	8984.127	1199.782	698.517	523.633
2000	10548.782	2951.857	14735.252	2873.755	4289.605	4082.334	707.925	761.996
2001	2865.053	11927.743	2755.652	9548.800	2063.127	2736.043	2336.628	539.778
2002	6674.521	5450.658	10824.009	3850.299	4325.186	1001.981	883.511	1345.346
2003	9401.375	7135.850	4823.148	5086.138	2405.049	1910.187	836.146	1388.223
2004	22865.653	12869.793	5354.714	3033.159	3190.419	1311.158	1123.429	1340.644
2005	2836.706	30899.441	11229.085	2927.505	1863.864	841.134	657.541	613.638
2006	10619.360	3196.279	20646.634	6686.155	1350.541	600.893	396.354	518.686
2007	13722.736	11904.443	3686.319	13650.123	3834.528	619.692	299.402	536.138
2008	6324.583	15318.705	6615.024	2906.164	5659.491	2232.003	295.969	358.718
2009	21720.304	9132.909	10457.988	4011.207	1844.250	2914.404	1035.851	362.547
2010	4359.496	20441.572	5100.731	4027.269	1178.808	837.906	945.070	485.887

Year	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8+
2011	8389.357	4157.842	12126.805	2620.974	1402.791	523.667	361.108	541.549
2012	5491.110	6052.392	2881.288	7442.494	1316.519	764.583	310.009	453.991
2013	6277.500	9586.966	4494.926	2395.068	3855.862	683.709	310.577	317.567
2014	4879.317	7569.532	6456.413	2358.146	1449.131	1393.046	350.105	369.394
2015	17062.390	4721.734	5122.951	3273.052	1245.001	659.271	584.741	292.927
2016	2981.093	18565.096	3810.393	2553.854	1229.387	509.378	407.220	451.724
2017	3614.921	6201.106	16705.645	3226.989	1578.918	682.113	243.671	402.254
2018	6346.667	6567.815	6543.666	12934.390	1891.634	616.832	258.340	209.799
2019	6028.654	10377.984	5622.130	5605.427	7528.813	785.873	293.900	237.821
2020	6499.891	5708.479	6277.637	3845.035	2843.776	3525.079	343.622	236.476
2021	4943.822	11224.038	5225.472	4918.213	2113.486	1649.830	1825.537	287.538

**Table 3.4. Natural mortality input data of the sprat used for the benchmark assessment.**

Year	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8+
1974	0.755	0.543	0.478	0.452	0.452	0.439	0.449	0.449
1975	0.762	0.566	0.503	0.476	0.476	0.462	0.472	0.472
1976	0.623	0.485	0.437	0.418	0.418	0.409	0.416	0.416
1977	0.816	0.572	0.485	0.464	0.464	0.451	0.46	0.46
1978	1.158	0.784	0.72	0.64	0.629	0.619	0.619	0.619
1979	1.269	0.835	0.766	0.771	0.713	0.718	0.733	0.733
1980	1.264	0.886	0.757	0.741	0.751	0.713	0.731	0.731
1981	1.13	0.717	0.676	0.638	0.641	0.668	0.619	0.619
1982	1.124	0.768	0.684	0.665	0.637	0.666	0.674	0.674
1983	0.867	0.676	0.608	0.59	0.576	0.564	0.561	0.561
1984	0.722	0.595	0.522	0.517	0.501	0.495	0.493	0.493
1985	0.636	0.517	0.483	0.468	0.45	0.434	0.443	0.443
1986	0.651	0.486	0.461	0.434	0.419	0.413	0.41	0.41
1987	0.656	0.485	0.439	0.421	0.416	0.416	0.405	0.405
1988	0.626	0.476	0.461	0.43	0.414	0.411	0.4	0.4
1989	0.515	0.404	0.375	0.369	0.361	0.358	0.354	0.354

Year	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8+
1990	0.371	0.308	0.303	0.297	0.291	0.293	0.286	0.286
1991	0.33	0.27	0.267	0.262	0.26	0.259	0.26	0.26
1992	0.346	0.28	0.268	0.264	0.261	0.259	0.257	0.257
1993	0.38	0.338	0.322	0.312	0.308	0.304	0.298	0.298
1994	0.378	0.334	0.317	0.305	0.303	0.299	0.298	0.298
1995	0.334	0.301	0.299	0.292	0.287	0.286	0.285	0.285
1996	0.305	0.293	0.279	0.277	0.271	0.27	0.27	0.27
1997	0.298	0.28	0.274	0.266	0.259	0.258	0.256	0.256
1998	0.307	0.286	0.28	0.277	0.269	0.267	0.268	0.268
1999	0.337	0.304	0.292	0.293	0.291	0.284	0.281	0.281
2000	0.376	0.316	0.318	0.313	0.309	0.306	0.298	0.298
2001	0.391	0.333	0.319	0.32	0.314	0.316	0.317	0.317
2002	0.405	0.341	0.337	0.33	0.33	0.329	0.33	0.33
2003	0.366	0.315	0.309	0.308	0.303	0.307	0.308	0.308
2004	0.345	0.316	0.296	0.289	0.292	0.29	0.291	0.291
2005	0.399	0.363	0.349	0.326	0.321	0.316	0.321	0.321
2006	0.429	0.375	0.369	0.36	0.341	0.335	0.336	0.336
2007	0.44	0.38	0.362	0.362	0.36	0.346	0.335	0.335
2008	0.466	0.382	0.373	0.363	0.367	0.369	0.35	0.35
2009	0.465	0.383	0.368	0.361	0.358	0.363	0.356	0.356
2010	0.504	0.43	0.401	0.39	0.387	0.383	0.386	0.386
2011	0.515	0.417	0.409	0.394	0.381	0.383	0.377	0.377
2012	0.487	0.38	0.357	0.356	0.347	0.343	0.345	0.345
2013	0.488	0.372	0.343	0.335	0.334	0.332	0.333	0.333
2014	0.491	0.378	0.356	0.338	0.332	0.333	0.338	0.338
2015	0.4	0.327	0.314	0.306	0.301	0.297	0.303	0.303
2016	0.376	0.336	0.309	0.298	0.295	0.291	0.292	0.292
2017	0.355	0.309	0.301	0.286	0.28	0.28	0.28	0.28
2018	0.342	0.296	0.288	0.285	0.276	0.271	0.274	0.274

Year	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8+
2019	0.336	0.297	0.284	0.279	0.279	0.267	0.267	0.267
2020	0.321	0.277	0.273	0.268	0.265	0.266	0.26	0.26
2021	0.31	0.276	0.269	0.266	0.263	0.258	0.262	0.262

**Table 3.5. Tuning fleet input data of the sprat used for the benchmark assessment.**

**Fleet 1. BIAS in October in SD 22-29 corrected by area surveyed (abundance in millions).**

Year	Fish. Effort	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8+
1991	1	46488	40299	43681	2743	8924	1851	1957	3117
1992	1	36519	26991	24051	9289	1921	2437	714	560
1993	1	NA	NA	NA	NA	NA	NA	NA	NA
1994	1	12532	44588	43274	17272	11925	5112	1029	1559
1995	1	NA	NA	NA	NA	NA	NA	NA	NA
1996	1	69994	130760	20797	23241	12778	6405	3697	1311
1997	1	NA	NA	NA	NA	NA	NA	NA	NA
1998	1	100615	21975	55422	36291	8056	4735	1623	1011
1999	1	4892	90050	15989	35717	38820	5231	3290	1738
2000	1	NA	NA	NA	NA	NA	NA	NA	NA
2001	1	12047	35687	6927	30237	4028	9606	6370	2407
2002	1	31209	14415	36763	5733	18735	2638	5037	4345
2003	1	99129	32270	24035	23198	8016	13163	4831	8536
2004	1	119497	47027	11638	7929	4876	2450	2389	3552
2005	1	7082	125148	48724	10035	5116	3011	2364	3325
2006	1	NA	NA	NA	NA	NA	NA	NA	NA
2007	1	NA	NA	NA	NA	NA	NA	NA	NA
2008	1	28805	45118	20134	5350	18820	5678	1241	1917

**Fleet 2. BIAS in October in SD 22-29 and 32 corrected by area surveyed (abundance in millions).**

Year	Fish. Effort	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8+
2000	1	72295	8611	53087	8052	16597	15982	1739	2753
2001	1	NA	NA	NA	NA	NA	NA	NA	NA
2002	1	NA	NA	NA	NA	NA	NA	NA	NA
2003	1	NA	NA	NA	NA	NA	NA	NA	NA
2004	1	NA	NA	NA	NA	NA	NA	NA	NA
2005	1	NA	NA	NA	NA	NA	NA	NA	NA
2006	1	83120	24175	147488	52014	10143	5143	2278	3491
2007	1	75613	39491	12088	40276	15871	1516	768	2379
2008	1	NA	NA	NA	NA	NA	NA	NA	NA
2009	1	134253	49826	39347	9935	9111	13065	4102	2176
2010	1	15367	88035	14904	9019	2161	2967	3707	1560
2011	1	34095	20175	68118	17115	8393	3072	1838	3188
2012	1	108251	28703	15212	43526	6640	3453	2135	4196
2013	1	38416	35889	17151	8465	15537	3171	1116	2739
2014	1	19021	33428	22062	11957	5857	9166	1771	2026
2015	1	162639	18894	22417	12790	4198	3964	3086	2164
2016	1	33849	119884	29659	11196	5441	2461	1506	1805
2017	1	48761	52739	103922	15961	7473	3698	1230	2445
2018	1	41907	24557	16383	39840	11997	3293	1434	1905
2019	1	17161	28807	15797	12692	29391	4002	1642	2404
2020	1	62659	19408	21467	9689	8402	17421	1226	1343
2021	1	100173	70693	23649	19445	7632	6306	12185	1910



**Fleet 3. BASS in May in SD 24-27 and 28 corrected by area surveyed (abundance in millions).**

Year	Fish. Effort	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8+
2001	1	8 225	35 735	12 971	37 328	5 384	4 635	4 526	600
2002	1	27 412	18 982	36 814	19 045	14 759	2 517	3 670	2 585
2003	1	26 469	16 471	8 423	15 533	5 653	7 170	1 660	3 607
2004	1	136 162	65 566	15 784	11 042	12 655	3 271	7 806	6 321
2005	1	4 359	88 830	23 557	7 258	3 517	2 781	1 830	2 243
2006	1	13 417	7 980	76 703	21 046	5 702	1 970	1 526	1 943
2007	1	51 569	28 713	6 377	36 006	7 481	1 261	533	698
2008	1	9 029	40 270	20 164	5 627	21 188	4 210	757	1 477
2009	1	39 412	26 701	36 255	10 549	6 312	14 106	5 341	964
2010	1	9 387	58 680	15 199	15 963	5 062	1 654	5 566	1 273
2011	1	18 092	6 791	66 160	16 689	10 565	4 077	2 399	3 382
2012	1	22 700	22 080	11 274	35 541	7 515	5 025	1 367	2 158
2013	1	24 877	35 333	18 393	11 358	14 959	3 385	2 164	950
2014	1	10 145	26 907	19 857	7 458	6 098	3 810	1 217	1 058
2015	1	70752	24660	29744	18935	8081	4074	2581	1721
2016	1	NA	NA	NA	NA	NA	NA	NA	NA
2017	1	32701	36292	132939	20630	6790	2250	809	942
2018	1	27209	25642	38632	69259	7251	2086	1025	619
2019	1	15958	28778	32532	49495	30131	3384	487	647
2020	1	38096	26252	29054	19630	18377	11756	473	376
2021	1	23212	45545	20134	18028	8525	7160	5361	911

**Fleet 4. Age 0 shifted to represent age 1 from BIAS in October in SD 22-29 and 32 corrected by area surveyed (abundance in millions).**

Year	Fish. Effort	Age 0 shifted to 1
2010	1	14528
2011	1	53562
2012	1	49130
2013	1	34941
2014	1	25347
2015	1	182073
2016	1	43534
2017	1	32784
2018	1	126748
2019	1	19371
2020	1	122062
2021	1	111155

### 3.3.2 Assessment model

Up to 2022, the sprat stock was evaluated using the stock assessment model XSA (Shepherd, 1999) as the primary model. Since the previous benchmark (ICES, 2013), SAM (Nielsen and Berg, 2014) has been run in parallel as a secondary assessment model. Usually, both models showed similar trends of stock and fishing mortality.

During WKBALTPEL, it was decided to test SAM as the primary assessment model for a few reasons, both practical and methodological. The more important methodological reasons are:

- SAM treats catches as observations with noise while in XSA catches are treated as exact
- SAM provides confidence intervals for the parameters and assessment estimates and enables probabilistic catch projections (in XSA it is also possible to provide approximate confidence intervals e.g. by bootstrapping, but it requires additional programming).
- SAM provides estimates of parameters based on maximum likelihood while XSA estimates are obtained iteratively based on the fixed-point theorem.

The practical reason for changing to SAM is that XSA is no longer supported in FLR and its DOS version can only be run under older Windows versions. For more recent Windows versions (e.g. Windows 10 pro), DOS is usually run on a virtual machine which requires special arrangements.

The input data were described in sections 3.3.1. The consistency of both catch-at-age and survey-at-age observations was verified (Figures 3.7-3.8), showing good consistency of the catch numbers and survey indices.

To tune SAM, two survey indices were available: the October acoustic survey (BIAS) and the May acoustic survey (BASS). They resulted in four tuning fleets:

- fleet1: October acoustic survey (BIAS) in the years 2000-2021 (gaps in years 2001-2005 and 2008) covering the ages 1-8 and subdivisions 22-29+32,
- fleet2: October acoustic survey (BIAS) in the years 1991-2008 covering the ages 1-8 and subdivisions 22-29 (years from this fleet which overlapped with above fleet1 were excluded),
- fleet3: May survey (BASS) in the years 2001-2021 covering the ages 1-8 and subdivisions 24-26+28,
- fleet4: October (BIAS) survey covering the age 0 sprat and subdivisions 22-29+32 in 2010-2021; the age 0 series was shifted to represent the age 1 the following year.

Following the recommendation of WGBIFS, some years were excluded from the tuning fleets due to poor area coverage.

The SAM assessments were performed using the *stockassessment* package in R; the analyses were reported as Word files using R-Markdown. In the produced Word docs for each SAM run, basic results and figures were included, e.g. distribution of residuals, leave-one-out analysis, retrospective runs and Mohn's rho, model fits to observations, catchabilities, AIC, and loglikelihood values (all this was calculated using *stockassessment* commands).

In addition, the mean square residual by fleet, and mean residual were calculated. Linear regressions were fitted to survey indices and SAM estimates of stock numbers and it was tested if the slopes of the regressions are significantly different from 1 (if they are, catchability may be considered dependent on year-class strength). The survey indices were plotted against SAM estimates of the stock numbers to check if fitted regressions are not biased by outliers. SAM provides confidence intervals for catchabilities but the provided regression analysis was an additional way of testing the catchabilities.

At the end of the Word file report, a summary of results (about half a page starting with line ## Summary) is attached.

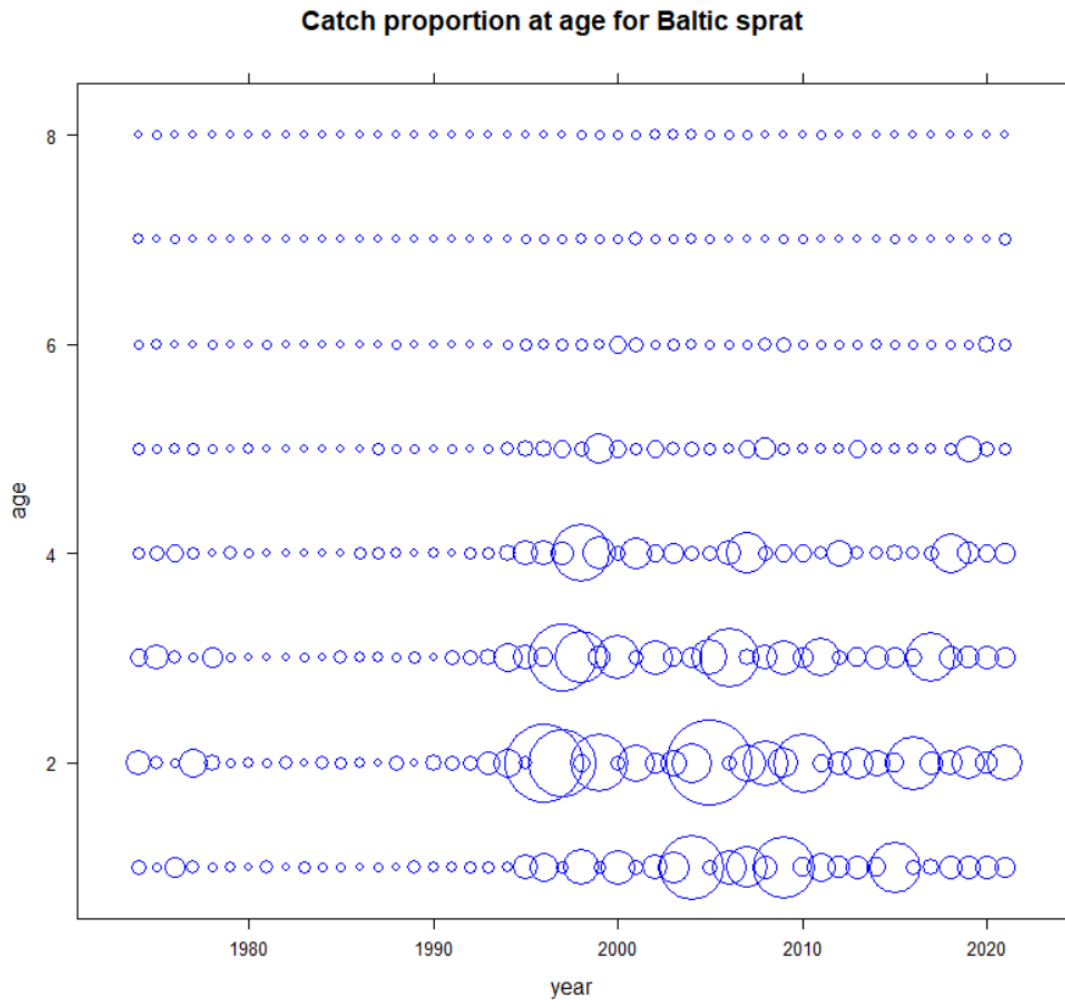


Figure 3.7. Canum consistency check visualised in bubbles plot.

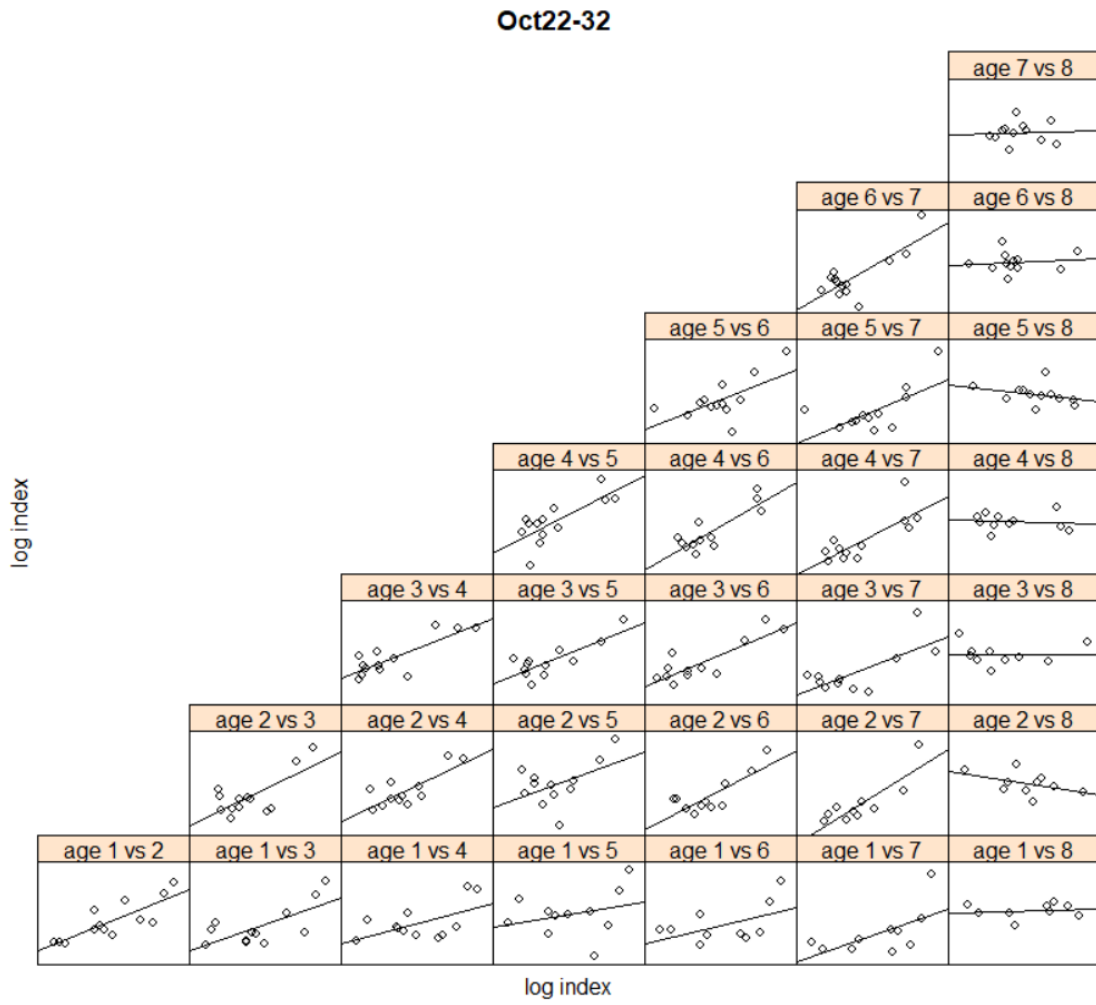


Figure 3.8a. Check for consistency in October 22-32 acoustic survey estimates.

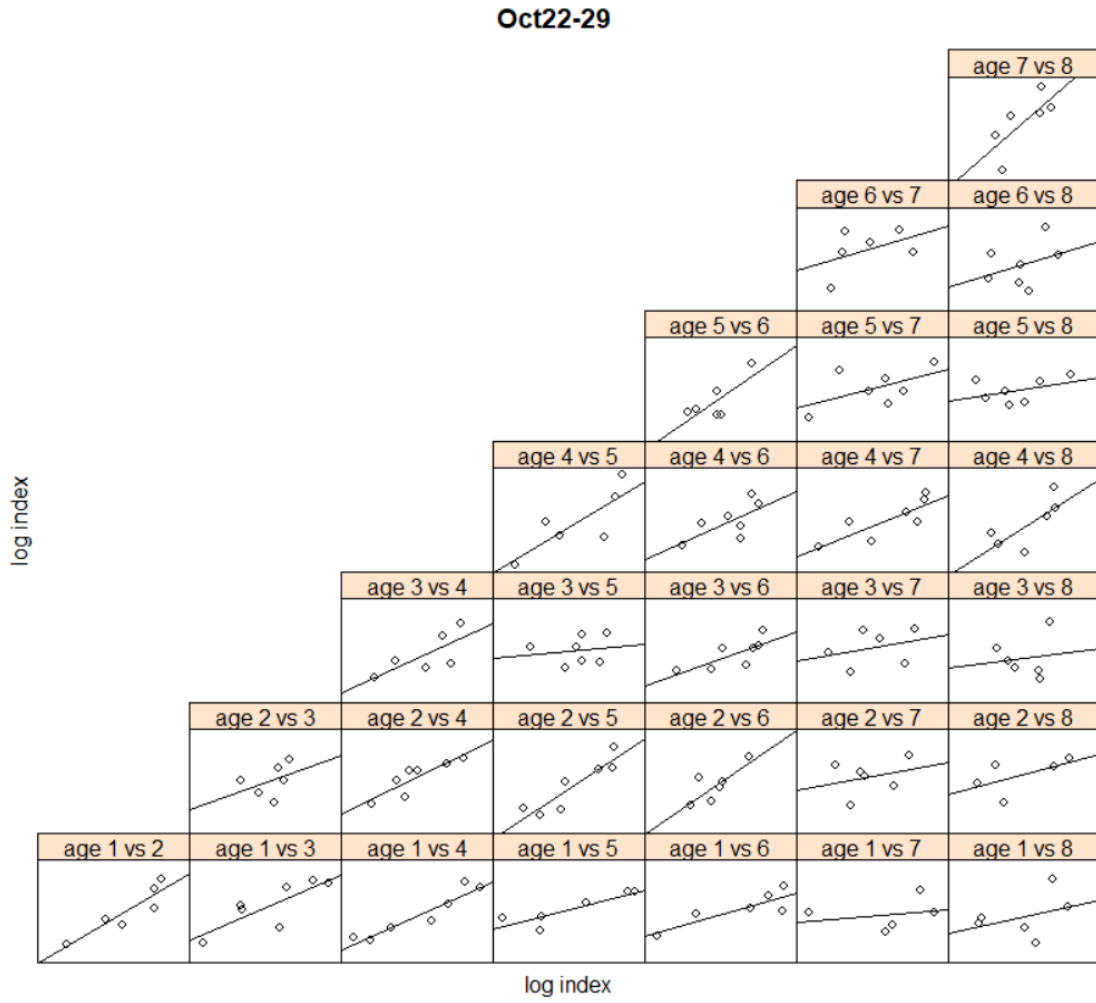


Figure 3.8b. Check for consistency in October 22-29 acoustic survey estimates.

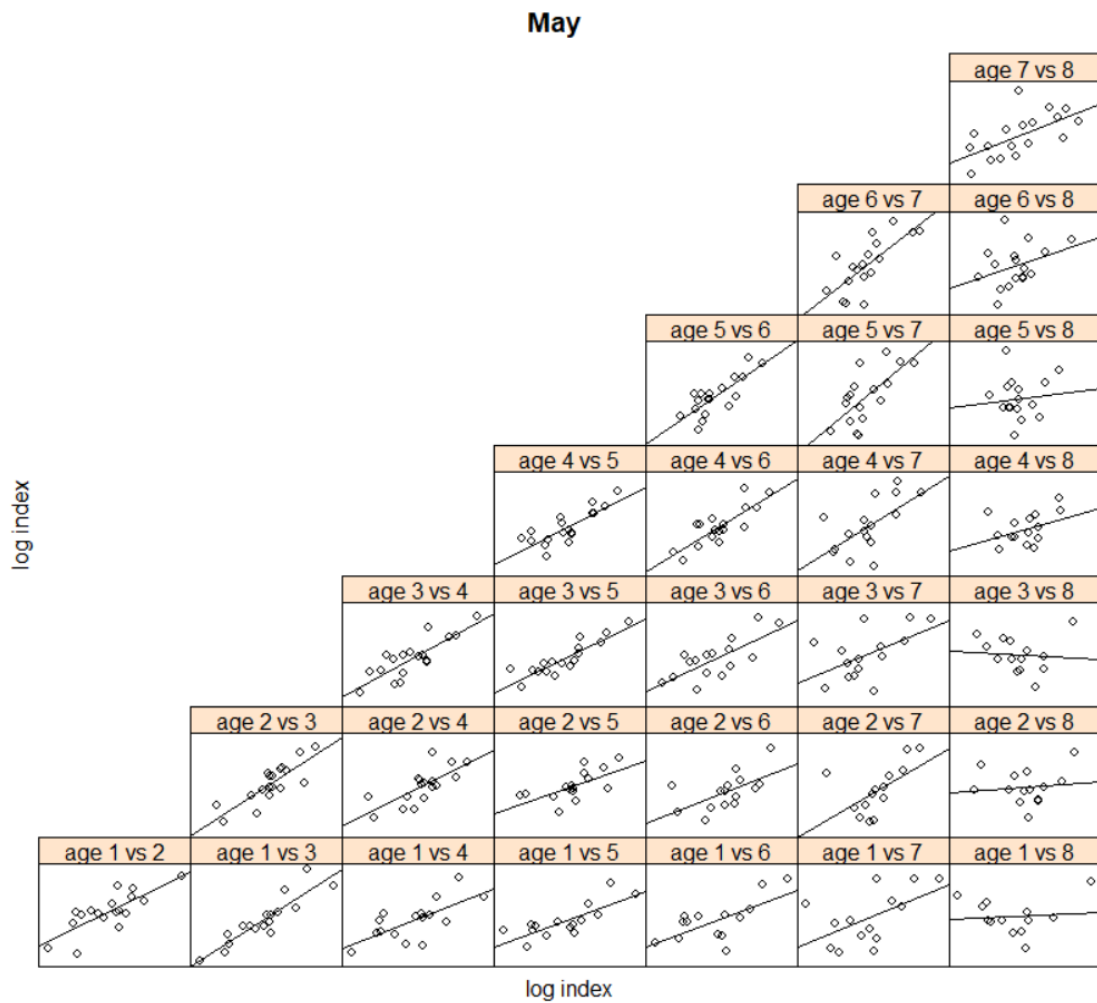


Figure 3.8c. Check for consistency in May acoustic survey estimates.

### 3.3.3 Model settings

The steps to derive the final SAM parameterisation were as follows:

1. The  $q$  (catchability) plateau was set at the two oldest ages (7-8 for sprat), and  $q$  was assumed dependent on the  $y$ - $c$  strength for all ages.
2. Based on the above, ages for which  $q$  was dependent on  $y$ - $c$  strength were found.
3. The  $q$  plateau was set at lower ages than in step 1, and effect on the model outcomes in terms of AIC, residuals distribution, retrospective patterns etc was evaluated.
4. Different covariance structures for each fleet were tested, e.g. "ID" (independent), "AR" AR(1) (autoregressive) by changing parameters: \$obsCorStruct in the configuration file.
5. The AIC, retrospective patterns (Mohn's  $\rho$ ), residuals distribution, leave-one-out analysis etc. were examined to select the final parameterisation.

Several SAM runs were performed. The five most important runs were sent to the group for examination before the meeting.

Basic runs performed before the meeting:

1. XSA based run:  $q$  plateau at ages 5-8, age 1 with  $q$  dependent on  $y$ - $c$  strength, results in the Word file spratSAMrmdA-xsabased.docx.

2. Run with q plateau at two oldest ages (7-8) and q for all ages dependent on y-c strength, results in the Word file spratSAMrmd-q-7-8-allPower-A.docx.
3. Run with q plateau at the two oldest ages (7-8), only age 1 with q dependent on y-c strength, results in the Word file spratSAMrmd-q-7-8-Power-1-A.docx.
4. Run with q plateau at ages 6-8, only age 1 with q dependent on y-c strength, results in the Word file spratSAMrmd-q-6-8-Power-1-A.docx.
5. As run 4 but "AR" structure was assumed for \$obsCorStruct (except fleet 4, i.e. October survey in SD 22-29, as SAM with such assumption did not provide solution): "AR" "AR" "AR" "ID" "AR", results in the Word file spratSAMrmd -q-6-8-AR in obsCorStruc-A-1.docx.

Run 4 was selected as „best“, as it had the second lowest AIC and quite good retrospective patterns expressed by low Mohn’s rho values. Run 5 with AR structure had an even lower AIC but a much worse retrospective pattern with two times higher rho values compared to Run 4.

During the meeting, a few other runs were tested and compared to Run 4. This included e.g. runs using tuning fleets from previous XSA assessment (subdivision 32 was not included in BIAS survey in former assessments) or a run in which the recruitment was modelled by the Beverton & Holt S-R relationship. It showed that the inclusion of subdivision 32 in the BIAS survey improved the retrospective pattern (markedly lower Mohn’s rho values). Differences between runs in SSB and  $F_{bar}$  estimates in the terminal year were rather low and ranged between  $\pm 5\%$ .

As using fleet4 for tuning enables the estimation of recruitment in the intermediate year (age 0 from last data year shifted to represent age 1 in the next year), an additional run in which fleet4 covers also year 2022 was performed. Its results were almost identical to the results of Run 4, but provided the estimation of recruitment in the intermediate year. Thus, that run was selected as the final run (Run 6 in the Figure below).

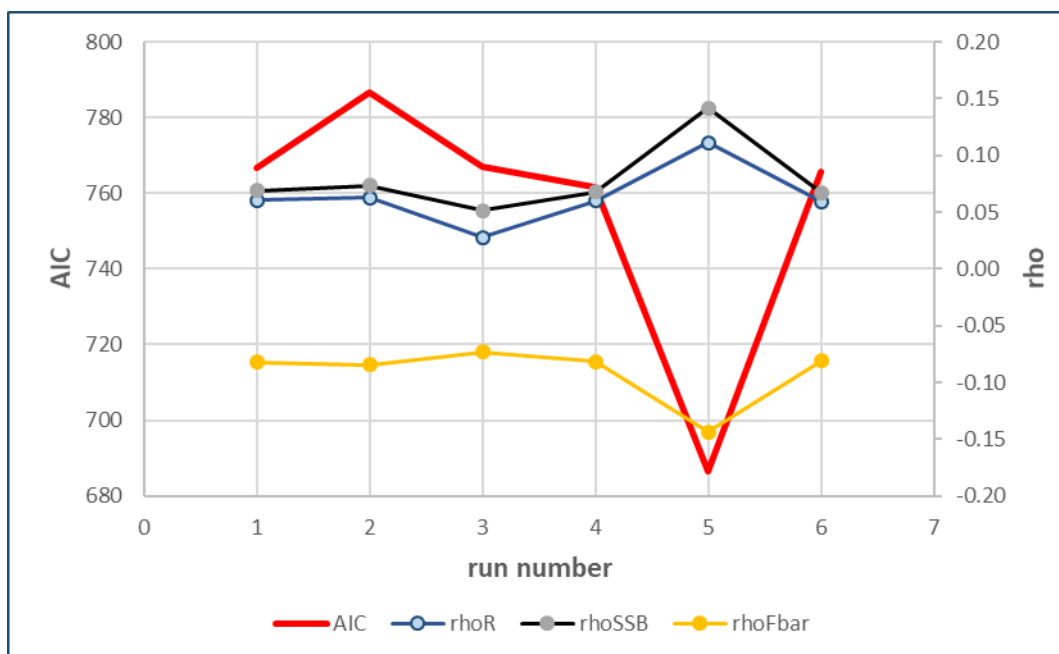


Figure 3.9. Values of AIC and Mohn’s rho obtained in selected SAM runs, Run 6 selected as final run.

### 3.3.4 Final model

The final parameterisation of SAM (Run 6):

- 4 tuning fleets were used.



- Catchability depended on year-class strength at age 1 for all fleets.
- Catchability plateau was set at age 6 (ages 6-8 assume the same  $q$ ).
- Recruitment was modelled as random walk.
- Covariance structure for each fleet was set as "ID" (independent).

The configuration file used for the final SAM assessment is shown in Table 3.6.

The final assessment results (using described above SAM options) are presented in Table 3.7 and Figure 3.10. More details of the final run are also presented in WD9: A proposal for FMSY reference points for Baltic sprat SDs 22-32

The distribution of residuals does not show a clear pattern except for age 1 in both age0 acoustic (fleet4) and October acoustic in sub-divisions 22-29 (fleet2). In these fleets, there is a tendency for negative residuals in the first years of the survey (Figure 3.11).

The leave-one-out analysis (Figure 3.12) shows little effect of excluding from tuning age 0 acoustic (fleet4) and October acoustic in sub-divisions 22-29 (fleet2). However, fleet4 is important as it provides the prediction of recruitment in the intermediate year.

Retrospective analysis shows some tendency to overestimate biomass and recruitment and underestimate fishing mortality (Figure 3.13). However, Mohn's rho values are acceptable and equal to 0.07, -0.08, and 0.06 for SSB,  $F_{bar}$ , and recruitment, respectively. The quality of the assessment in terms of retrospective deviations is higher than in the previous XSA assessment, where Mohn's rho values ranged from -0.15 to 0.15.

Up to the early 1990s, three assessment units (AUs) were used to assess sprat in the Baltic: sprat in subdivisions 22-25, sprat in subdivisions 26+28, and sprat in subdivisions 27, 29-32. Within the INSPIRE project (years 2014-2017) sprat assessments were performed by former assessment units using spatially different predation mortality from cod. These assessments used data up to 2015. The basic aim of such analysis was to check if the dynamics of sprat in the old AUs is similar, so that the merging in the 1990s of former assessment units into one stock of sprat in the Baltic (subdivisions 22-32) is still justified. It appeared that the sum of sprat's SSB by former AUs estimated with XSA was very similar to the SSB estimated by ICES (using data up to 2015) for sprat in the whole Baltic. Similarly, the average fishing mortality by AUs was very close to the  $F$  estimated by ICES for Baltic sprat. Thus, the conclusion from this analysis was that merging of the three AUs into one AU of sprat in the Baltic seems to be justified from an assessment's point of view.

**Table 3.6. Configuration file for final sprat assessment with SAM.**

```

#
$minAge
# The minimum age class in the assessment
1
$maxAge
# The maximum age class in the assessment
8
$maxAgePlusGroup
# Is last age group considered a plus group for each fleet (1 yes, or 0 no).
1 1 1 0
$keyLogFsta
# Coupling of the fishing mortality states (nomally only first row is used).
0 1 2 3 4 5 6 6
-1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1
$corFlag
# Correlation of fishing mortality across ages (0 independent, 1 compound symmetry, 2 AR(1), 3 separable AR(1).
2
$keyLogFpar
# Coupling of the survey catchability parameters (nomally first row is not used, as that is covered by fishing mortality).
-1 -1 -1 -1 -1 -1 -1 -1
0 1 2 3 4 5 5 5
6 7 8 9 10 11 11 11
12 13 14 15 16 17 17 17
18 -1 -1 -1 -1 -1 -1 -1
$keyQpow
# Density dependent catchability power parameters (if any).
-1 -1 -1 -1 -1 -1 -1 -1
0 -1 -1 -1 -1 -1 -1 -1
1 -1 -1 -1 -1 -1 -1 -1
2 -1 -1 -1 -1 -1 -1 -1
3 -1 -1 -1 -1 -1 -1 -1
$keyVarF
# Coupling of process variance parameters for log(F)-process (nomally only first row is used)
0 0 0 0 0 0 0 0
-1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1
$keyVarLogN
# Coupling of process variance parameters for log(N)-process
0 1 1 1 1 1 1 1
$keyVarObs
# Coupling of the variance parameters for the observations.
0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3
4 -1 -1 -1 -1 -1 -1 -1
$sobsCorStruct
# Covariance structure for each fleet ("ID" independent, "AR" AR(1), or "US" for unstructured). | Possible values are: "ID" "AR"
"US"
"ID" "ID" "ID" "ID" "ID"
$keyCorObs
# Coupling of correlation parameters can only be specified if the AR(1) structure is chosen above.
# NA's indicate where correlation parameters can be specified (-1 where they cannot).
#1-2 2-3 3-4 4-5 5-6 6-7 7-8
NA NA NA NA NA NA NA
NA NA NA NA NA NA NA
NA NA NA NA NA NA NA
NA NA NA NA NA NA NA
-1 -1 -1 -1 -1 -1 -1
$stockRecruitmentModelCode
# Stock recruitment code (0 for plain random walk, 1 for Ricker, 2 for Beverton-Holt, and 3 piece-wise constant).
0
$noScaledYears
# Number of years where catch scaling is applied.
0
$keyScaledYears
# A vector of the years where catch scaling is applied.

```

```

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

$FbarRange
# lowest and highest age included in Fbar
3 5
$KeyBiomassTreat
# To be defined only if a biomass survey is used (0 SSB index, 1 catch index, 2 FSB index, 3 total catch, 4 total landings and 5 TSB
index).
-1 -1 -1 -1 -1
$ObsLikelihoodFlag
# Option for observational likelihood | Possible values are: "LN" "ALN"
"LN" "LN" "LN" "LN" "LN"
$FixVarToWeight
# If weight attribute is supplied for observations this option sets the treatment (0 relative weight, 1 fix variance to weight).
0
$fracMixF
# The fraction of t(3) distribution used in logF increment distribution
0
$fracMixN
# The fraction of t(3) distribution used in logN increment distribution
0
$fracMixObs
# A vector with same length as number of fleets, where each element is the fraction of t(3) distribution used in the distribution of that
fleet
0 0 0 0 0
$constRecBreaks
# Vector of break years between which recruitment is at constant level. The break year is included in the left interval. (This option is
only used in combination with stock-recruitment code 3)
$predVarObsLink
# Coupling of parameters used in a prediction-variance link for observations.
-1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1
NA NA NA NA NA NA NA NA
    
```

**Table 3.7. Final estimates of stock size and fishing mortality from SAM**

R(age	1)	Low	High	SSB	Low	High	Fbar(3-5)	Low	High
1974	75943	51605	111757	1059	779	1440	0.311	0.224	0.432
1975	23195	15886	33865	842	619	1146	0.302	0.218	0.419
1976	148520	101643	217016	639	469	871	0.299	0.213	0.418
1977	76205	52118	111424	925	671	1274	0.288	0.205	0.404
1978	21536	14649	31661	696	505	959	0.249	0.176	0.35
1979	60736	41235	89458	380	274	527	0.225	0.159	0.317
1980	20022	13357	30012	282	203	391	0.242	0.171	0.342
1981	108496	72998	161256	283	205	391	0.189	0.133	0.268
1982	19914	13363	29676	347	248	486	0.194	0.138	0.273
1983	126242	85180	187099	347	252	478	0.127	0.09	0.179
1984	62969	43197	91791	561	409	769	0.148	0.108	0.203
1985	33882	23462	48929	615	461	821	0.161	0.119	0.216
1986	24387	16844	35307	550	422	718	0.179	0.135	0.238

R(age	1)	Low	High	SSB	Low	High	Fbar(3-5)	Low	High
1987	45404	31686	65061	474	369	609	0.226	0.171	0.299
1988	4652	3245	6669	467	361	605	0.233	0.178	0.305
1989	90457	63716	128422	451	356	571	0.239	0.185	0.309
1990	58199	41890	80857	749	588	956	0.195	0.154	0.247
1991	64795	48612	86363	945	770	1159	0.18	0.146	0.221
1992	86934	65404	115551	1008	836	1215	0.202	0.166	0.246
1993	87130	63372	119796	1276	1040	1566	0.218	0.178	0.267
1994	36669	27676	48584	1419	1181	1705	0.272	0.226	0.328
1995	181831	132239	250022	1128	946	1345	0.352	0.291	0.426
1996	164623	124262	218093	1622	1338	1965	0.398	0.333	0.474
1997	31641	23352	42873	1729	1427	2094	0.433	0.361	0.519
1998	162945	124823	212709	1168	993	1374	0.472	0.398	0.559
1999	27585	21200	35894	1269	1072	1500	0.434	0.367	0.512
2000	131605	101557	170543	1033	885	1206	0.424	0.359	0.502
2001	41637	32549	53263	1039	889	1213	0.424	0.36	0.499
2002	84957	66702	108209	900	779	1041	0.419	0.356	0.493
2003	126361	99031	161233	850	735	982	0.412	0.35	0.485
2004	258503	201475	331673	957	819	1118	0.463	0.392	0.546
2005	32262	25306	41130	1340	1118	1607	0.419	0.355	0.495
2006	117341	91467	150532	1048	883	1244	0.369	0.31	0.438
2007	144116	113410	183136	925	796	1076	0.444	0.376	0.524
2008	69430	54591	88303	1025	875	1199	0.437	0.371	0.514
2009	202962	158808	259393	966	831	1125	0.456	0.385	0.539
2010	47545	37380	60474	1040	869	1245	0.402	0.337	0.478
2011	95180	74632	121386	865	730	1025	0.326	0.273	0.39
2012	96308	75021	123636	865	737	1014	0.313	0.261	0.374
2013	84252	66284	107090	891	757	1049	0.355	0.297	0.424
2014	60340	47502	76648	772	656	909	0.373	0.312	0.446
2015	234577	180730	304469	787	670	924	0.344	0.287	0.414

R(age 1)	Low	High	SSB	Low	High	Fbar(3-5)	Low	High	
2016	62600	48051	81556	1186	971	1447	0.313	0.26	0.378
2017	72318	56610	92384	1191	995	1426	0.327	0.272	0.393
2018	98689	76801	126815	1016	863	1197	0.355	0.296	0.426
2019	65611	50744	84833	909	773	1070	0.407	0.338	0.489
2020	113551	88102	146352	888	753	1047	0.374	0.308	0.454
2021	103047	77144	137646	1089	890	1332	0.353	0.283	0.442
2022	37889	20296	70731						

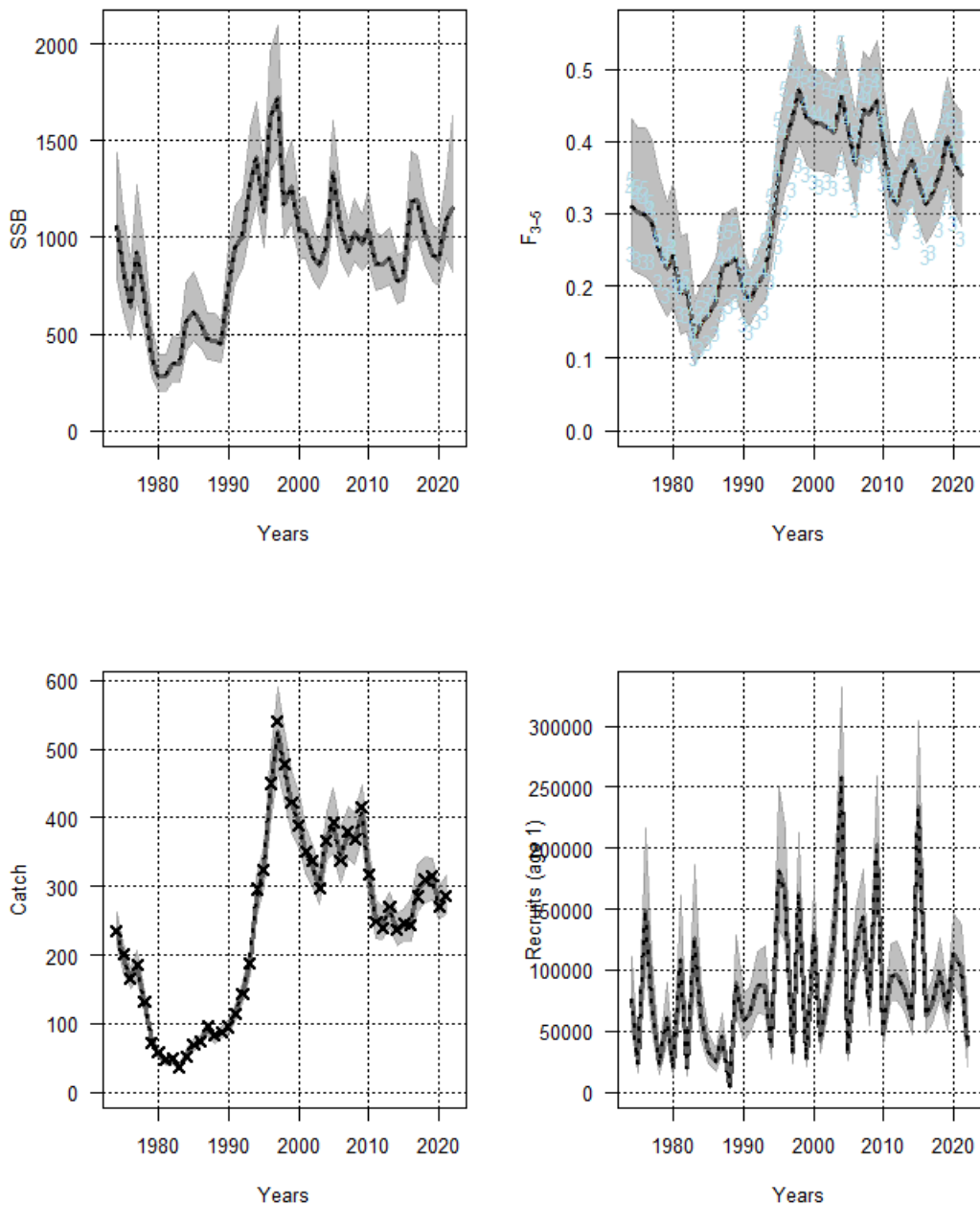


Figure 3.10. Summary of SAM assessment: SSB, F<sub>bar</sub>, catches, and recruitment.

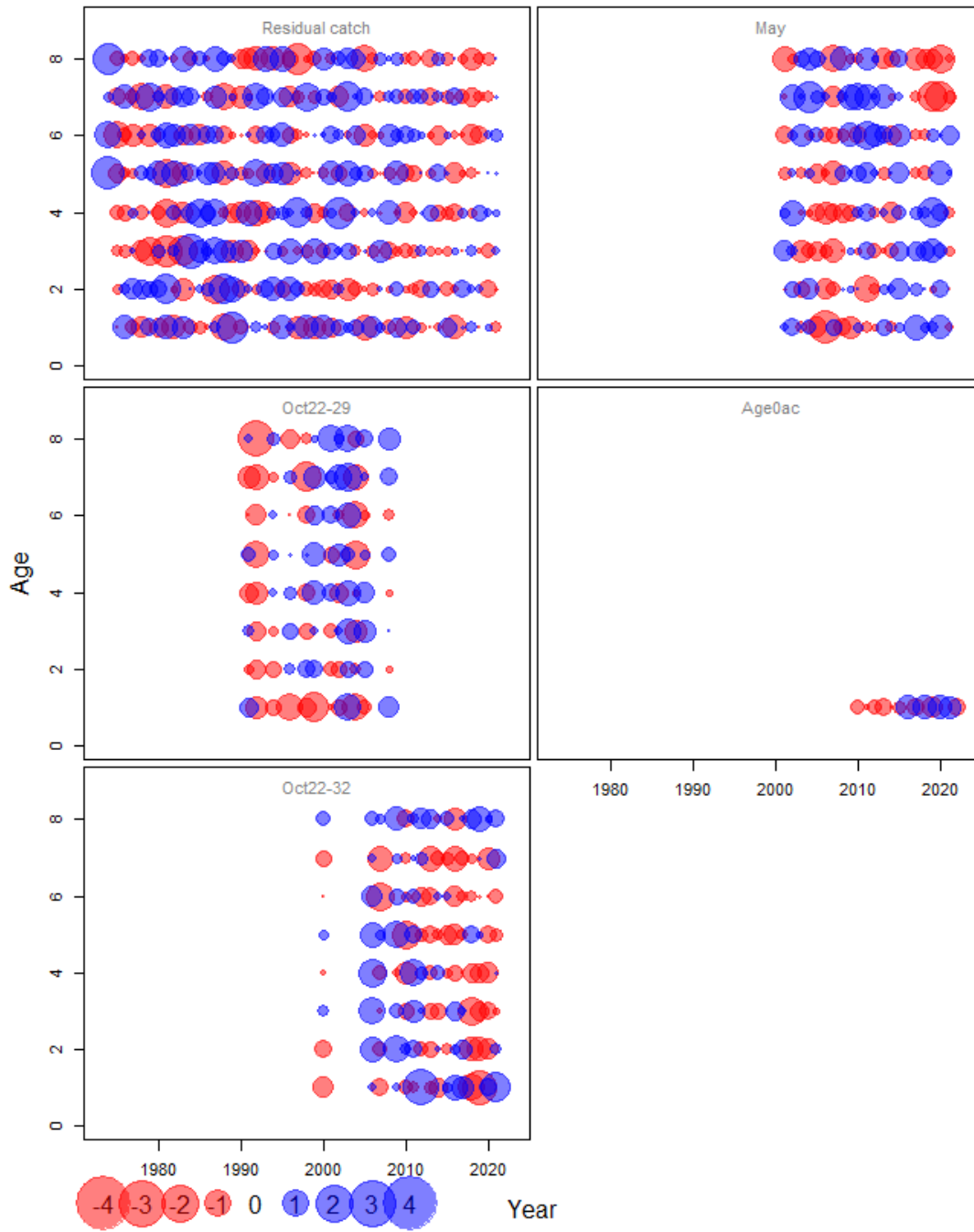


Figure 3.11. Distribution of SAM residuals in catches and by fleet.

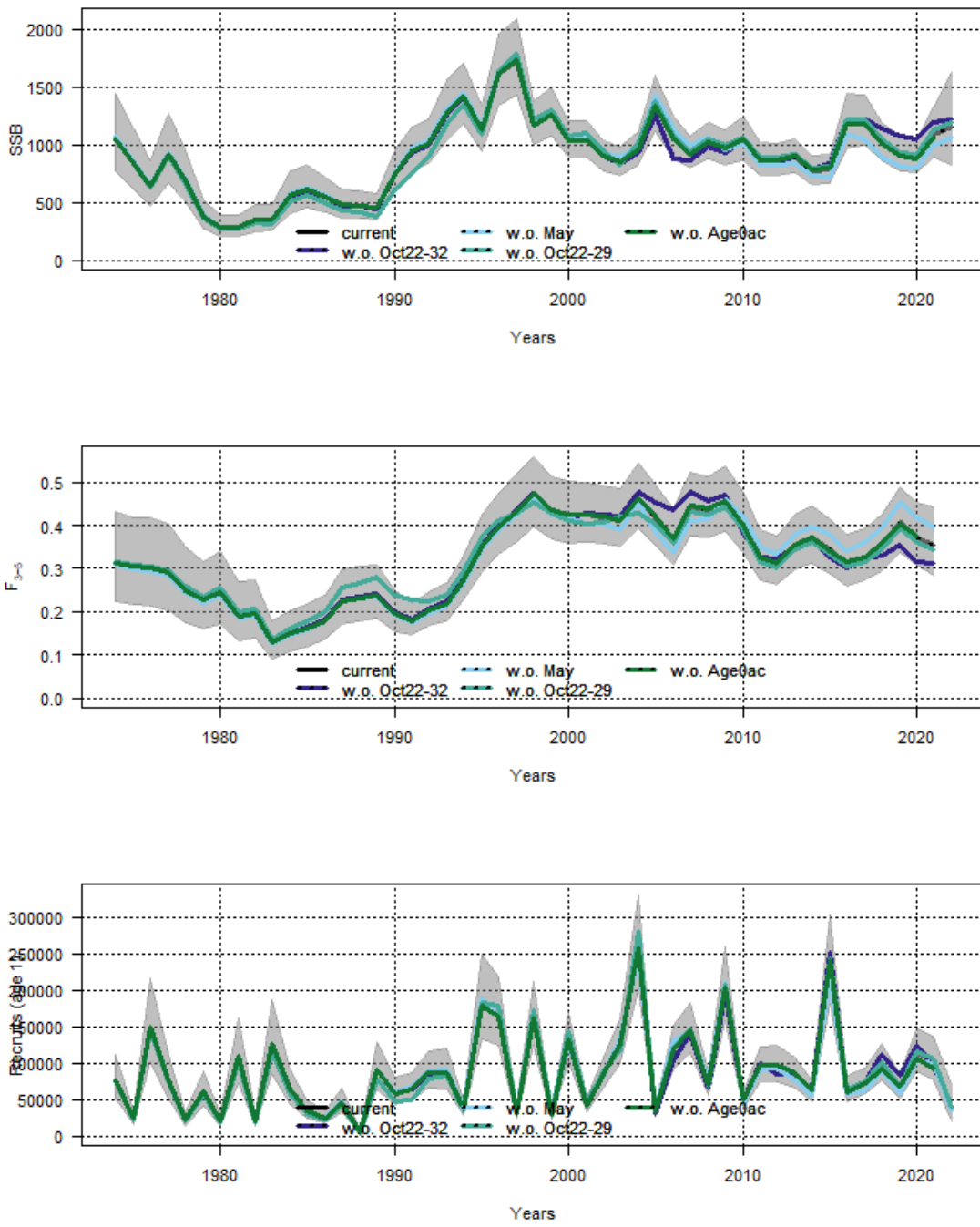


Figure 3.12. Leave-one-out analysis in SAM.

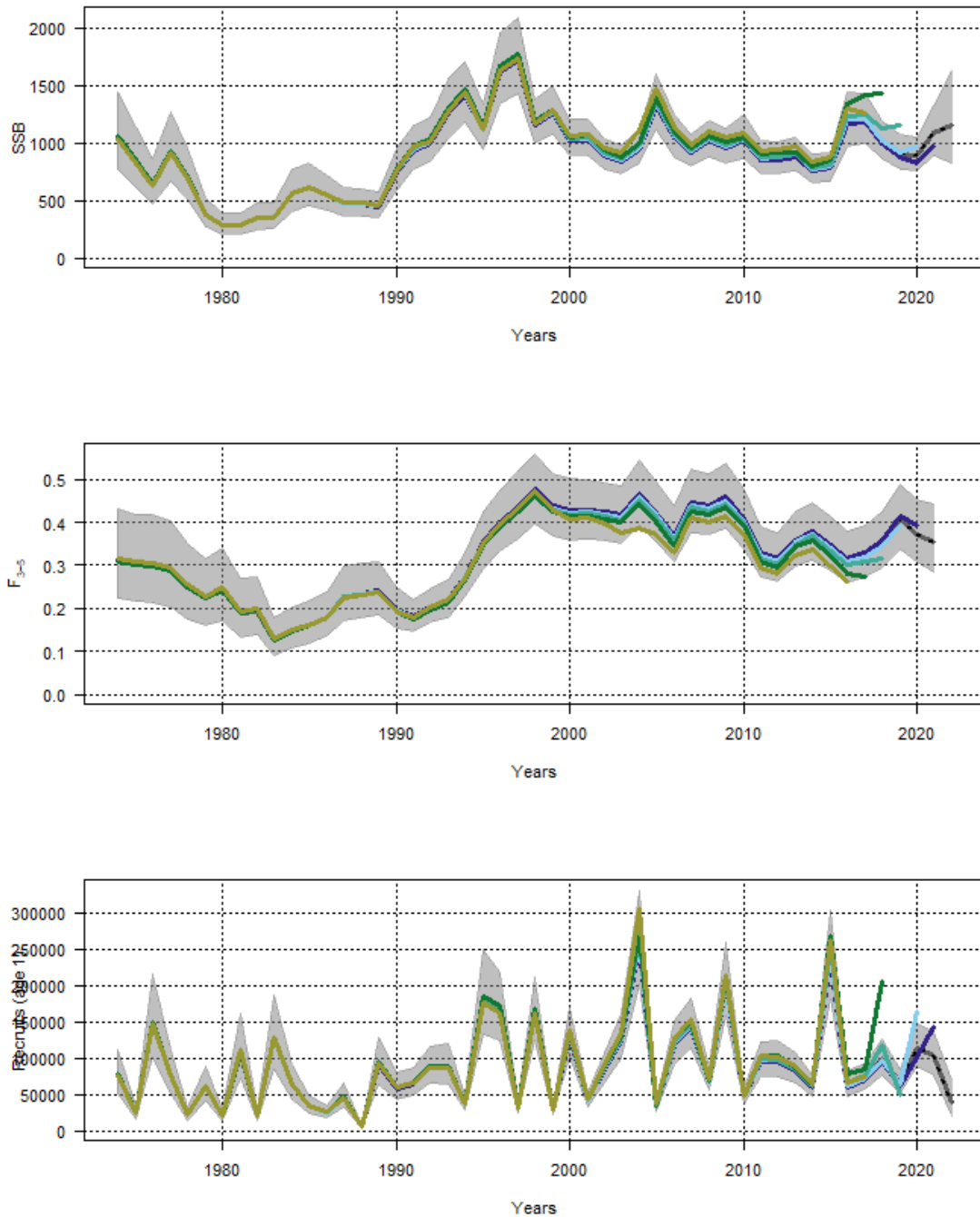


Figure 3.13. Retrospective estimates of biomass, fishing mortality, and recruitment in SAM assessments.

### 3.4 Short-term projections

Following the use of SAM for stock assessment, it was decided to use also stochastic SAM forecast for short-term predictions instead of the so far used deterministic MFDP software.

The input to short-term projections is based on standard assumptions, i.e. the three years averages of weight at age, maturity, fishing (selection), and natural mortalities. However, in the cases when a clear trend in quantities in question of these standard assumptions is observed, values from the most recent year could be used. Examples of such an approach may be a change in  $F_{\text{bar}}$  with selection which is based on a three years average but may be scaled to terminal year  $F_{\text{bar}}$  when a trend in  $F$  is observed; alternatively, only the most recent year for natural mortality could



be used in the terminal year when a trend in predation mortality is observed (e.g. change in cod biomass).

Initial stock size is obtained from the distribution of the stock size estimated in SAM at the start of the intermediate year.

Recruitment in the intermediate year is estimated from SAM. The recruitment in the next two years is obtained through sampling with replacement from estimates of SAM recruitments in the period from 1991 onwards.

## 3.5 MSY reference points

Details of the derivation of MSY and related reference points are presented in WD9 in Annex 5. The reference points were estimated with EqSim.

### 3.5.1 Settings

The settings used in the EqSim runs are summarised in Table 3.8. For biological data and fisheries selectivity, the most recent 10 and 5 years, respectively, were found to be most representative, and re-sampling from these years was used in EqSim. In all EqSim simulations, a combination of the Beverton and Holt function and segmented regression with a freely estimated inflection point was used as the stock-recruitment relationship (Figure 3.14).

**Table 3.8. Sprat. Summary of settings used in EqSim runs.**

Input	Details
Biological data	Mean of the last 10 years (2012-2021)
Fishing selectivity	Mean of the last 5 years (2017-2021)
sigmaSSB	0.10 (from SAM)
sigmaF	0.10 (from SAM)
$F_{cv}$	0.25 (ICES, 2015)
$F_{phi}$	0.30 (ICES, 2015)
Recruitment	No autocorrelation. Whole time-series used.
Stock-recruitment relationship	Combination of Beverton and Holt and segmented regression

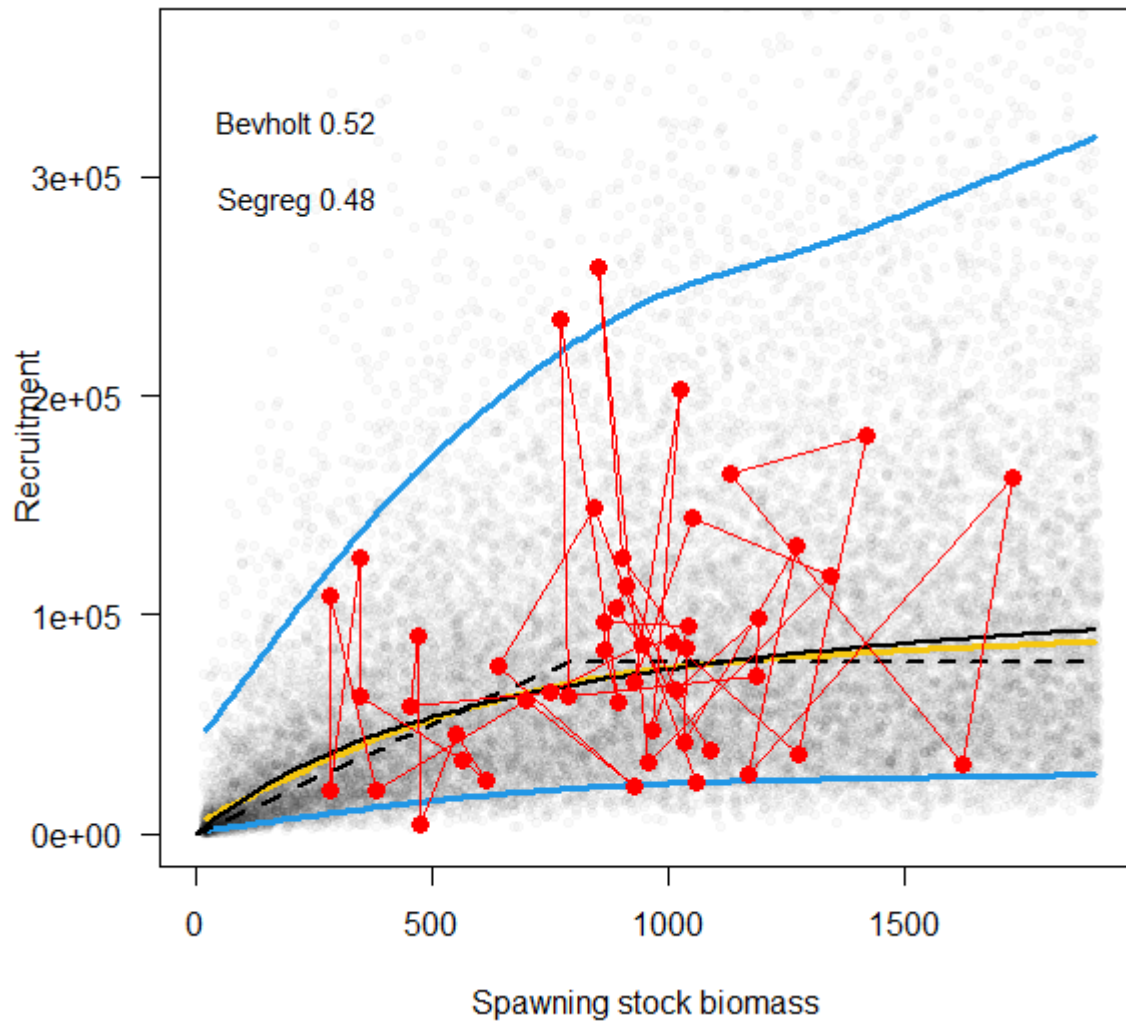


Figure 3.14. Stock-recruitment relationship of sprat. Spawning stock biomass is shown in k tonnes, recruitment in Mio numbers. The numbers represent the weights of the different functions in explaining the S-R pattern of sprat (dashed line: segmented regression (Segreg), solid line: Beverton and Holt (Bevholt)).

### 3.5.2 Estimation of $B_0$ , $B_{lim}$ and $B_{pa}$

Following approach from the last benchmark (ICES, 2013) and inter-benchmark assessment (ICES, 2020),  $B_{lim}$  was defined as the spawning stock biomass which produces 50% of the maximal recruitment. Maximal recruitment was estimated using the equations of Horbowy & Luzeńczyk (2012) and Horbowy & Hommik (2022) where equilibrium recruitment ( $R_{eq}$ ), yield ( $Y_{eq}$ ), and spawning stock biomass ( $B_{eq}$ ) at fishing mortality  $F$  may be derived as:

$$R_{eq}(F) = \frac{SPR(F) - a}{b * SPR(F)} \quad (1a)$$

$$Y_{eq}(F) = YPR(F) * R_{eq}(F) = YPR(F) \frac{SPR(F) - a}{b * SPR(F)} \quad (1b)$$

$$B_{eq}(F) = SPR(F) * R_{eq}(F) = \frac{SPR(F) - a}{b} \quad (1c)$$

The SPR and YPR denote stock-per-recruit, and yield-per-recruit, respectively;  $a$  and  $b$  are parameters of the B&H S-R relationship of the form  $R=B/(a+b*B)$ .

Above approach for estimation of equilibrium recruitment, biomass, and yield is similar to the one used by Albertsen & Trijoulet (2020), except that equations 1a-1c provide analytical estimates

of the quantities in question while Albertsen & Trijoulet (2020) used numerical solution to find these values.

The  $R_0$  was estimated at 105 billion and the biomass which produces 50%  $R_0$  is 459 kt (=  $B_{lim}$ ).

$B_{pa} = B_{lim} * e^{\sigma_{SSB} * 1.645} = 541\,071$  with  $\sigma_{SSB} = 0.1$  based on the last assessment year.

Three other options for setting  $B_{lim}$  were presented in WD9 in Annex 5 but the one presented above was finally selected.

### 3.5.3 Estimating $F_{MSY}$ , $F_{p05}$ and $F_{pa}$

To estimate the unconstrained  $F_{MSY}$ , the EqSim was run without the advice rule (i.e. no MSY  $B_{trigger}$ ), with assessment and advice error using the values  $(F_{cv}, F_{\phi}) = (0.25, 0.30)$  as suggested by WKMSYREF3 (ICES, 2015), and with a combination of Beverton-Holt and the segmented regression (Figure 3.14). When allowing the program to use the full range, and combinations of, bootstrap simulated a and b parameters in the S-R function, the results presented unrealistically high catches at low fishing mortalities. The resulting  $F_{MSY}$  values were therefore not considered reliable. The extreme values of the parameters were thus removed and only values within the 5<sup>th</sup> and 95<sup>th</sup> percentile kept.

The resulting unconstrained  $F_{MSY}$  obtained (median MSY for lanF) was  $F_{MSY} = 0.34$ .

To ensure consistency between the precautionary and the MSY frameworks,  $F_{MSY}$  is not allowed to be above  $F_{p,05}$ ; therefore, if the initial  $F_{MSY}$  value is above  $F_{p,05}$ ,  $F_{MSY}$  is reduced to  $F_{p,05}$ .  $F_{p,05}$  was calculated by running EqSim with assessment/advice error, with advice rule, and with a segmented regression with breaking point fixed at  $B_{pa}$  to ensure that the long-term risk of  $SSB < B_{lim}$  of any  $F$  used does not exceed 5% when applying the advice rule.  $F_{p,05}$  was estimated to be 0.35. Therefore, as explained above,  $F_{pa} = F_{p,05} = 0.35$ . Corresponding  $F_{MSY}$  ranges are shown in Table 3.9.

$F_{lim}$  was estimated as  $F$  corresponding to 50% probability for  $SSB > B_{lim}$ . This resulted in a  $F_{lim}$  of 0.58, which corresponds to  $F_{pa}$  of 0.49 ( $F_{pa} = F_{lim} * e^{-1.645 * 0.1}$ ).

**Table 3.9. Sprat.  $F_{MSY}$  ranges.**

Reference point	Value	Technical basis
$F_{MSYlower}$	0.26	$F_{MSYlower}$ (EqSim)
$F_{MSY}$	0.34	$F_{MSY}$ (EqSim)
$F_{MSYupper}$	0.35	$F_{MSYupper}$ (EqSim)

### 3.5.4 Estimating MSY $B_{trigger}$

MSY  $B_{trigger}$  is a lower bound of the SSB distribution when the stock is fished at  $F_{MSY}$  (ICES, 2021). As stated in the ICES technical guidelines, recent fishing mortality estimates need to be considered to set MSY  $B_{trigger}$  as for most stocks that lack data on fishing at  $F_{MSY}$ , MSY  $B_{trigger}$  is set at  $B_{pa}$ . Here, the stock has been fished above  $F_{MSY}$  (0.34) for the last 5 years. Following the ICES technical guidelines our MSY  $B_{trigger}$  will be equal to  $B_{pa}$ , MSY  $B_{trigger} = 541\,071$  tonnes.

### 3.5.5 Comparison with previous reference points

Present estimates of BRPs are quite similar to previous ones (12% increase in  $B_{lim}$  and 10% increase in  $F_{MSY}$ ). The difference in the estimates of  $F_{MSY-upper}$  and  $F_{p05}$  larger; both values are now estimated at 0.35, while they were previously estimated to be equal to 0.41.

## 3.6 Future research and roadmap

Further work on misreporting of herring and sprat and its effect on the assessment and management of both stock should be encouraged.

## 3.7 Progress Feco

Feco was generally discussed after presentation of its concept by Maciej Tomczak. However, it was no time at the meeting to develop environmental indicator which could be used in Feco estimates.

## 3.8 References

- Albertsen, C. M., & Trijoulet, V. 2020. Model-based estimates of reference points in an age-based state-space stock assessment model. *Fisheries Research*, 230, [105618]. <https://doi.org/10.1016/j.fishres.2020.105618>
- Horbowy, J., Luzeńczyk, A. 2012. The estimation and robustness of FMSY and alternative fishing mortality reference points associated with high long-term yield. *Can. J. Fish. Aquat. Sci.* 69: 1468–1480
- Horbowy, J., & Hommik, K. 2022. Analysis of Fmsy in light of life-history traits—Effects on its proxies and length-based indicators. *Fish and Fisheries*, 00, 1–17. <https://doi.org/10.1111/faf.12640>
- ICES. 1986. Report of the Working Group on Assessment of Pelagic Stocks in the Baltic. Copenhagen, 7-17 April 1986. C.M. 1986/Assess:20.
- ICES. 2013. Report of the Benchmark Workshop on Baltic Multispecies Assessments (WKBALT), 4–8 February 2013, Copenhagen, Denmark. ICES CM 2013/ACOM:43. 399 pp.
- ICES. 2022. Baltic Fisheries Assessment Working Group (WGBFAS). ICES Scientific Reports. 4:44. 659 pp. <http://doi.org/10.17895/ices.pub.19793014>
- ICES. 2023a. Working Group on Baltic International Fish Survey (WGBIFS; outputs from 2022 meeting). ICES Scientific Reports. 05:10. 529pp. <https://doi.org/10.17895/ices.pub.22068821>
- ICES. 2023b. Working Group on Multispecies Assessment Methods (WGSAM; outputs from 2022 meeting). ICES Scientific Reports. 5:12. 233 pp. <https://doi.org/10.17895/ices.pub.22087292>
- Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*. 158: 96–101.
- Santos, J. N. S., Araújo, F. G., and Silva, D. S. 2009. Length correction for early-juvenile Brazilian herring *Sardinella janeiro* (Eigenmann, 1894) after preservation in formalin, ethanol and freezing. *Neotropical Ichthyology*, 7: 87–92. Sociedade Brasileira de Ictiologia.
- Shepherd, J. G. 1999. Extended survivors analysis: an improved method for the analysis of catch-at-age data and abundance indices. *ICES Journal of Marine Science*, 56: 584–591.

## 4 Herring (*Clupea harengus*) in Subdivision 28.1 (Gulf of Riga)

Working documents related to this stock

- WD10\_Maturity-at-age analysis for Gulf of Riga herring\_f
- WD11\_Analysis of Gulf of Riga herring trap net tuning series\_f
- WD12\_GoR. Assessment model settings\_f
- WD13\_Reference points for GoR herring\_f

### 4.1 Issue list

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed / are these available / where should these come from?	Research/WG input needed	Main Person	Conclusions and outcomes
Life history parameters	Maturity-at-age. Time constant maturity has been used for the whole time-series.	Investigate changes in maturity by year.	Commercial sampling data from Q1, available from Estonia and Latvia.		Kristiina, Ivars	New time varying maturity starting from 1995.
Tuning series	Fisheries dependent trapnet fleet. Trapnet effort has been kept constant in the analysis since 2015, as there are problems reproducing the previous trapnet effort calculations.	Investigate into tarpnet effort changes. Is the constant effort since 2015 plausible or not. Try to reproduce the effort calculations.	Detailed effort data. Available from Estonian side for the last 4 years (2019-2021).		Kristiina	Based on Estonian data effort has not been constant for the last 4 years. We were not able to reproduce the old effort calculations. Trapnet fleet tuning series excluded from the assessment.
	Acoustics survey (GRAHS)	Recalculation of GRAHS survey index on StoX.	Data available in the ICES acoustics data-based starting from 2011. Older data not uploaded yet.	WGBIFS	Elor	Data upload from years 1999-2010 was not managed in time for the benchmark. Data has to be double checked, even the data that has already been uploaded to the database. Old survey index estimates were used in the assessment.
Assessment method	Assessment model. A change to SAM model framework.	Configuration of SAM model.	CANUM, WECA, maturity, mortality etc.	Anders Nielsen	Kristiina	Successful SAM configuration was achieved.

## 4.2 Stock ID and sub-stock structure

Gulf of Riga herring (GoR) is a separate population of Baltic herring (*Clupea harengus*) that is met in the Gulf of Riga (ICES Subdivision 28.1). It is a slow-growing herring with one of the smallest length- and weight-at-age in the Baltic and thus differs considerably from the neighbouring herring stock in the Baltic Proper (Subdivisions 25–28.2, 29 and 32) (Kornilovs, 1994; ICES, 2001). The differences in otolith structure serve as a basis for discrimination of Baltic herring populations (ICES, 2005; Ojaveer *et al.*, 1981; Raid *et al.*, 2005). The discrimination between the herring populations takes place during the aging process of sampled fish when fishes are aged and also assigned their population belonging on the basis of otolith structure. The stock does not migrate into the Baltic Proper; only minor part of the older herring leaves the gulf after spawning season in summer –autumn period but afterwards returns to the gulf. There is evidence, that the migrating fishes mainly stay close to the Irbe Strait region in Subdivision 28.2 and do not perform longer trips. The extent of this migration depends on the stock size and the feeding conditions in the Gulf of Riga. In 1970s and 1980s when the stock was on a low level the amount of migrating fishes was considered negligible. Since the beginning of 1990s when the stock size increased, also the number of migrating fishes increased and the catches of the Gulf of Riga herring outside the Gulf of Riga in Subdivision 28.2 are taken into account in the assessments.

## 4.3 Ecosystem drivers

The Gulf of Riga is a separate semi-enclosed ecosystem of the Baltic Sea characterized by low salinity of about 5 psu and separated from the Baltic Proper by a strong hydrological front in the Irben Strait. The gulf is shallow (deepest point is about 60 m) with a surface area of 16 000 km<sup>2</sup> and with the average volume of 424 km<sup>3</sup> which constitutes 3.9% of the total area and 2.1% of volume the BS (Berzinsh, 1995). That effects on the residence of marine species in the Gulf of Riga where herring is the dominant species. Both trawl fishery and coastal fisheries occur in the gulf. The trawl fishery in the gulf targets herring while coastal fishery targets also a number of fresh-water species. There is some bycatch of sprat in the periods when the sprat stock is on a high abundance level. There is also a lack of natural predators of herring in the gulf since cod are present in the Gulf of Riga only in these periods when the cod stock is on a very high level (last time in early 1980s) and the abundance of salmon and sea trout is relatively low.

The investigations of herring spawning grounds in 1980s showed that their overall spawning area has decreased compared with the situation in 1950s (Raid, 1990). That happened due to disappearance of demersal vegetation from larger depths as a result of increased eutrophication of the gulf that led to increased mortality of eggs. Since then, the status of the spawning grounds has not been investigated. Estonia has performed the mapping of herring spawning grounds in its waters of the Gulf of Riga in 2011. However, it could be stated that the pollution of the gulf has considerably decreased since the end of 1980s when changes in industry and agriculture took place and several sewage treatment plants were built.

The year-class strength of Gulf of Riga herring strongly depends on the severity of winter. It has been stated already in the 1960s that after mild winters rich year classes emerge (Rannak, 1971). After mild winters spawning starts earlier and the spawning activity is more evenly distributed over the spawning season, which results in lower mortality of eggs on the spawning grounds. Additionally, after mild winters the zooplankton is more abundant providing better feeding conditions for herring larvae. The relationships with average water temperature in April, when the spawning starts, and the abundance of *Copepoda* in May, when the hatching of larvae begins, were used to predict recruitment until 2006.

However, in the more recent RCT3 predictions the weight of zooplankton abundance in the prediction of recruitment has considerably decreased due to appearance of two very rich year classes. Zooplankton abundance in May in those years was only slightly above the average and thus these years stand out of line in the relationship between zooplankton abundance and year-class strength.

Therefore, during the ICES Workshop of Recruitment processes of herring in the Baltic Sea (ICES, 2007) other factors explaining the year-class strength were analysed. It was stated that the average water temperature of 0–20 m depth layer in May and the biomass of the copepod *Eurytemora affinis* have significant relationship with year-class strength of Gulf of Riga herring. Therefore, for prediction of 2006 year class at age 1 in 2007 we used new data mentioned above. The same procedure was used in since 2008.

In 2011 the analysis of factors determining year-class strength was performed and a paper at ICES Annual science conference in Gdańsk was presented (Putnis *et al.*, 2011). Two additional significant relationships were found for the herring year-class strength. It was shown that since 2000 the year-class strength strongly depends on the feeding conditions during the herring feeding season. The feeding conditions were characterized as the average Fulton's condition factor for ages 2–5. In 2000, 2002, and 2005 when very rich year classes appeared the Fulton's condition factors were among the highest in 2000–2010. Apparently in good feeding years the feeding competition between older herring and the young-of-the-year decreases and the latter have bigger chance to survive. A strong negative relationship between neighbouring year classes was also found. The very rich year classes were usually followed by poor or below average year classes. Since the one-year old herring does not spawn and starts feeding much earlier than the mature herring it strongly affects the amount of food for the young-of-the-year, especially in the end of spring- beginning of summer during the new generation is in larval stage. In 2012 the found relationships were tested in RCT3 but were found insignificant due to high variation ratio. Since 2012 the geometric mean over recent climate period with higher stock abundance and recruitment dynamic (1990-present) is used for incoming year-class abundance estimation.

## 4.4 Stock Assessment

### 4.4.1 Data

Below we outline fishery-dependent (landings), fishery-independent, and biological data that are used as input data in the assessment model.

There was no specific data call made for this benchmark. As only two countries are fishing on the Gulf of Riga herring stocks, in addition to uploading the data to InterCatch, catch-at-age information are exchanged directly between countries. These catch-at-age files include quarterly data by country of number-at-age and mean weight-at-age. The original catch-at-age exchange files were located, and these were available starting from 2003. Older catch-at-age and weight-at-age information was gathered from old WGBFAS reports (years 1977-2001). Original catch-at-age files were needed to retrieve information about age class 0.

#### 4.4.1.1 Landings and discards

The total reported landings by country used for this benchmark were kept the same as in the latest 2022 assessment conducted WGBFAS meeting. However, it was noted that the official catch statistics differ slightly in some years (beginning of timeseries) with the catch estimates calculated by multiplying the catch in numbers with weight-at-age. ( $CANUM * WECA$ ). To some extent this issue is related to the fact that catch numbers up to 1995 are rounded values, causing some discrepancies, however, in some years the differences are too big to be just caused by the

rounding values. For example, in 1978 the official statistics estimates catch to be 16 728 tonnes while canum\*weca gives a value of 14 584 tonnes, which is 12.8% lower compared to the official number. The comparison of catch estimates are shown in Table 4.1.

**Table 4.1. Gulf of Riga herring. Comparison of catch estimates from official statistics and calculation of canum\*weca.**

Year	Catch,t (canum*weca)	Catch, t (official sta- tistics)	Difference, t	Difference, %
1977	23274	24186	912	3.8%
1978	14584	16728	2144	12.8%
1979	17537	17142	-395	-2.3%
1980	14651	14998	347	2.3%
1981	16291	16769	478	2.8%
1982	13407	12777	-630	-4.9%
1983	16440	15541	-899	-5.8%
1984	15525	15843	318	2.0%
1985	15956	15575	-381	-2.4%
1986	16529	16927	398	2.3%
1987	12908	12884	-24	-0.2%
1988	18264	16791	-1473	-8.8%
1989	16415	16783	368	2.2%
1990	15272	14931	-341	-2.3%
1991	14744	14791	47	0.3%
1992	21800	20000	-1800	-9.0%
1993	22218	22200	-18	-0.1%
1994	24265	24300	35	0.1%
1995	32659	32656	-3	0.0%
1996	32582	32584	2	0.0%
1997	39861	39843	-18	0.0%
1998	29426	29443	17	0.1%
1999	31409	31403	-6	0.0%
2000	34064	34069	5	0.0%
2001	38799	38785	-14	0.0%
2002	39701	39701	0	0.0%



Year	Catch,t (canum*weca)	Catch, t (official sta- tistics)	Difference, t	Difference, %
2003	40803	40803	0	0.0%
2004	39116	39115	-1	0.0%
2005	32227	32225	-2	0.0%
2006	31232	31232	0	0.0%
2007	33741	33742	1	0.0%
2008	31144	31137	-7	0.0%
2009	32553	32554	1	0.0%
2010	30175	30174	-1	0.0%
2011	29647	29639	-8	0.0%
2012	28115	28115	0	0.0%
2013	26513	26511	-2	0.0%
2014	26253	26253	0	0.0%
2015	32856	32851	-5	0.0%
2016	30865	30865	0	0.0%
2017	28063	28058	-5	0.0%
2018	25746	25747	1	0.0%
2019	28923	28921	-2	0.0%
2020	33216	33215	-1	0.0%
2021	35760	35758	-2	0.0%

#### 4.4.1.2 Biological Information

##### 4.4.1.2.1 Catch in numbers

Previous CANUM consisted of ages 1-8+. The new CANUM also includes age 0. In addition, catch-at-age values were corrected for years 2003 and 2008. In 2003 the age 1 numbers were corrected as previously the age 1 numbers reflected both age 1 and age 0. For 2008 inconsistencies were found between previously show catch-at-age and values that were calculated based on the original data exchange files. Old and updated CANUM differences are shown in Figure 4.1. Updated catch in numbers values are in Table 4.2.

**Table 4.2. Gulf of Riga herring. Catch-at-age in numbers (10<sup>3</sup>).**

Year/age	0	1	2	3	4	5	6	7	8+
1977	800	69500	885100	141400	109700	35300	15700	16000	600
1978	7600	112000	97300	403900	39200	35900	9300	3200	5700
1979	15400	76700	176500	103800	342500	22100	19300	6800	5500
1980	18500	101000	125900	99600	55400	133100	10500	8600	2500
1981	10700	62500	172500	112000	83000	51400	71700	7400	3500
1982	1400	80000	96000	116900	68800	43000	29900	24500	3300
1983	3100	49700	225300	138300	77700	38900	23300	15500	9600
1984	1900	44000	152100	255100	96300	56700	32500	14700	11900
1985	4400	23200	283900	203900	121700	31800	23700	8000	6100
1986	1000	9200	106700	246900	110600	66500	19600	8000	5800
1987	1000	70000	49000	110000	205000	75000	32000	5000	2000
1988	1400	6000	197700	112700	112400	144600	38700	27800	5900
1989	15100	61100	47400	492700	143000	76300	53900	6500	5400
1990	12500	88100	83100	67100	263500	66800	27600	14600	4100
1991	18500	119500	234000	94500	40800	180500	40500	35400	40800
1992	12100	150300	339100	369300	91300	33200	157400	19000	47600
1993	8600	192200	381400	298100	224400	66800	19000	78800	26900
1994	11760	164230	288440	368870	263500	192700	46080	9410	56150
1995	18100	232400	316900	363000	426900	277200	170900	39300	51500
1996	31700	428800	450100	281400	247600	291000	183800	105600	57000
1997	31700	204200	930700	559700	345400	242800	186700	90600	61100
1998	19600	239360	282060	505410	274890	172470	114020	90230	67650
1999	31400	361890	446500	157050	316480	157200	83650	60670	81050
2000	49700	259030	552300	359430	123730	258070	83980	35120	53370
2001	38700	819480	461570	378160	261040	81170	120980	56040	70710
2002	29057	304160	1182680	360540	202120	118950	36310	48060	44940
2003	5930	591660	396178	922839	231178	107441	70509	19995	58637
2004	50863	166756	1342017	306214	505774	129160	64392	33204	73423
2005	44630	384871	205390	833206	213430	171555	55243	27450	28925

Year/age	0	1	2	3	4	5	6	7	8+
2006	70251	787870	600122	113606	467376	100900	70418	16470	20007
2007	28897	305069	1145972	441269	83886	305940	59687	33710	24165
2008	40183	583363	341051	703895	165817	22389	119082	13798	26776
2009	55660	274301	765448	200530	494726	107356	20478	100014	28994
2010	48129	469192	407892	515483	109991	275715	55632	7764	75734
2011	48443	88964	327256	391007	278589	170847	128611	31572	63420
2012	76397	458920	123970	276010	196090	245430	39330	90650	33980
2013	17708	435220	596630	95600	143650	86850	128500	21350	57920
2014	50932	76960	553760	443440	68530	115750	62060	80660	58830
2015	108856	277380	141080	575230	394950	68160	82500	63190	117450
2016	36183	467310	287890	110350	427240	291430	43770	50850	94760
2017	61159	291780	449000	219830	59410	251400	183300	24030	94910
2018	29515	357867	295664	329437	150533	46463	149032	88866	36412
2019	64518	174379	629505	255381	267814	117162	48007	116436	60657
2020	41046	623754	285022	512507	192367	158621	85216	23743	109093
2021	136985	314882	794199	268629	384044	148641	123598	49741	70121



Figure 4.1. Gulf of Riga herring. Updated catch in numbers (red) and previously used estimates (blue).

#### 4.4.1.2.2 Weight-at-age

The same WECA as used in 2022 assessment was applied, adding the weight for age group to the weight-at-age matrix. Weight-at-age in stock is assumed same as in catch (WEST=WECA). Updated weight-at-age values are in Table 4.3.

**Table 4.3. Gulf of Riga herring. Weight-at-age in catch and stock (kg).**

Year/age	0	1	2	3	4	5	6	7	8+
1977	0.00290	0.01320	0.01600	0.02270	0.02690	0.02950	0.03120	0.02940	0.05080
1978	0.00530	0.00980	0.01770	0.02190	0.02730	0.03110	0.03040	0.03810	0.05040
1979	0.00630	0.01220	0.01620	0.02340	0.02760	0.02980	0.03400	0.03680	0.03600
1980	0.00710	0.01450	0.02010	0.02410	0.03210	0.03930	0.04560	0.05330	0.07110
1981	0.00760	0.01210	0.02160	0.02880	0.03340	0.03900	0.04390	0.04990	0.05950
1982	0.00540	0.01410	0.02140	0.02870	0.03570	0.03720	0.04510	0.05030	0.06837
1983	0.00570	0.01380	0.01930	0.02760	0.03790	0.04160	0.05090	0.06100	0.09130
1984	0.00540	0.01000	0.01500	0.02150	0.02810	0.03430	0.03910	0.04910	0.05590
1985	0.00600	0.01290	0.01720	0.02080	0.02780	0.03580	0.04870	0.05310	0.06650
1986	0.00600	0.01260	0.01980	0.02560	0.03140	0.04020	0.04620	0.06390	0.07090
1987	0.00600	0.01010	0.01540	0.01970	0.02630	0.03030	0.03790	0.04310	0.09050
1988	0.00660	0.01170	0.01860	0.02100	0.02730	0.03680	0.04340	0.05860	0.07500
1989	0.00670	0.01200	0.01480	0.01660	0.01960	0.02300	0.03150	0.03820	0.03640
1990	0.01140	0.01460	0.01780	0.01980	0.02690	0.03060	0.03310	0.05220	0.05540
1991	0.00690	0.01190	0.01540	0.01780	0.01990	0.02140	0.02250	0.02690	0.03360
1992	0.00630	0.01120	0.01360	0.01770	0.02150	0.02360	0.02500	0.02640	0.03590
1993	0.00640	0.01250	0.01360	0.01610	0.02010	0.02470	0.02630	0.02750	0.03520
1994	0.00410	0.01120	0.01460	0.01620	0.01880	0.02150	0.02520	0.02630	0.03000
1995	0.00540	0.01040	0.01360	0.01640	0.01790	0.02090	0.02290	0.02630	0.02910
1996	0.00390	0.01050	0.01250	0.01570	0.01770	0.01890	0.02150	0.02350	0.02800
1997	0.00490	0.00970	0.01240	0.01490	0.01780	0.01910	0.01960	0.02120	0.02420
1998	0.00660	0.01010	0.01330	0.01690	0.01820	0.02030	0.02130	0.02250	0.02400
1999	0.00490	0.01310	0.01550	0.01890	0.02210	0.02310	0.02450	0.02650	0.02890
2000	0.00631	0.01250	0.01650	0.02010	0.02290	0.02540	0.02640	0.02820	0.02960
2001	0.00523	0.01020	0.01600	0.02050	0.02300	0.02450	0.02770	0.02830	0.03070
2002	0.00495	0.01000	0.01530	0.01930	0.02360	0.02500	0.02710	0.02800	0.03090

Year/age	0	1	2	3	4	5	6	7	8+
2003	0.00468	0.00758	0.01530	0.01995	0.02226	0.02476	0.02632	0.02678	0.02760
2004	0.00445	0.00863	0.01012	0.01651	0.02103	0.02422	0.02676	0.02709	0.03310
2005	0.00525	0.01198	0.01393	0.01583	0.01930	0.02411	0.02536	0.02871	0.03080
2006	0.00541	0.00857	0.01319	0.01776	0.01913	0.02284	0.02656	0.02752	0.02960
2007	0.00562	0.00891	0.01166	0.01544	0.02020	0.01957	0.02369	0.02715	0.02780
2008	0.00541	0.00976	0.01493	0.01728	0.02047	0.02389	0.02331	0.02845	0.03270
2009	0.00584	0.00916	0.01399	0.01755	0.01907	0.02177	0.02068	0.02441	0.02940
2010	0.00452	0.00913	0.01380	0.01685	0.01942	0.02089	0.02369	0.02307	0.02600
2011	0.00448	0.01232	0.01586	0.01838	0.02152	0.02377	0.02540	0.02568	0.02877
2012	0.00545	0.00940	0.01593	0.02026	0.02317	0.02581	0.02771	0.02994	0.03340
2013	0.00582	0.00965	0.01465	0.01966	0.02266	0.02572	0.02820	0.02952	0.03190
2014	0.00562	0.00981	0.01384	0.01760	0.02158	0.02356	0.02534	0.02709	0.03020
2015	0.00576	0.00892	0.01502	0.01822	0.02108	0.02297	0.02516	0.02723	0.02950
2016	0.00599	0.00864	0.01516	0.01810	0.02039	0.02227	0.02388	0.02596	0.02830
2017	0.00514	0.00866	0.01473	0.01852	0.02093	0.02251	0.02412	0.02481	0.02760
2018	0.00649	0.00965	0.01532	0.01909	0.02159	0.02298	0.02452	0.02561	0.02840
2019	0.00592	0.00871	0.01357	0.01809	0.02066	0.02320	0.02366	0.02477	0.02620
2020	0.00602	0.00899	0.01535	0.01890	0.02123	0.02310	0.02499	0.02473	0.02600
2021	0.00539	0.00862	0.01379	0.01775	0.01963	0.02148	0.02310	0.02470	0.02530

#### 4.4.1.2.3 Maturity

New maturity ogives for years 1995-2021 were calculated using maturity data collected from commercial trawls in months January-April. The maturity ogive was produced by modelling maturity as a binomial GLM with a logit link:

$$\text{logit}(M) = \log(M / 1 - M)$$

where  $M$  is the probability of being mature. In this instance, age and year, alongside their interactions were included in the full model. Year was treated as factor. Maturity ogives were produced as predictions from the fitted models. To reduce the effect of interannual variability, the raw estimated time-series was smoothed for age classes 1-4, as it was assumed that from age 5 all individuals are mature. The main difference between previously used maturity ogive and the new time-varying maturity ogive is that roughly 25% of age 1 individuals are mature while previously age 1 was considered not mature, and in the new maturity ogive slightly lower percentage of age 2 is considered mature. More details are found in the WD-1.

There is no data available to derive maturity estimates from 1977-1994, hence for this period it we use the previously agreed maturity ogive. New smoothed maturity ogive estimates for period 1995-2021 and shown in Table 4 and Figure 2 together with the previously agreed maturity ogive for time period 1977-1994.

**Table 4.4. Gulf of Riga herring. Maturity ogive.**

Year/age	0	1	2	3	4	5	6	7	8+
1977-1994	0	0	0.93	0.98	0.98	1.0	1.0	1.0	1.0
1995	0	0.254	0.706	0.941	0.991	1.0	1.0	1.0	1.0
1996	0	0.251	0.702	0.939	0.990	1.0	1.0	1.0	1.0
1997	0	0.249	0.698	0.938	0.990	1.0	1.0	1.0	1.0
1998	0	0.246	0.694	0.936	0.989	1.0	1.0	1.0	1.0
1999	0	0.244	0.690	0.934	0.988	1.0	1.0	1.0	1.0
2000	0	0.242	0.686	0.932	0.987	1.0	1.0	1.0	1.0
2001	0	0.240	0.681	0.930	0.986	1.0	1.0	1.0	1.0
2002	0	0.239	0.676	0.927	0.985	1.0	1.0	1.0	1.0
2003	0	0.238	0.670	0.924	0.984	1.0	1.0	1.0	1.0
2004	0	0.237	0.664	0.921	0.983	1.0	1.0	1.0	1.0
2005	0	0.237	0.658	0.918	0.982	1.0	1.0	1.0	1.0
2006	0	0.237	0.651	0.915	0.981	1.0	1.0	1.0	1.0
2007	0	0.237	0.645	0.911	0.979	1.0	1.0	1.0	1.0
2008	0	0.238	0.640	0.908	0.978	1.0	1.0	1.0	1.0
2009	0	0.239	0.635	0.905	0.977	1.0	1.0	1.0	1.0
2010	0	0.241	0.631	0.902	0.975	1.0	1.0	1.0	1.0
2011	0	0.245	0.628	0.899	0.973	1.0	1.0	1.0	1.0
2012	0	0.250	0.626	0.896	0.972	1.0	1.0	1.0	1.0
2013	0	0.257	0.626	0.893	0.970	1.0	1.0	1.0	1.0
2014	0	0.265	0.626	0.890	0.968	1.0	1.0	1.0	1.0
2015	0	0.274	0.627	0.886	0.965	1.0	1.0	1.0	1.0
2016	0	0.284	0.629	0.883	0.963	1.0	1.0	1.0	1.0
2017	0	0.295	0.632	0.881	0.962	1.0	1.0	1.0	1.0
2018	0	0.306	0.636	0.879	0.960	1.0	1.0	1.0	1.0
2019	0	0.317	0.639	0.877	0.959	1.0	1.0	1.0	1.0

Year/age	0	1	2	3	4	5	6	7	8+
2020	0	0.328	0.643	0.875	0.958	1.0	1.0	1.0	1.0
2021	0	0.338	0.647	0.873	0.956	1.0	1.0	1.0	1.0

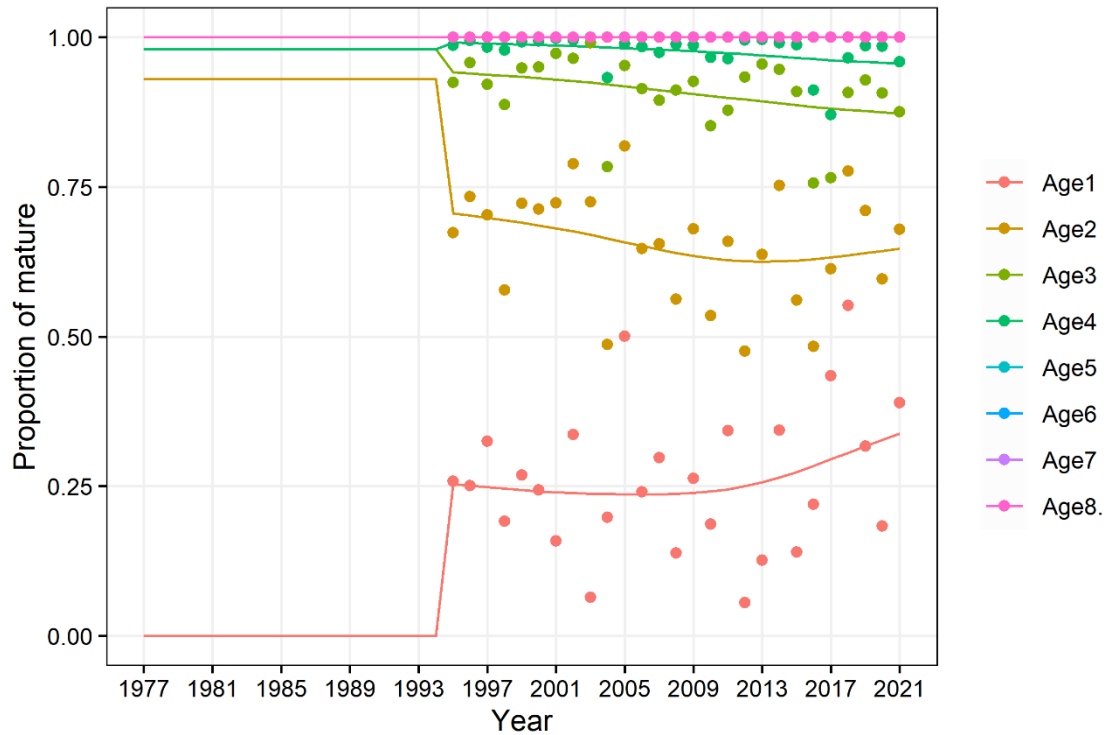


Figure 4.2. Gulf of Riga herring. Maturity ogive estimates. Dots are the raw estimates for period 1995-2021. Lines correspond to previous estimates (period 1977-1994) and smoothed values (1995-2021).

#### 4.4.1.2.4 Natural Mortality

There is no new information available about the natural mortality values. The natural mortality value was taken to be the same as that used previously: 0.20. Constant natural mortality  $M = 0.20$  is used for all the years except for the period 1979–1983 when a value of  $M = 0.25$  is used due to presence of cod in the Gulf of Riga.

#### 4.4.1.3 Tuning indices

Based on the results conducted before the benchmark data evaluation meeting (detailed description in WD-2) it was decided that the new assessment will use only one tuning series, the fisheries independent Estonian-Latvia joint hydro-acoustic survey (GRAHS), which includes ages 1-8+.

The reasons of excluding the previously used commercial trapnet tuning series are the following (more details in WD-2):

1. We are not able to reproduce the “effort” time-series in the trapnet tuning series
2. The “effort” in trapnet tuning series has been constant since 2015, however analysis showed that this assumption is not correct
3. “Effort” showed in the trapnet tuning series corresponds to number of trapnets directed towards Gulf of Riga herring fishing. This number however does not show the true effort of a passive gear (number of gears \* days-at-sea). In addition, even if an effort estimates as gear\*DAS would be available, it wouldn’t still have a direct link to changes in Gulf of

Riga herring stock, as the trapnet fishery also targets Central Baltic herring, and is even more aimed towards those fish. Therefore, it has been shown that in years when CBH is low in trapnet catches the effort (gear\*DAS) increases, which is unrelated to the GoR stock status.

#### 4.4.1.3.1 Hydroacoustic tuning series

The group explored the possibilities to recalculate the acoustics abundance estimates using StoX. However, as the whole time-series was not made available in time for the benchmark, the recalculation of the abundance estimates in StoX was not possible for the whole time-series and previously used acoustics abundance estimates were applied.

The latest acoustics abundance estimates were validated by WGBIFS. This led to minor changes in numbers for years 2010, 2020 and 2021. In addition, previously age group 8 was used a true age, however this was recalculated to be considered as a plusgroup (8+) to accommodate SAM settings. When the oldest age in catch and tuning index are the same, and in catch it is a plusgroup then it also has to be a plusgroup in the tuning index. The updated acoustics abundance estimates are in Table 4.5, and the comparison between updated abundance estimated with previous one is shown in Figure 4.3.

**Table 4.5. Gulf of Riga herring. Hydroacoustic survey abundance estimates.**

Year/age	1	2	3	4	5	6	7	8+	
1999	1	5292	4363	1343	1165	457	319	208	98
2000	1	4486	4012	1791	609	682	336	151	243
2001	1	7567	2004	1447	767	206	296	56	173
2002	1	3998	5994	1068	526	221	87	165	128
2003	1	12441	1621	2251	411	263	269	46	193
2004	1	3177	10694	675	1352	218	195	94	137
2005	1	8190	1564	4532	337	691	92	75	83
2006	1	12082	1986	213	937	112	223	36	49
2007	1	1478	3662	1265	143	968	116	103	39
2008	1	9231	2109	4398	816	134	353	6	23
2009	1	6422	4703	870	1713	284	28	223	44
2010	1	5077	2311	1730	244	593	107	12	50
2011	1	3162	5289	2503	2949	597	865	163	162
2012	1	5957	758	1537	774	1035	374	308	193
2013	1	9435	5552	592	1240	479	827	187	427
2014	1	1109	3832	2237	276	570	443	466	370
2015	1	3221	539	1899	1110	255	346	181	325
2016	1	4542	1081	504	1375	690	152	113	103



Year/age		1	2	3	4	5	6	7	8+
2017	1	3231	3442	874	402	1632	982	137	752
2018	1	11216	4529	3607	776	338	1439	755	381
2019	1	4912	7007	2237	1335	475	228	681	265
2020	1	9947	2659	3641	1234	1131	403	201	805
2021	1	6171	4891	1054	2161	815	670	257	405



Figure 4.3. Gulf of Riga herring acoustics abundance estimate comparison between updated estimates (red) and previously used (blue).

#### 4.4.1.4 Summary of data input to assessment model

Table 4.6. Gulf of Riga herring. Summary of data inputs to SAM.

Dataset	Year	Ages	Details
Catch numbers	1977-2021	0-8+	Catch-at-age in numbers.
Survey: acoustics	1999-2021	1-8+	Acoustics survey abundance estimates. Survey conducted in end of July/beginning of August
Catch weights	1977-2021	0-8+	Mean weight-at-age in the total catch.
Stock weights	1977-2021	0-8+	= catch weights
Natural mortality	1977-2021	0-8+	Time and age invariant, $M=0.2$ , except in years 1979-1983 $M=0.25$ due to presence of cod in GoR.
Maturity	1977-2021	0-8+	Proportion mature. Commercial-trawl derived maturity ogives 1995-2021, smoothed and time-varying. Non time-varying for 1977-1994.

Proportion of natural mortality before spawning	1977-2021	0-8+	0.1 for years 1977-1998, 0.2 from 1999 and onwards.
Proportion of fishing mortality before spawning	1977-2021	0-8+	0.3 for all years and ages.

## 4.4.2 Assessment model

The current assessment model used for Gulf of Riga herring is XSA. The primary motivation for a new assessment model is that the XSA software is old and not maintained. The XSA model has been converted to FLR, however active maintaining is still a problem.

It is proposed here that the state-space assessment model (SAM; Nielsen and Berg, 2014) should be considered. SAM has been run alongside XSA as an exploratory assessment since 2018 at the Working Group of Baltic Sea Fisheries Stock Assessment (WGBFAS).

### 4.4.2.1 SAM model configuration

The input data to the SAM model is summarised in Table 6. The base settings for the model were set as the default configuration of SAM model. One addition had to be included for the model convergence, the process variance parameters for  $\log(F)$  process were decoupled for ages 0, 1 and 2+ (baserun called *GoR\_BP\_base*). Different steps taken to reach to the final model settings are shown in Table 4.7. The final model settings were chosen based on the model AIC values and one-step-ahead residual visual inspection. Final configuration settings used within SAM are listed in Table 4.8 (settings not listed are set to default values, the whole configuration file is listed in section 4.10).

Comparison between the one-step-ahead residuals (OSA residuals) of the base run (*GoR\_BP\_base*) and final run (*GoR\_BP\_v2.2.3qF\_s*) are shown in Figure 4.4. With different configuration settings following issues were resolved:

1. Clear year effect in survey OSA residual plot
2. Larger OSA residual values in the beginning of timeseries.

Summary overview of final SAM assessment model output is shown in Figure 4.5. The final model is available in [www.stockassessment.org](http://www.stockassessment.org) under name "GoR\_BP\_v2.2.3qF\_s". The full description of the process by which final SAM assessment configuration was decided can be found in WD-3.

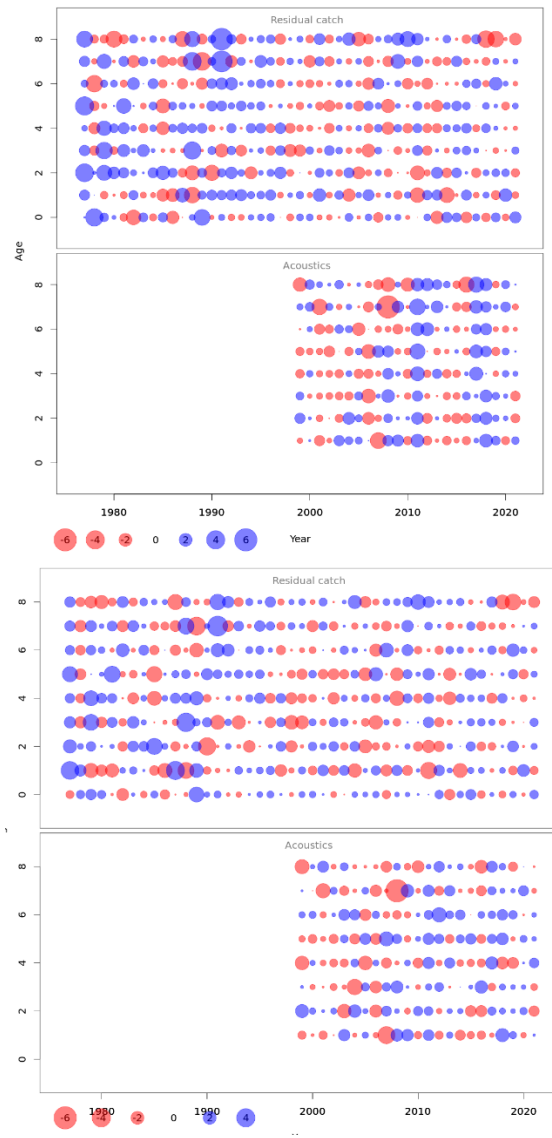


Figure 4.4. Gulf of Riga herring. Comparison of OSA residuals between base run (*GoR\_BP\_base*) and final run (*GoR\_BP\_v2.2.3qF\_s*).

**Table 4.7. Gulf of Riga herring. SAM model different runs and configuration settings.**

Model run	Parameter	setting	AIC	par
GoR_BP_base	Coupling of F state process for each age	0 1 2 3 4 5 6 7 7 -1 -1 -1 -1 -1 -1 -1 -1 -1	619.92	17
	\$keyLogFsta			
	Correlation of F across ages	2		
	\$corFlag			
	Coupling of the survey catchability	-1 -1 -1 -1 -1 -1 -1 -1 -1		
	\$keyLogFpar	-1 0 1 2 3 4 5 6 7		
	Coupling of process variance params for log(F) process	0 1 2 2 2 2 2 2 2 -1 -1 -1 -1 -1 -1 -1 -1 -1		
	\$keyVarF			
	Coupling of R and survival process variance	0 1 2 2 2 2 2 2		
	\$keyVarLogN			
GoR_BP_vol1	Coupling of the variance param. for the observations	0 0 0 0 0 0 0 0 0 -1 1 1 1 1 1 1 1 1	602.39	19
	\$keyVarObs			
	Covariance structure	ID ID		
	\$obsCorStruct			
GoR_BP_v2	\$stockRecruitmentModelCode	0 (plain random walk)	550.0	21
	\$fbarRange	3-7		
	\$matureModel	0 (MO is used as known)		
GoR_BP_v2.2	\$keyVarObs	0 0 0 0 0 0 0 0 0 -1 1 2 2 2 2 2 2 3	532.9	24
	\$keyVarF	0 1 2 2 2 3 4 5 5 -1 -1 -1 -1 -1 -1 -1 -1 -1		
GoR_BP_v2.2.1q	\$obsCorStruct	ID AR	543.27	17
	\$keyCorObs	NA NA NA NA NA NA NA NA -1 0 0 0 0 1 1 1		
GoR_BP_v2.2.2	\$keyLogFpar	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 0 0 0 0 0 0 0 0	536.23	20
	\$keyLogFsta	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 0 0 1 1 1 2 3 3		
GoR_BP_v2.2.3q	\$keyLogFpar	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 0 1 2 2 2 2 2 2	532.43	19

GoR_BP_v2.2.3qF_s	\$Fbar	2 6	558.27	19
	\$KeyVarLogN	0 1 1 1 1 1 1 1 1		

**Table 4.8. Gulf of Riga herring. Settings used for final SAM assessment model (GoR\_BP\_v2.2.3qF\_s). Default values were used for settings not listed.**

Configuration setting	Details
Assessment age range	0-8+
Is maximum considered a plus group	1 1
Coupling of F state process for each age	0 1 2 3 4 5 6 7 7
Correlation of F across ages	2 (i.e., AR(1))
Coupling of the survey catchability parameters	-1 0 1 2 2 2 2 2 2
Coupling of process variance parameters for log(F) process	0 1 2 2 2 3 4 5 5
Coupling of R and survival process variance	0 1 1 1 1 1 1 1 1
Coupling of the variance parameters for the observation	0 0 0 0 0 0 0 0 0 -1 1 2 2 2 2 2 2 3
Covariance structure	ID AR
Coupling of the correlation parameters for the AR(1) structure	NA NA NA NA NA NA NA NA -1 0 0 0 0 1 1 1
F <sub>bar</sub> range	2-6

The retrospective analysis shows that the final model assessment run has still same retrospective pattern as was observed previously, underestimating SSB and overestimating F, however the respective Mohn’s rho values are smaller compared to estimates based on XSA assessment (Table 4.9).

**Table 4.9. Gulf of Riga herring. Retrospective analysis. Mohn’s rho estimates based on final SAM model and XSA assessment (WGBFAS 2022).**

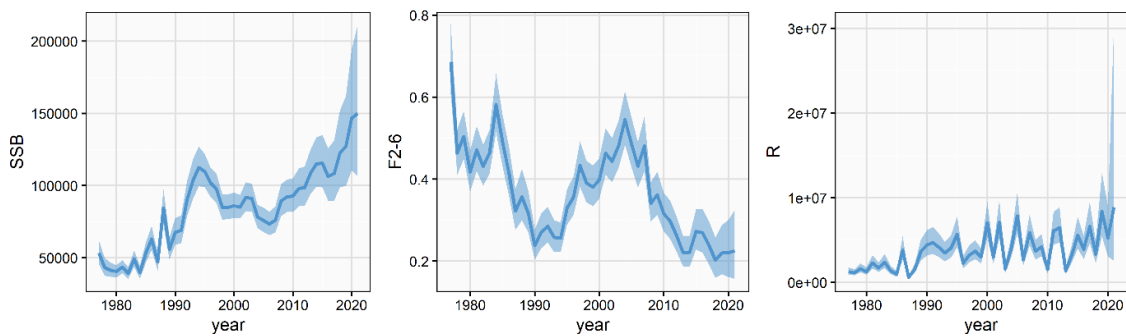
Mohn’s Rho	SAM (benchmark 2023)	XSA (WGBFAS 2022)
SSB	-8%	-15%
F <sub>bar</sub>	11%	26%
Recruitment	-22%	-12%

A comparison of the final model results to XSA assessment estimates from WGBFAS 2022 (ICES, 2022) is given in Figure 4.6. The overall trends between the two assessment results are mostly the same. The scale difference in SSB estimates are partially due to the fact that updated maturity ogive was used in SAM for time period 1995-2021, but even without the difference in maturity ogive (time period 1977-1994) SAM assessment estimates slightly lower SSB compared to XSA

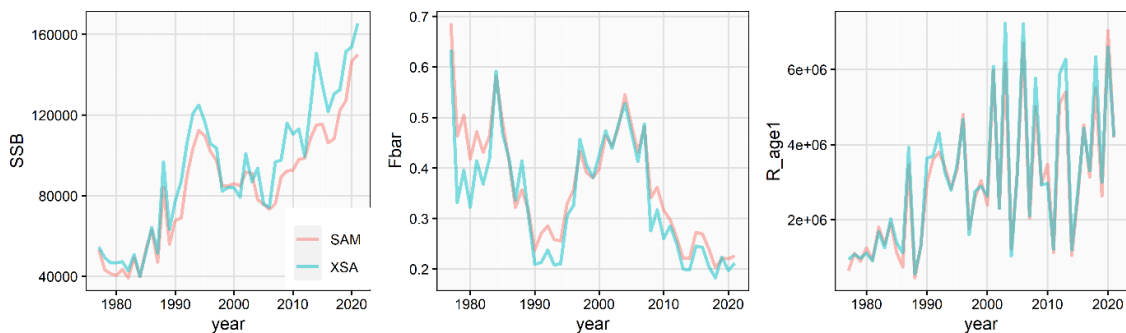
assessment. Inclusion of the new maturity ogive resulted in smoother SSB estimates as individual cohorts now become mature over several years as opposed to single year when they turn to age 2 (93%) as indicated in by the old maturity ogives.

One of the main differences in the input data between SAM and XSA is that age 0 is included in the catch matrix, which means that recruitment in SAM model happen at age 0 while in XSA at age 1. For comparison we have plotted (Figure 4.6) the age 1 estimates from SAM to compare with XSA output. Similarly, to SSB, SAM estimates for age 1 are in overall slightly lower compared to XSA.

Another update compared to XSA assessment was the revision of ages included in the  $F_{bar}$ . Previously the  $F_{bar}$  included ages 3-7, while closer inspection of the catch distribution by age and catch-curve analyses indicated that more suitable  $F_{bar}$  should already include age 2. In Figure 4.7 catch-curve analysis by cohorts are shown. While in the beginning of time-series starting the  $F_{bar}$  from age 3 seemed suitable, nowadays the highest catches (in numbers) are at age 2. The upper bound in  $F_{bar}$  was set to age 6. The corresponding  $F_{bar}$  age range (2-6) constitutes on average 80% of the catches. In Figure 4.6 the XSA assessment  $F_{bar}$  was adjusted to ages 2-6 to be comparable with SAM assessment output.



**Figure 4.5.** Gulf of Riga herring. SAM assessment model output. SSB (left),  $F_{bar}$  (middle) and recruitment (right).



**Figure 4.6.** Gulf of Riga herring. Comparison of SAM assessment model and XSA assessment model output (SSB – left,  $F_{bar}$  – middle,  $R_{age1}$  – right). Note that the  $F_{bar}$  for XSA assessment was updated to ages 2-6 to be compatible with SAM assessment  $F_{bar}$ . In addition, recruitment happens in SAM at age 0, while for comparison we have plotted age 1.

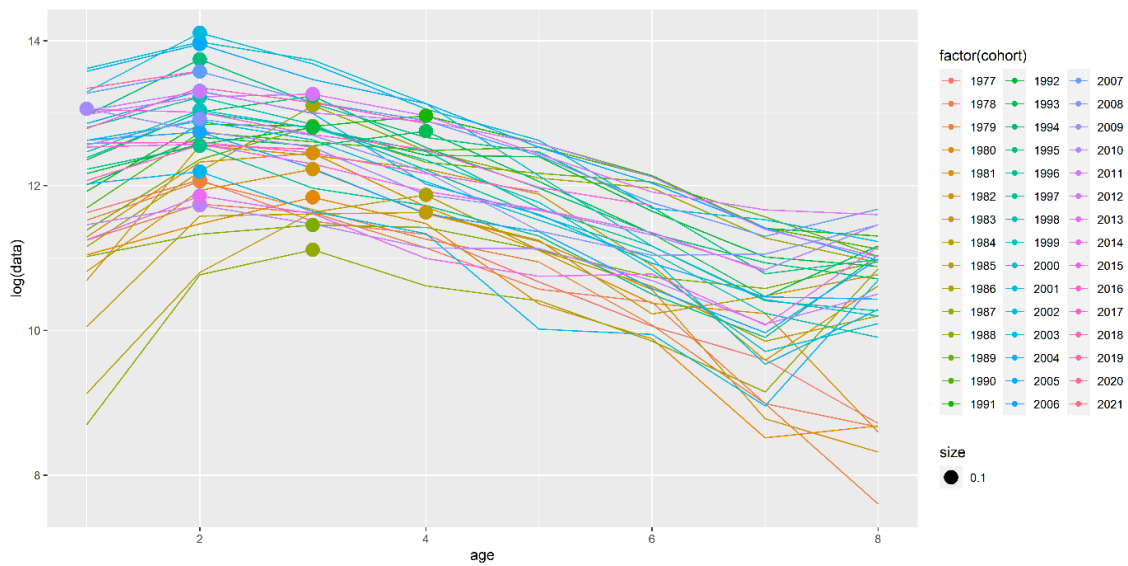


Figure 4.7. Gulf of Riga herring. Catch curve analysis. Different colours represent cohort years, and the points show the highest abundance by age in catch in a cohort.

### 4.5 Short-term projections

As the assessment model was changed to SAM the forecasting software will also be changed to SAM to ensure consistency between the assessment and forecast projection. SAM allows for a stochastic projection whereas MFDP only provides a deterministic projection.

The new settings and input to the SAM forecast are listed in Table 10. Majority of these inputs are the same as used in the current procedure.

Table 4.10. Gulf of Riga herring. Forecast settings for SAM projection.

Input	Details
Initial stock size	Simulated from the estimated distribution at the start of the intermediate year (including covariances)
Maturity	Mean of final 3 data years
Natural mortality	Constant value of 0.2
F and M before spawning	0.2 and 0.3, respectively
Mean weights-at-age in the catch	Mean of final 3 data years
Mean weights-at-age in the stock	Mean of final 3 data years
Exploitation pattern	Mean of final 3 data years, scaled by $F_{bar}$ (2-6) to the level of the last year in the case of obvious trend
Intermediate year assumption	Decided each year at WG
Recruitment assumption	Recruitment for the intermediate year onwards is sampled, with replacement, from 1990 to the final year -1 of catch data

## 4.6 MSY reference points

### 4.6.1 Settings

The settings used in the EqSim runs are summarised in Table 4.11. For both biological data and fisheries selectivity, the most recent 5 years were found to be most representative, and re-sampling from the last 5 years was used in EqSim.

In all EqSim simulations, a segmented regression was used as the stock-recruitment relationship. The breakpoint of the segmented regression was fixed, as fits from Beverton-Holt SRR and smooth hockey-stick produced a liner fit.

**Table 4.11. Gulf of Riga herring. Summary of settings used in EqSim runs.**

Input	Details
Biological data	Mean of the last 5 years
Fishing selectivity	Mean of the last 5 years
sigmaSSB	0.172 (from SAM)
$F_{cv}$	0.25 (ICES, 2015)
$F_{phi}$	0.30 (ICES, 2015)
Recruitment	No autocorrelation. Whole time-series used.
Stock-recruitment relationship	Segmented regression. Breakpoint fixed et $B_{pa}$ .

### 4.6.2 Estimating $B_{lim}$ , $F_{lim}$ and PA reference points

$B_{lim}$  is an important reference point from which other precautionary reference points are derived. To determine  $B_{lim}$ , the full assessment data series should be used to determine stock type in terms of the SSB-recruitment relationship. For Gulf of Riga herring there is no clear SRR. Beverton-Holt and smooth hockey-stick SRR produced a straight line, same problem was exhibited during the last benchmark when reference points were determined (ICES, 2015). Based on the ICES guidelines (ICES, 2021) the stock falls into *type 3*, however this *type* assumes that  $B_{lim}$  may be close to the highest observed SSB, this is considered not plausible for this stock, as current SSB is the highest observed and recruitment is above average. Therefore, defining a  $B_{lim}$  based on ICES stock *types* is not possible.

In cases where  $B_{pa}$  can be estimated but  $B_{lim}$  cannot, a proxy for  $B_{lim}$  is considered based on the inverse of the standard factor for calculating  $B_{pa}$  from  $B_{lim}$  (i.e. a  $B_{lim}$  proxy equal to  $B_{pa}/1.4$ ).  $B_{pa}$  is a stock status reference point above which the stock is considered to have full reproductive capacity. Following the procedure from last reference point calculations (ICES, 2015),  $B_{pa}$  is defined separately from  $B_{lim}$ , and used as the fixed breakpoint in segmented regression.

$B_{pa}$  was calculated with following steps:

- i. Calculate median SSB and R based on the whole time-series
- ii. Calculate average SSB based on data points that are  $\leq$  medianSSB and  $\geq$  medianR. Calculated average SSB =  $B_{pa}$ .

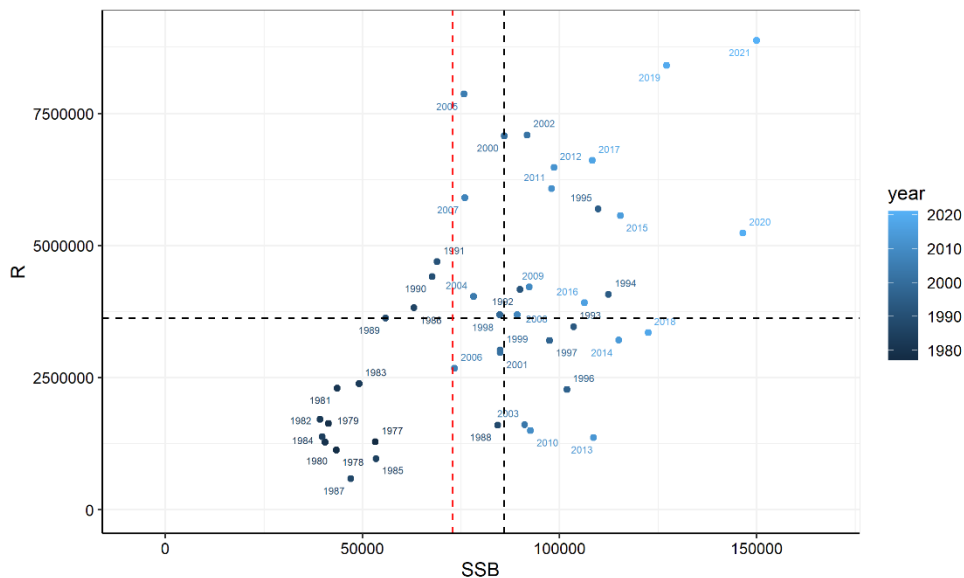


The median SSB and R, and the average SSB are shown in Figure 4.8. The corresponding values are:

MedianSSB = 85 994 tonnes

MedianR = 3 626 236 indiv.

averageSSB<sub>ssb=<medianSSB&R>=median</sub> = 72 907 tonnes



**Figure 4.8. Gulf of Riga herring. Stock-recruitment relationship. Black dashed vertical line = median SSB, horizontal black dashed line = median R, red dashed line = average SSB.**

Based on these assumptions:

$B_{pa} = \text{averageSSB} = 72\,907$  tonnes, and

$B_{lim} = B_{pa}/1.4 = 52\,076$  tonnes.

To estimate  $F_{lim}$ , EqSim was run without assessment/advice error, without advice rule, and with a segmented regression with a breakpoint fixed at  $B_{lim}$  to model recruitment in EqSim. The resulting  $F_{lim}$  ( $F_{50}$ ) obtained was 0.491.

### 4.6.3 Determining the SRR for $F_{MSY}$ calculations

The stock-recruitment relationship is crucial in the estimation of  $F_{MSY}$  and the risk of falling below precautionary biomass reference point. As mentioned above and seen in Figure 4.8, in overall recruitment appears to increase with SSB for all observed SSB values. Based on this and following the ICES guidelines (ICES, 2015) the stock will categorise under *type 3*. However, as explained above, the stock doesn't conform exactly to this *type*, as the current SSB is the highest observed, and there is no indication that the stock is impaired currently. This stock has exhibited high recruitment values since the beginning of 2000s, and it is more likely that the high SSB values are the result of high recruitments not the other way around (e.g. Figure 4.9 shows how SSB is related to R with a 2-year time lag).

To determine the most suitable SRR to be used in the estimation of  $F_{MSY}$ , we took an example from previous reference point workshop. We tested multiple SRR assumptions (more details in WD-4), in addition we compared these results with deterministic yield-per-recruit and spawner-per-recruit models. The comparison with YPR and SPR reference points was done, as these are commonly used  $F_{MSY}$  proxies (e.g.  $F_{35\%SPR}$  and  $F_{40\%SPR}$  are commonly used  $F_{MSY}$  proxies in the North

Pacific Fishery Management Council (Geromont and Butterworth, 2015)), which are used when no clear SRR is apparent.

Comparison tests showed that when assuming segmented regression with a fixed breakpoint at  $B_{pa}$  will lead to unconstrained  $F_{MSY}$  estimate which is between  $F_{35\%SPR}$  and  $F_{40\%SPR}$  estimates. While assuming a fixed breakpoint at a lower level than  $B_{pa}$  will lead to  $F_{MSY}$  estimates which are closer to  $F_{30\%SPR}$  estimate.

Based on equilibrium estimations and the comparisons with deterministic YPR and SPR reference points, it was deemed that having a segmented regression with a fixed breakpoint at  $B_{pa}$  as an SRR lead to reliable  $F_{MSY}$  estimate. In addition, we tested, that the  $F_{MSY}$  estimated with these assumptions for a  $F_{bar}=3-7$  will lead to  $F_{MSY}$  estimate (0.31) which is very similar to the current  $F_{MSY}$  estimate (0.32). Assuming  $B_{pa}$  as a breakpoint in the segmented regression is not typical procedure, however this conforms to the suggestions in ICES guidelines for stocks where recruitment appears to increase with SSB, that the change point of the segmented regression should be the average of all observed SSBs. In this case, the median SSB (85 994 tonnes) is slightly higher compared to the set breakpoint ( $B_{pa}=72\ 907$  tonnes).

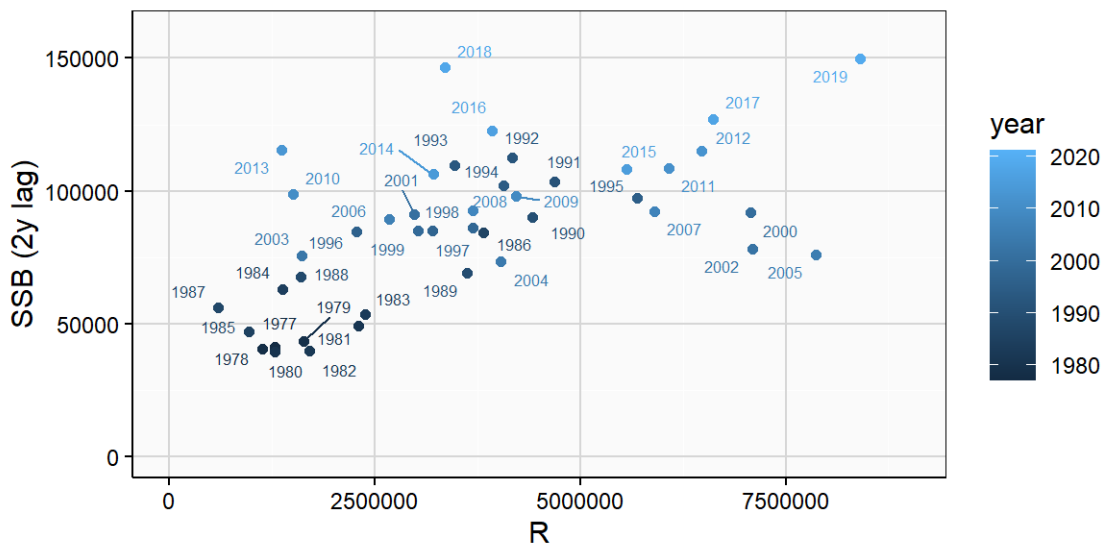


Figure 4.9. Gulf of Riga herring. Recruitment and SSB pairs with 2-year lag in SSB.

#### 4.6.4 Estimating $F_{MSY}$ , $F_{p.05}$ and $F_{pa}$

To estimate the unconstrained  $F_{MSY}$ , the EqSim was run without the advice rule (i.e. no MSY  $B_{trigger}$ ), with assessment and advice error using the values  $(F_{cv}, F_{phi}) = (0.25, 0.30)$  as suggested by WKMSYREF3 (ICES, 2015), and with a segmented regression with a breakpoint fixed at  $B_{pa}$  (Figure 10). The resulting unconstrained  $F_{MSY}$  obtained (median MSY for lanF) was  $F_{MSY} = 0.279$ .

To ensure consistency between the precautionary and the MSY frameworks,  $F_{MSY}$  is not allowed to be above  $F_{p.05}$ ; therefore, if the initial  $F_{MSY}$  value is above  $F_{p.05}$ ,  $F_{MSY}$  is reduced to  $F_{p.05}$ .  $F_{p.05}$  was calculated by running EqSim with assessment/advice error, with advice rule, and with a segmented regression with breaking point fixed at  $B_{pa}$  to ensure that the long-term risk of  $SSB < B_{lim}$  of any  $F$  used does not exceed 5% when applying the advice rule.  $F_{p.05}$  was estimated to be 0.353. Therefore, as explained above,  $F_{pa} = F_{p.05} = 0.353$ . Corresponding  $F_{MSY}$  ranges are shown in Table 12.

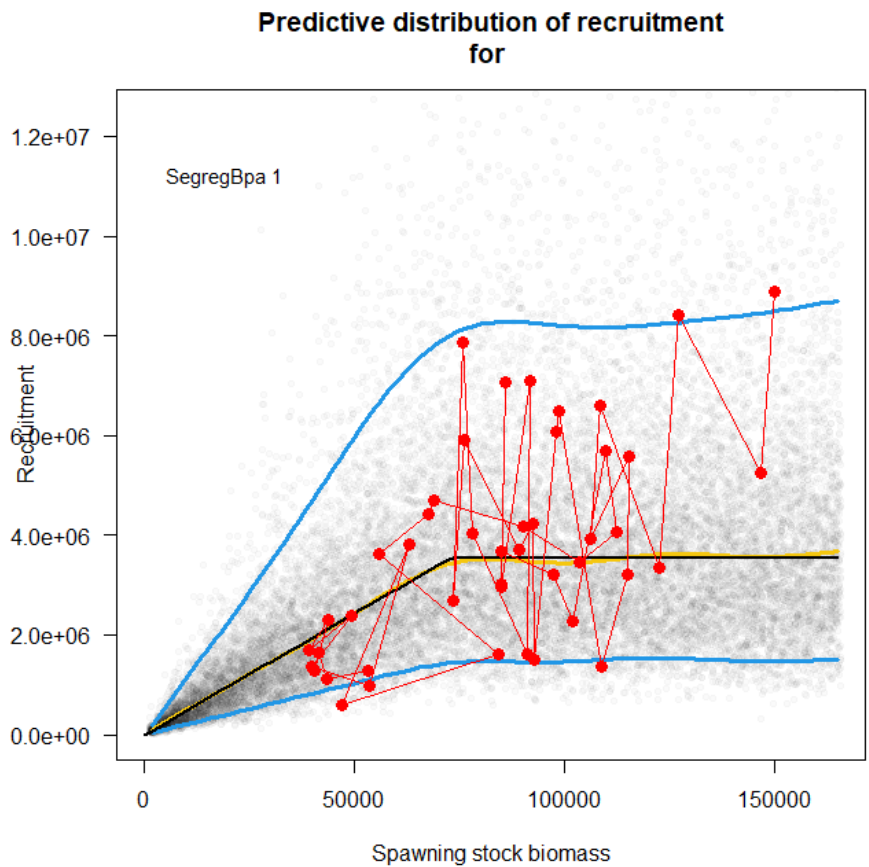


Figure 4.10. Gulf of Riga herring. Segmented regression using a fixed breakpoint at  $B_{pa}$  to fit stock-recruitment relationship.

Table 4.12. Gulf of Riga herring.  $F_{MSY}$  ranges.

Reference point	Value	Technical basis
$F_{MSYlower}$	0.21	$F_{MSY lower}$ (EqSim)
$F_{MSY}$	0.28	$F_{MSY}$ (EqSim)
$F_{MSYupper}$	0.33	$F_{MSY upper}$ (EqSim)

### 4.6.5 Estimating MSY $B_{trigger}$

MSY  $B_{trigger}$  is a lower bound of the SSB distribution when the stock is fished at  $F_{MSY}$  (ICES, 2021). As stated in the ICES technical guidelines, recent fishing mortality estimates need to be considered to set MSY  $B_{trigger}$  as for most stocks that lack data on fishing at  $F_{MSY}$ , MSY  $B_{trigger}$  is set at  $B_{pa}$ . Here, the stock has been fished below  $F_{MSY}$  (0.279) for the last 5 years. Next step is to look if the 5<sup>th</sup> percentile of  $B_{MSY} > B_{pa}$ . This is not the case and following the ICES technical guidelines our MSY  $B_{trigger}$  will be equal to  $B_{pa}$ , MSY  $B_{trigger} = 72\ 907$  tonnes.

#### 4.6.6 Comparison with previous reference points

The main difference between the previous and the new reference points obtained here is that the new values are lower ( $F_{MSY}$ ,  $F_{P.05}$ ,  $F_{lim}$ ) mostly due to change in  $F_{bar}$ , which now includes ages 2-6 compared to previous inclusion of ages 3-7.

$B_{pa}$  and  $B_{lim}$  and  $MSY B_{trigger}$  have increased compared to previous values. These differences are mainly caused by the categorisation of stock *Type*, while previously  $B_{lim}$  was defined as  $B_{loss}$  (stock *type* 5), currently obtained  $B_{lim}$  is inferred from  $B_{pa}$ .

### 4.7 Future research and roadmap

*Goal: Investigate the possibility to extend the time-series back in time.*

- Until 2003 the time-series used in the Gulf of Riga herring started from year 1970. However, in 2003, seven years of data (1970-1976) were discarded as the XSA assessment model estimated high fishing mortality for the earlier time period and these fishing mortality estimates were deemed not plausible for a pelagic stock (ICES, 2003). In theory, necessary data for years 1970-1976 is available.
- Data exchange files between Estonia and Latvia are available from 2003.

*Goal: Update the hydroacoustics abundance estimates*

- Re-calculate the hydroacoustic abundance estimates with StoX. For this to happen, data for years 1999-2010 need to be uploaded to the ICES acoustics database.
- Explore the possibility to calculate acoustics abundance estimates by mathematical models (GAM, VAST etc), which would take into account different factors that can affect the estimates ( e.g. year effect, water temperature, vessel etc).

### 4.8 Progress Feco

The progress on the work to calculate Feco for the stocks benchmarked was briefly presented during the benchmark. The work has been partly postponed, but will continue during spring 2023.

### 4.9 References

- Anders Nielsen, Casper W. Berg. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research* 158, 96-101. doi:10.1016/j.fishres.2014.01.014
- Berzinsh, V. 1995. Hydrology. In: Ojaveer, E. (ed) *Ecosystem of the Gulf of Riga between 1920 and 1990*. Estonian Academy Publishers. Tallinn. pp. 7-311.
- Geromont H, Butterworth DS. 2015. A review of assessment methods and the development of management procedures for data-poor fisheries. *FAO Report. The Marine Resource Assessment and Management Group (MARAM), University of Cape Town, South Africa.*
- ICES. 2001. Report of the Study Group on the herring assessment units in the Baltic Sea. ICES C.M. 2001/ACFM:10.
- ICES. 2003. Report of the Baltic Fisheries Assessment Working Group (WGBFAS), 7 -16 April 2003, ICES Headquarters, Copenhagen. ICES CM 2003\ACFM:21. 522 pp.
- ICES. 2005. Report of the Baltic Fisheries Assessment Working Group (WGBFAS), 12-21 April 2005, Hamburg, Germany. 589 pp.

- ICES, 2007 WGRP. Report of the Working Group on Recruitment Process, September 2007. ICES CM 2007/OCC:11. 12pp.
- ICES. 2008. Report of the Baltic Fisheries Assessment Working Group (WGBFAS), 8 -17 April 2008, ICES Headquarters, Copenhagen. ICES CM 2008\ACOM:06. 692 pp.
- ICES. 2015. Report of the Joint ICES-MYFISH Workshop to consider the basis for FMSY ranges for all stocks (WKMSYREF3), 17–21 November 2014, Charlottenlund, Denmark. ICES CM 2014/ACOM:64. 156 pp. <https://doi.org/10.17895/ices.pub.5661>
- ICES. 2021. ICES fisheries management reference points for category 1 and 2 stocks; Technical Guidelines. In Report of the ICES Advisory Committee, 2021. ICES Advice 2021, Section 16.4.3.1. <https://doi.org/10.17895/ices.advice.7891>
- ICES. 2022. Baltic Fisheries Assessment Working Group (WGBFAS). ICES Scientific Reports. 4:44. 659 pp. <http://doi.org/10.17895/ices.pub.19793014>
- Kornilovs, G. 1994. Yearly length distribution of herring in the Gulf of Riga in relation to population structure of the stock. ICES C.M. 1994/J:9.
- Putnis, I., Müller-Karulis, B., and Kornilovs, G. 2011. Changes in the reproductive success of Gulf of Riga herring. ICES CM 2011/H:13.
- Ojaveer, E., Jevtjukhova, B., Rechlin, O. and Strzyzewska, K. 1981. Results of investigations of population structure and otoliths of Baltic spring spawning herring. ICES C. M. 1981/J:19.
- Raid, T. 1990. Herring Spawning Grounds in the North-eastern Baltic: Recent Changes and Present Situation. Proc. Int. Herring Symposium, Oct. 1990, Anchorage, Alaska: 629-638.
- Raid, T., Kornilovs, G., and Shpilev, H. 2005. Stock diversity of herring in the Northern Baltic: is the separate assessment of herring natural stock units possible? ICES CM 2005/K:22.
- Rannak, L. 1971. On the factor determining the abundance of recruitment in the spring spawning herring of the eastern Baltic. ICES CM 1971 "special meeting on cod and herring in the Baltic", Herring, 11.

## 4.10 Annex 1. Gulf of Riga herring SAM model full settings.

```
# Configuration saved: Tue Jan 17 11:17:27 2023
#
# Where a matrix is specified rows corresponds to fleets and columns to ages.
# Same number indicates same parameter used
# Numbers (integers) starts from zero and must be consecutive
# Negative numbers indicate that the parameter is not included in the model
#
$minAge
# The minimum age class in the assessment
0
$maxAge
# The maximum age class in the assessment
8
```

`$maxAgePlusGroup`

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

1 1

`$keyLogFsta`

# Coupling of the fishing mortality states processes for each age (normally only  
# the first row (= fleet) is used).

# Sequential numbers indicate that the fishing mortality is estimated individually  
# for those ages; if the same number is used for two or more ages, F is bound for  
# those ages (assumed to be the same). Binding fully selected ages will result in a  
# flat selection pattern for those ages.

0 1 2 3 4 5 6 7 7

-1 -1 -1 -1 -1 -1 -1 -1 -1

`$corFlag`

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

# 0: independent means there is no correlation between F across age

# 1: compound symmetry means that all ages are equally correlated;

# 2: AR(1) first order autoregressive - similar ages are more highly correlated than  
# ages that are further apart, so similar ages have similar F patterns over time.

# if the estimated correlation is high, then the F pattern over time for each age

# varies in a similar way. E.g if almost one, then they are parallel (like a

# separable model) and if almost zero then they are independent.

# 3: Separable AR - Included for historic reasons . . . more later

2

`$keyLogFpar`

# Coupling of the survey catchability parameters (nomally first row is  
# not used, as that is covered by fishing mortality).

-1 -1 -1 -1 -1 -1 -1 -1 -1

-1 0 1 2 2 2 2 2 2

`$keyQpow`

# Density dependent catchability power parameters (if any).

```
-1 -1 -1 -1 -1 -1 -1 -1 -1
```

```
-1 -1 -1 -1 -1 -1 -1 -1 -1
```

```
$keyVarF
```

```
# Coupling of process variance parameters for log(F)-process (Fishing mortality
```

```
# normally applies to the first (fishing) fleet; therefore only first row is used)
```

```
0 1 2 2 2 3 4 5 5
```

```
-1 -1 -1 -1 -1 -1 -1 -1 -1
```

```
$keyVarLogN
```

```
# Coupling of the recruitment and survival process variance parameters for the
```

```
# log(N)-process at the different ages. It is advisable to have at least the first age
```

```
# class (recruitment) separate, because recruitment is a different process than
```

```
# survival.
```

```
0 1 1 1 1 1 1 1 1
```

```
$keyVarObs
```

```
# Coupling of the variance parameters for the observations.
```

```
# First row refers to the coupling of the variance parameters for the catch data
```

```
# observations by age
```

```
# Second and further rows refers to coupling of the variance parameters for the
```

```
# index data observations by age
```

```
0 1 1 1 1 1 1 1 1
```

```
-1 2 3 3 3 3 3 3 4
```

```
$obsCorStruct
```

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

```
"ID" "AR"
```

```
$keyCorObs
```

```
# Coupling of correlation parameters can only be specified if the AR(1) structure is chosen above.
```

```
# NA's indicate where correlation parameters can be specified (-1 where they cannot).
```

```
#0-1 1-2 2-3 3-4 4-5 5-6 6-7 7-8
```

```
NA NA NA NA NA NA NA NA
```

```
-1 0 0 0 0 1 1 1
```

`$stockRecruitmentModelCode`

# Stock recruitment code (0 for plain random walk, 1 for Ricker, 2 for Beverton-Holt, 3 piece-wise constant, 61 for segmented regression/hockey stick, 62 for AR(1), 63 for bent hyperbola / smooth hockey stick, 64 for power function with degree < 1, 65 for power function with degree > 1, 66 for Shepher, 67 for Deriso, 68 for Saila-Lorda, 69 for sigmoidal Beverton-Holt, 90 for CMP spline, 91 for more flexible spline, and 92 for most flexible spline).

0

`$noScaledYears`

# Number of years where catch scaling is applied.

0

`$keyScaledYears`

# A vector of the years where catch scaling is applied.

`$keyParScaledYA`

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

`$fbarRange`

# lowest and highest age included in Fbar

2 6

`$keyBiomassTreat`

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

-1 -1

`$obsLikelihoodFlag`

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

"LN" "LN"

`$fixVarToWeight`

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

0

`$fracMixF`

# The fraction of t(3) distribution used in logF increment distribution

0



\$fracMixN

# The fraction of t(3) distribution used in logN increment distribution (for each age group)

0 0 0 0 0 0 0 0

\$fracMixObs

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

0 0

\$constRecBreaks

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

\$predVarObsLink

# Coupling of parameters used in a prediction-variance link for observations.

-1 -1 -1 -1 -1 -1 -1 -1 -1

NA -1 -1 -1 -1 -1 -1 -1 -1

\$hockeyStickCurve

#

20

\$stockWeightModel

# Integer code describing the treatment of stock weights in the model (0 use as known, 1 use as observations to inform stock weight process (GMRF with cohort and within year correlations))

0

\$keyStockWeightMean

# Coupling of stock-weight process mean parameters (not used if stockWeightModel==0)

NA NA NA NA NA NA NA NA NA

\$keyStockWeightObsVar

# Coupling of stock-weight observation variance parameters (not used if stockWeightModel==0)

NA NA NA NA NA NA NA NA NA

\$catchWeightModel

# Integer code describing the treatment of catch weights in the model (0 use as known, 1 use as observations to inform catch weight process (GMRF with cohort and within year correlations))

0

\$keyCatchWeightMean

# Coupling of catch-weight process mean parameters (not used if catchWeightModel==0)

NA NA NA NA NA NA NA NA NA NA

\$keyCatchWeightObsVar

# Coupling of catch-weight observation variance parameters (not used if catchWeightModel==0)

NA NA NA NA NA NA NA NA NA NA

\$matureModel

# Integer code describing the treatment of proportion mature in the model (0 use as known, 1 use as observations to inform proportion mature process (GMRF with cohort and within year correlations on logit(proportion mature)))

0

\$keyMatureMean

# Coupling of mature process mean parameters (not used if matureModel==0)

NA NA NA NA NA NA NA NA NA NA

\$mortalityModel

# Integer code describing the treatment of natural mortality in the model (0 use as known, 1 use as observations to inform natural mortality process (GMRF with cohort and within year correlations))

0

\$keyMortalityMean

#

NA NA NA NA NA NA NA NA NA NA

\$keyMortalityObsVar

# Coupling of natural mortality observation variance parameters (not used if mortalityModel==0)

NA NA NA NA NA NA NA NA NA NA

\$keyXtraSd

# An integer matrix with 4 columns (fleet year age coupling), which allows additional uncertainty to be estimated for the specified observations

## Annex 1: WKBBALTPEL Resolution

A Benchmark Workshop on Baltic Pelagic stocks, chaired by External Chair Jim Ianelli, USA, and ICES Chair, Johan Lövgren, Sweden, and attended by invited external experts Alex Hansell, USA, and, Vanessa Trijoulet, Denmark, will be established. BWKBALTPEL will meet on 14-18 November 2022 for a data evaluation workshop (DEWK), and on 13-17 February 2023. Both meetings will take place at ICES HQ (Copenhagen with hybrid meeting access for all participants). If additional time is needed to agree to reference points and the short-term forecast, the benchmark can agree to additional meeting days. Stakeholders are invited to contribute data in advance of the data evaluation workshop (including data from non-traditional sources) and to contribute to data preparation and evaluation of data quality. BWKBALTPEL will evaluate the proposed assessments methods for the stocks listed in the table below and will work to:

- 1) As part of the data evaluation workshop:
  - a. Consider the quality of data proposed for use in the assessment;
  - b. Consider stock identity and migration issues;
  - c. Consider environmental drivers and impacts, including multispecies interactions;
  - d. Identify and analyse the quality of data proposed to use in the calculation of Feco<sup>1</sup>;
  - e. Make a proposal to the benchmark on the use and treatment of data for each assessment, including discards, surveys, life history, etc.
- 2) In preparation for the assessment methods workshop:
  - a) Following the DEWK, produce working documents to be reviewed during the Benchmark assessment meeting at least 14 days prior to the meeting.
- 3) As part of the assessment methods workshop, agree to and thoroughly document the most appropriate, data, methods and assumptions for:
  - a) Obtaining population abundance and exploitation level estimates (conducting the stock assessment);
  - b) Estimating fisheries and biomass reference points that are in line with ICES principles and objectives when deriving reference points, including the calculation of Feco<sup>\*</sup>.

If additional time is needed to conduct the work and agree to reference points, a short additional reference point workshop will be scheduled to conduct this work;

  - c) Agreeing on the settings to be used to conduct short-term forecast.

---

<sup>1</sup> Bentley, J.W., Lundy, M.G., Howell, D., Beggs, S.E., Bundy, A., De Castro, F., Fox, C.J., Heymans, J.J., Lynam, C.P., Pedreschi, D. and Schuchert, P., 2021. Refining fisheries advice with stock-specific ecosystem information. *Frontiers in Marine Science*, 8, p.602072.

Howell, D., Schueller, A.M., Bentley, J.W., Buchheister, A., Chagaris, D., Cieri, M., Drew, K., Lundy, M.G., Pedreschi, D., Reid, D.G. and Townsend, H., 2021. Combining ecosystem and single-species modeling to provide ecosystem-based fisheries management advice within current management systems. *Frontiers in Marine Science*, 7, p.607831.

- 4) As part of the assessment methods workshop, knowledge about environmental drivers, including multispecies interactions, and ecosystem impacts should be integrated in the methodology.
- 5) A full suite of diagnostics (regarding data, retrospective behaviour, model fit, predictive power etc.) should be examined as a whole to evaluate the appropriateness of any model developed and proposed for use in generating advice.
- 6) Update the stock annex as appropriate;
- 7) Develop recommendations for future improvements of the assessment methodology and data collection.

Stock	Description	Current model	Proposed model	ICES stock category	Coordinators	Assessor
her.27.25-2932	Herring ( <i>Clupea harengus</i> ) in subdivisions 25–29 and 32, excluding the Gulf of Riga (central Baltic Sea)	XSA	Stock Synthesis	1	Julita Gutkowska; Simon Smolinski	Mikaela Bergenius Nord, Massimiliano Cardinale
her.27.28	Herring ( <i>Clupea harengus</i> ) in Subdivision 28.1 (Gulf of Riga)	XSA	SAM	1	Tiit Raid	Kristiina Hommik; Ivars Putnis
spr.27.22-32	Sprat ( <i>Sprattus sprattus</i> ) in subdivisions 22–32 (Baltic Sea)	XSA	SAM	1	Olavi Kaljuste	Jan Horbowy

The Benchmark Workshop will report by 15 March 2023 for the attention of ACOM.

## Annex 2: List of participants

Member	Dept/Institute	Email
Alexander Hansell	Northeast Fisheries Science Center	alex.hansell@noaa.gov
Claus Reedtz Sparrevohn	Danish Pelagic Producers' Organisation	crs@pelagisk.dk
Elor Sepp	Department of Fish Biology and Fisheries	elor.sepp@ut.ee
Eros Quesada	SLU Department of Aquatic Resources Institute of Marine Research	eros.quesada@slu.se
Francesco Masnadi	University of Bologna	francesco.masnadi@studio.unibo.it
Harald Gjøsaeter	Institute of Marine Research	Harald.Gjoesaeter@hi.no
Ivars Putnis	Institute of Food Safety, Animal Health and Environment	ivars.putnis@bior.lv
Jan Horbowy	National Marine Fisheries Research Institute	horbowy@mir.gdynia.pl
Johan Lövgren	SLU Department of Aquatic Resources	johan.lovgren@slu.se
Jukka Pönni	Natural Resources Institute Finland	jukka.ponni@luke.fi
Julie Olivia Coad Davies	DTU Aqua, National Institute of Aquatic Resources	joco@aqua.dtu.dk
Julita Gutkowska	National Marine Fisheries Research Institute	jgutkowska@mir.gdynia.pl
Katja Ringdahl	SLU Department of Aquatic Resources	katja.ringdahl@slu.se
Kristiina Hommik	University of Tartu	kristiina.hommik@ut.ee
Maciej Tomczak	SLU Department of Aquatic Resources	maciej.tomczak@slu.se
Malin Skog	Swedish Pelagic Federation Producers Organisation	malin.skog@pelagic.se
Maria Golovaneva	Thünen-Institute of Baltic Sea Fisheries	maria.golovaneva@thuenen.de
Maria Pierce	Thünen-Institute of Baltic Sea Fisheries	maria.pierce@thuenen.de
Marie Storr-Paulsen	DTU Aqua, National Institute of Aquatic Resources	mstp@aqua.dtu.dk
Massimiliano Cardinale	SLU Department of Aquatic Resources	massimiliano.cardinale@slu.se
Mikaela Bergenius Nord	SLU Department of Aquatic Resources	mikaela.bergenius.nord@slu.se
Morten Vinther	DTU Aqua, National Institute of Aquatic Resources	mv@aqua.dtu.dk
Nicolas Goñi	Natural Resources Institute Finland	nicolas.goni@luke.fi
Noa Steiner	Christian-Albrechts-University of Kiel	nsteiner@ae.uni-kiel.de
Nuno Prista	SLU Department of Aquatic Resources	nuno.prista@slu.se

Member	Dept/Institute	Email
Olavi Kaljuste	SLU Department of Aquatic Resources	olavi.kaljuste@slu.se
Ruth Fernandez	International Council for the Exploration of the Sea	ruth.fernandez@ices.dk
Stefanie Haase	Institute of Marine Ecosystem and Fishery Science	stefanie.haase@thuenen.de
Szymon Smolinski	National Marine Fisheries Research Institute	ssmolinski@mir.gdynia.pl
Tiit Raid	Estonian Marine Institute	Tiit.Raid@ut.ee
Valerio Bartolino	SLU Department of Aquatic Resources	valerio.bartolino@slu.se
Vanessa Trijoulet	DTU Aqua, National Institute of Aquatic Resources	vttri@aqua.dtu.dk

## Annex 3: Reviewer report for WKBBALTPEL

Data evaluation workshop (DEWK) was held: November 14 - 18th

The benchmark workshop was held: February 13 - 17th

A series of hybrid meetings also occurred on January 10th, 16th and 23rd. At these meetings, the assessment teams provided updates and additional information on data/model decisions.

A series of meetings took place after the benchmark workshop to finalize the reference points for all the stocks on 3rd and 13th of March.

### **External chair's comments:**

The ICES chair led the WKBBALTPEL workshop, according to a rather flexible agenda where each of the three stocks under consideration were given time slots each day but where sufficient flexibility allowed for experts to present their work according to convenience. The reviewers were active asking for clarifications where appropriate and the discussions were to-the-point and helped inform the meeting about the work done. Both chairs took notes throughout which they compared and complemented as appropriate.

Much work had been done between the data evaluation meeting and the benchmark workshop, especially for one of the stocks, the Central Baltic herring. This also allowed for the reviewers to do much evaluation work before the benchmark workshop.

Unfortunately, as is often the case during benchmark workshops, some issues were raised at the final day, which could not easily be solved with few hours left of the meeting, resulting in a decision to hold a follow-up online meeting shortly after the benchmark. Such situations are not easily avoided, given the dynamics of the meeting, where experts of one stock present their work while other experts are working on their tasks, with limited possibility to follow all details of the work presented. Even in those cases where detailed working papers are produced prior to the meeting, time constraints make it difficult for the experts to allocate enough time for a thorough scrutinization of the presented work. A third factor, which probably played a role at the present meeting, was that relative novel methods, not often used in previous benchmarks and assessments, made it demanding to fully comprehend the work done and the consequences it would have for assessment and management.

The external chair shares the conclusions and recommendations of the external reviewers, outlined below.

### **External reviewers' comments**

#### **General comments:**

Most of the data decisions were made during the DEWK. A summary of the review and decisions made during the meeting are available. The current review mainly focuses on the benchmark meeting but still includes a short data section per stock.

The three stocks considered at the benchmark switched assessment model from Extended Survivor Analysis (XSA) to Stock Synthesis (SS) or State-space assessment model (SAM). The primary reason for the change is because XSA will not be maintained in the near future. We commend all the assessment teams for their effort to update the modeling platforms. For Baltic sprat

and Gulf of Riga herring (GoRH), SAM was already run along XSA for multiple years as exploratory runs and moving to SAM was straightforward. For Central Baltic herring (CBH), the incentive for changing to SS was to move in the near future to a spatial assessment model for this stock. SS can easily incorporate spatial data ( e.g. fleets, tagging and multiple areas) and therefore, was considered a good platform for this stock. However, in this benchmark SS was fit to the same inputs used in the previous XSA. This iteration of SS will be used into the next benchmark and will allow the experts to get comfortable with the model. Three SS runs with different assumptions of natural mortality were used to form an ensemble. Some concerns were raised briefly during the meeting regarding the fact that the models used in the ensemble are very similar (same configurations) except for the natural mortality assumption and that it seemed better to ensemble very different models ( e.g. a biomass production model with an age-structured model) rather than using common sensitivity runs as dimension of the ensemble. However, this method seemed commonly accepted by SS experts and was therefore accepted at the benchmark.

Mixing occurs between CBH and GoRH, CBH also mixes with western Baltic herring (whose one of the reviewers is the stock assessor of). Despite mixing being on the issue list for both stocks, mixing was not considered at this benchmark due to lack of data and the need for collaborative work between the different expert groups. These issues will be discussed at an upcoming workshop (WKSIDAC2) and collaborations will be developed between the different expert groups towards accounting for mixing and making it consistent among stocks in the future.

$F_{eco}$  was considered at the benchmark; however, not enough progress was made on this approach prior to the meeting and thus,  $F_{eco}$  was neither used or reviewed.

Significant time at the benchmark was spent discussing reference points and agreement could not be reached at the meeting. Two additional virtual meetings were needed to come to consensus on reference points. For all three stocks, recruitment relationships were not well defined, which made it difficult to estimate reference points because the current ICES framework requires  $F_{MSY}$  reference points. The reviewers recommended that if a stock recruitment relationship cannot be estimated, future benchmarks should consider using a  $F_{MSY}$  proxy ( e.g.  $F_{40\%}$ , the fishing mortality that is expected to conserve 40% of maximum spawning potential). Proxies are routinely used in other areas (USA, Canada, New Zealand) and support sustainable fishing. The difficulty is that the ICES framework as it currently is does not allow to get away from  $MSY$  reference points. Several ICES workshops are planned in the near future to redefine the calculation of reference points in ICES so this problem might be obsolete in the next few years.

For the experts using EqSim for reference point estimation, having a general script on the Share-Point containing all possible options down the decision tree available to all groups would be valuable. Despite the package being used for many years, the knowledge of experts is limited (for good reasons) when it comes to precise methods such as estimating the ranges when  $F_{MSY}=F_{p05}$ . A general script online (instead of going from one expert to another and being modified meanwhile) would make the process much smoother.

Major changes to the assessment that were not previously accepted should be flagged either before the benchmark or at the beginning of the benchmark. Suggesting major changes ( e.g. recruitment stanzas) should not be first mentioned on the last day of the meeting.

Due to time limits, the settings of the short-term forecast were not discussed for any of the stocks during the benchmark. According to the reviewers, this is not a big problem given that the decisions could be taken by the expert group, which is more knowledgeable regarding the stock biology.

For some of the stocks, discussion with experts happened prior to benchmark and outside of the extra pre-benchmark meetings by emails or comments on working documents. In this



benchmark, the reviewers found this very helpful. However, this assumes that the reviewers are willing to spend time outside of the meetings to start the review.

Reviewers' recommendations are highlighted in italic and final conclusions in bold.

### **Central Baltic Herring:**

Prior to this benchmark, this stock was assessed with an XSA. The assessment team decided to move toward Stock Synthesis (SS), which is an integrated statistical catch at age model that has a strong track history of providing catch advice for a wide variety of stocks. This assessment approach is particularly popular on the west coast of the United States. Plans were made to create a bridge run of the XSA to compare with the new SS model. However, due to time limitations at the meeting this run was not presented.

### **Data:**

The assessment model was fit to the same data as past XSA models. Data inputs to the model included: 1) a single commercial fleet; 2) acoustic survey; 3) age and weights from the fishery and survey. The reference run included time varying natural mortality, which was informed by SMS. Ideally, the assessment team wanted to include non-aggregated data to better understand spatial differences in the stock; however, this data was not available and will be explored at the next benchmark.

### **Model:**

Before the benchmark meeting the assessment, team participated in several hybrid meetings with the rest of the working group and reviewers. In these meetings initial model plans were discussed. Multiple weeks before the benchmark a draft report of the assessment and MSE were presented to the reviewers. All the necessary input files and R code to run the assessment/produce the diagnostics were also provided. Additionally, via email correspondence the lead analyst was responsive to reviewer questions as well as provided additional analyses as requested. Discussion before the meeting included: 1) fixing a minor issue in how age 0 is treated in the model; 2) discussions about correlation of R0 and steepness; 3) including one step ahead (OSA) residuals as a diagnostic; 4) conducting a sensitivity run looking at time varying fleet selectivity. The reviewers commend the thoroughness and responsiveness of the assessment team and believe these correspondences before the meeting improved the benchmark process.

The assessment team started with the old XSA inputs and then followed a decision tree to inform different model adaptations. Diagnostics used followed best practices (Carvalho et al., 2017; 2021; Winker et al. 2018) and included: 1) Convergence; 2) RMSE; 3) Retro; 4) MASE; 5) Jittering the starting values; 6) Plausibility. Based on reviewer feedback before the meeting and the development of code to estimate OSA residuals by the reviewers, the assessment team also produced OSA residuals. The OSA residuals were slightly larger and showed some correlations in time for both fleets. Some extra runs were requested by the reviewers before the meeting to try to improve these residuals; however, they were unsuccessful. To our knowledge, this OSA diagnostic has not been used before to evaluate SS runs and to estimate the weight of a SS model in the ensemble. *The reviewers suggest OSA residuals are routinely looked at as a diagnostic for SS.*

Selectivity of the acoustic survey in the assessment was discussed in detail at the benchmark meeting. Two options were presented to the group: 1) fixing selectivity for old ages (5-8) or 2) letting the model freely fit these ages. Freely fitting these ages led to a dome shaped selectivity. Both models passed all diagnostics and produced similar results. Based on expert knowledge of the survey and peer reviewed literature both options were considered as biologically plausible. Ultimately, the working group recommended that selectivity should be fixed for the older ages

because this model produced better predictions (MASE) for both the survey and commercial fishery age comps.

**The reference run passed all diagnostic tests, which is uncommon for SS, indicating the model is stable and usable as baseline run to set up the ensemble model.**

The ensemble model was run following the method developed by the SS experts. It is based on a baseline model, which passed all the diagnostic tests in SS and is believed to best represent the uncertainty in the stock natural mortality. **The ensemble model is therefore considered good for use in assessment and advice for CBH.**

#### Reference Points:

A MSE was run on the output of the ensemble model. One MSE run accounted for “implementation error”, which was defined as the difference of actual catch and TAC for the last three years (16%), so that the mean of the error is +16% and the standard deviation is 0.149. The other run is called “without implementation error”, but it corresponds to what is usually assumed in ICES with an implementation error with mean 0 and standard deviation (here same as before, e.g. 0.149). After seeing both runs, the working group agreed the run using implementation error was more appropriate given the uncertainty in future Russian catches. A significant portion of the meeting was spent discussing Russian catch because it accounts for ~ 20% of CPH landings.

Another meeting had to be scheduled after the benchmark because the working group could not agree on how reference points should be calculated. There was concern that a regime shift had occurred for central Baltic herring in 1990 and that it was inappropriate to use the entire time-series of recruitment in the calculation of reference points. However, proper analysis had not been conducted at the time of the benchmark to confirm this hypothesis.

A presentation at the benchmark was given about depensatory recruitment of CBH “Allee effect”, which provided useful context when estimating  $B_{lim}$ . Originally,  $B_0$  was calculated using three years of biology information. At request from the benchmark group,  $B_0$  was recalculated using a 10-year average of biological parameters because the group had concerns over low  $M$  in recent years (consequence of the decline of Baltic cod, only predator in the SMS model). Several different reference points were compared: Breakpoint,  $B_{loss}$ ,  $B_{15\%}$ ,  $B_{10\%}$ , Allee effect, 40%  $B_{MSY}$ , and 50%  $B_{MSY}$ . The team suggested that because of a poor stock recruit relationship  $B_{lim}$  was chosen by taking 15% of  $B_0$ . 15% was preferred over 10% of  $B_0$  because there was concern 10% was close to the “Allee effect”. Given that  $B_{MSY}$  is often less uncertain than  $B_0$ , which is often estimated outside the bounds of observed biomass estimates, some extra checks were requested by the reviewers to compare  $B_{lim}$  (15% $B_0$ ) to  $B_{MSY}$ . The experts calculated  $B_{lim}$  to correspond to around 51% $B_{MSY}$ , so very close to what would be usually used in the US to estimate  $B_{lim}$ . This reassured the reviewers on the choice of 15% $B_0$  for  $B_{lim}$ .

*The reviewers recommend to always check  $B_{lim}$  as a function of  $B_{MSY}$ , but also to check that  $MSYB_{trigger}$  and  $B_{lim}$  are distant enough to allow enough time for the stock to react to changes in management below  $MSYB_{trigger}$ .*

The final MSE results were presented in a table on March 13th, which made it easy for the benchmark group to see the different options. The first option “in the green” was recommended by the working group and reviewers. This option allowed the most amount of fishing, while ensuring  $B > B_{MSY}$  and  $F < F_{MSY}$ . This option results in an  $F_{MSY}$  proxy of  $F_{B35\%}$  and a  $MSYB_{trigger}$  reference point of 60% of  $B_{35\%}$ .

**Gulf of Riga Herring:**

The benchmark for Gulf of Riga herring (GoRH) aimed at considering three main issues: investigating the trap net tuning series, updating the maturity ogive values and changing the assessment model from XSA to SAM.

**Data:**

The trapnet tuning series was discarded at the DEWK due to an impossibility in reproducing the series back in time (more details in Annex). As a result, the assessment includes one commercial fleet and one acoustic tuning series (BIAS). The tuning series was modified at the beginning of the benchmark so that the age 8 was a plus group, a setting that was necessary for it to be correctly modeled in SAM. The survey was cross-checked with the numbers obtained in the WGBIFS report and the indices were the same.

The maturity at age estimates were updated. The raw values varied, so the GoRH experts favored using a smoothing procedure to get more realistic values of the maturity ogive for use in the assessment model. This method is commonly used in ICES. Prior to the benchmark meeting, some tests were made on trying to do this procedure directly in SAM by treating the maturity at age as a random effect. Unfortunately, the model was overfitting the raw values and no smoothing was occurring. It was therefore chosen by the GoRH experts to do the smoothing externally and to use the smoothed values directly as inputs in SAM. The reviewers double-checked that the overfitting was still happening with the final model configuration and it was the case so doing the smoothing outside of the model was considered the best option no matter the model configuration.

Some retrospective analysis on the smoothed values was done prior to the benchmark and showed that the spline estimates vary back in time as the data is peeled. Doing the smoothing every year at each data update will therefore result in a likely change in the perception of the stock (SSB) due to the maturity ogive being re-estimated and possibly modified back in time compared to the one used the year before. *The reviewers therefore suggest to clearly explain this in the advice sheet since Figure 2 might show a large variation in SSB from year to year.*

The average fishing mortality ( $F_{bar}$ ) range (being 3-7 prior to the benchmark) has not been re-considered for this stock in a long time. A catch curve was plotted and analyzed during the benchmark and showed that the age 2 individuals have been fully selected since the late 1990's. Given that ages 7-8+ do not represent a large proportion of the catch, the  $F_{bar}$  range was chosen to be 2-6. This new range was accepted as it will better represent the fishing catch selection.

**Model:**

Some difficulties were encountered during the benchmark in estimating a consistent recruitment at age 0. This was due to a problem in the model configuration that was giving too much flexibility and resulted in recruitment being sometimes lower than numbers at age 1. This problem was solved by fixing the configuration of the model (same survival variance for ages 1-8+).

The age 0 data in the catch was used in the model despite the information being poor and not really informative (large estimated standard deviation on the age 0 catch observation). This decision was taken given that a sensitivity run without age 0 showed the same outputs and diagnostics. While it will not make a difference model-wise, the GoRH experts preferred keeping the data in since age 0 are caught in the fishery.

The model consists of 9 ages (0-8+) and is run for the period 1977-2021. The inclusion of age 0 is new compared to the previous assessment. The tuning series includes data for 1999-2021 and ages 1-8+. SSB is modelled at the time of spawning (spring) and  $M=0.2$  for all ages and years.

The tuning of the model was done thoroughly by inspecting the model diagnostics (residuals and retrospective patterns) starting from a configuration that was improved from the original XSA. Many model configurations were tested and different sensitivity runs were performed. The different runs were made available on the *stockassessment.org* platform, which simplified the review significantly. For instance, sensitivity to natural mortality was done by looking at different values of  $M$  than 0.2 (average  $M$  from CBH,  $M$  based on life history parameters). These runs were just scaling up and down the stock with similar to worse model diagnostics.  $M=0.2$  was therefore kept as natural mortality for this stock. At this point, the run “GoR\_BP\_v2.2” on *stockassessment.org* was favoured.

It was noticed by the reviewers that the catchability estimates for the tuning series were estimated to be very similar for all ages for the favoured run. It was therefore decided to see if reducing the number of catchability parameters could improve the model (robustness, diagnostics or retros). Extra runs were performed but it was difficult to choose a preferred run based on model diagnostics. Therefore, the herring survey experts were asked for their opinions. They thought that it would make sense to have different catchability estimates for ages 1, 2 and 3+, which was relatively close to what was estimated when all were freely estimated. This latter run (“GoR\_BP\_v2.2.3qF\_s”) presented a similar AIC to GoR\_BP\_v2.2 and similar retrospective patterns and Mohn’s  $\rho$ . It was therefore chosen as the final model.

**The final assessment model GoR\_BP\_v2.2.3qF\_s shows reasonable residuals and retrospective patterns. The final model is therefore considered good for use in assessment and advice for GoRH.**

#### Reference points:

The choice was made by experts to use EqSim to estimate the MSY reference points for GoRH. Some trials were made by the reviewers to fit a stock recruitment relationship (SRR) in SAM so that SAM could be used for estimation of reference points. Unfortunately, there was no clear SRR for this stock and the fits resulted in a linear increase in recruitment to infinity. As a result, it was not possible to estimate reference points in SAM (no equilibrium yield could be reached, the biomass never getting to an equilibrium).

Variations in fishing selectivity and weight at age in recent years were investigated to choose the years that should be considered in EqSim for these inputs. Weights were consistent in the recent years while selectivity of older ages decreased recently (last 3 years). It was chosen to use the last 5 years in case the dome-shape selectivity does not perdure in the future. The average of the last 5 years for selectivity has a logistic shape.

The  $F_{cv}$  and  $F_{phi}$  (0.25 and 0.3, respectively) used in EqSim were the ones for the Baltic Sea stocks taken from the report of the Joint ICES-MYFISH Workshop to consider the basis for  $F_{MSY}$  ranges for all stocks (WKMSYREF3, ICES 2015). This seems reasonable since it comes from a meta-analysis.

The different reference points were estimated following common steps used by experts using EqSim.  $B_{pa}$  was estimated as a function of  $B_{lim}$  using the standard deviation of  $\log(SSB)$  in 2021 from SAM (0.172) rather than the default 0.2 value.

Given the difficulty in estimating a consistent stock-recruitment relationship for this stock, it was not possible to reach a consensus at the benchmark so the reference points were further discussed in additional meetings in March. This difficulty affects both the estimation of  $B_{lim}$  and  $F_{MSY}$ . It became clear that the biological reference point should be used to inform the stock-recruitment relationship so that the assumptions are consistent.

Three options were presented at the follow up meetings for determining reference points:

- Option 1: first define  $B_{pa}$  and use this value to infer  $B_{lim}$  and then use a segmented regression with inflection point at  $B_{pa}$  for the estimation of MSY reference points.
- Option 2: assume ICES type 2 with  $B_{lim}$  estimated as the lowest SSB that produces median recruitment (SSB in 1989), and then use a segmented regression with inflection point at  $B_{lim}$  for the estimation of MSY reference points.
- Option 3:  $B_{lim}$  equals  $B_{loss}$  and then use a segmented regression with inflection point at  $B_{lim}$  for the estimation of MSY reference points.

All three options result in different stock-recruitment relationships and therefore in different values of  $F_{MSY}$ .

After discussion on 13 March, Option 1 was chosen because this approach was used for this stock in the past and the stock appears to be highly productive; the values from this option were between  $F_{35\%SPR}$  and  $F_{40\%SPR}$  reference points, values often used for  $F_{MSY}$  proxy when a stock-recruitment relationship cannot be estimated.

*The reviewers recommend considering in the future MSY proxies for this stock if the stock-recruitment relationship continues to be unclear.*

### **Baltic Sea sprat:**

The benchmark for Baltic Sea sprat aimed at considering three issues: updating the natural mortality estimates to the new SMS keyrun, investigating and possibly updating the estimates of misreporting, and changing the assessment model from XSA to SAM.

#### **Data:**

Similarly to CBH, only the Danish catches were updated to correct for misreporting. The new estimates of  $M$  from the last SMS keyrun were used as input to the assessment model.

It was agreed during the DEWK that different runs be tested regarding the autumn BIAS survey. Similarly to CBH, BIAS has surveyed a new area (SD32) for the past 20 years. The sprat experts agreed on testing different data used for this survey. The conclusion was to use the BIAS indices as two separate surveys, one for SD 22-29&32 for 2000-2021, excluding the years 2001-2004 and 2008; and one for SD 22-29 for the years 1991-2008 excluding the years the survey SD 22-29&32 is used (see model section below for more details).

The age 0 survey is shifted to age 1 for use in the assessment. During the benchmark it was chosen to use the estimates in the intermediate year (2022) to inform the recruitment in 2022. This is new compared to how the XSA was parameterized. The reviewers consider this an improvement since the recruitment estimates in the intermediate year can be used in the short-term forecast instead of an average or sampled recruitment. Some work was done and presented at the DEWK regarding a larval drift model to correct the age 0 survey for the assessment but the model was not finalized in time for the benchmark.

Except for the changes above, the data used were the same as the previous assessment, i.e., commercial catches 1974-2021 for ages 1-8+; BIAS SD 22-29&32, SD 22-29, and May survey (2001-2021) for ages 1-8+; and age 0 survey 2010-2022 shifted to age 1.

#### **Model:**

The model consists of 8 ages (1-8+) and is run for the period 1974-2022. As mentioned above, the inclusion of the intermediate year is new compared to the previous assessment. SSB is modeled at the time of spawning (summer).

A RMarkdown script was developed by the sprat experts to produce a working document for all model configurations tested, collecting code and figures. The documents were uploaded to the sharepoint for review. In addition, the fit of specific model versions was provided on the sharepoint for extra checks by the reviewers.

Initially, five model configurations were tested and presented at the beginning of the benchmark meeting. The configuration started with a configuration equivalent to the old XSA run and developed by testing different survey catchability assumptions and observations correlation structures. At this point, the run 4 was privileged since it was the run with best retrospective patterns and second best AIC. The age composition residuals for this run 4 were not perfect but could not be improved. This was double-checked by the reviewers who did not manage to get a better configuration than the one proposed by the sprat experts without compromising on the other model diagnostics (AIC, retros, etc.).

Extra runs were requested by the reviewers where the BIAS October survey is implemented as agreed at the DEWK. This resulted in 3 additional runs, i.e., run 6 with survey 22-29&32 only, run 7 with survey SD 22-29 only with shortened time-series, and run 8 SD 22-29 only with full time-series. The runs 4 and 6 were very similar but the retrospective patterns and Mohn's rho were better for the run 4, which was therefore still favoured.

The run 4 includes an extra parameter per survey for the age 1 that allows the index equation to not be linearly proportional to the fish numbers but to be proportional to a power (called here  $\alpha$ , such as  $\alpha \neq 1$ ) of it as follows:  $I = qN^\alpha$ . This assumption was also used in past XSA assessments. After additional checks by the sprat experts and the reviewers, the inclusion of the power parameter was accepted. Including the power parameter was biologically realistic, was statistically significant, improved AIC and sensitivity runs showed not including it had minimal effect on stock trends. This setting in SAM results in a perfect correlation between  $\alpha$  and the catchability at age 1 for each survey that cannot be avoided but does not seem to alter the model robustness.

Finally, the sprat experts mentioned they wanted to use the age 0 survey index for 2021 as an indication for recruitment at age 1 in 2022 for the short-term forecast. It was brought to the attention of the experts that, in SAM, it is possible to run the model to the intermediate year if a survey is available in that year. Therefore, a final run 10 was performed that runs up to 2022 with extra data for age 0 survey in 2022. This run resulted in a very small difference in model outputs compared to run 4 (that did not include the age zero survey in the terminal year + 1) and had very similar retrospective patterns. It was therefore favoured as the final model given that it enables the recruitment in the intermediate year to be estimated rather than assumed for the short-term forecast.

To keep full flexibility when estimating the reference points, the sprat experts did not want to take advantage of having a stock recruitment relationship directly estimated in the assessment model (run 9), which would allow the internal estimation of reference points in SAM. The reviewers understand the experts' choice but would have favoured the estimation of reference points consistent with the assessment model. *The reviewers recommend that in future updates of the assessment the analysts re-explore estimating reference points in SAM.*

**The final assessment model run10 shows reasonable residuals and retrospective patterns. The final model is therefore considered good for use in assessment and advice for Baltic Sea sprat.**

#### Reference points:

Significant discussions occurred around setting  $B_{lim}$  with four options considered to set it. It was ultimately determined that the SSB that produces 50% of the maximum equilibrium recruitment (recruitment at SPR with  $F=0$ ) should be used. This was the same option as the one used in the last benchmark for this stock.

The choice was made by experts to use EqSim to estimate the MSY reference points for Baltic sprat.

The experts chose to use the last 5 years of the assessment model (excluding the intermediate year 2022) as setting for the biological data and fishing selectivity in EqSim. Similarly to GoRH, the default from the Baltic stocks were used for setting  $F_{cv}$  and  $F_{phi}$ . At a follow up meeting and to be consistent with CBH (that also uses SMS to inform  $M$ ), it was changed and the last 10 years of the model were used as setting for biological data.

The different reference points were estimated following common steps used by experts using EqSim.  $B_{pa}$  was estimated from  $B_{lim}$  using the standard deviation of  $\log(SSB)$  in 2021 from SAM (0.101). The extreme values obtained from the SRR simulations in EqSim were trimmed to avoid unrealistic values (option in EqSim). The experts used both Beverton-Holt and the segmented regression with freely estimated inflection point to estimate MSY reference points given that they both are equally chosen as best in EqSim (52% Beverton-Holt vs. 48% segmented regression).

The sequence of fishing mortality considered in EqSim was increased as requested by the reviewers to improve the accuracy of the MSY estimation.

The reference points values were validated on the 13th of March.

## Annex 4: Updated Biological Reference Points for Central Baltic Herring

The WKBBALTPEL group proposed a set of target and trigger reference points derived from MSE with implementation error set to 0.165 with standard deviation equal to 0.149 (see relevant section of the main WKBBALTPEL report). This procedure had been followed previously for a *Pandalus* stock (pra.27.3a4, ICES, 2022) for which ICES provides catch advice. At WKBBALTPEL  $B_{lim}$  was defined as 15% of  $B_0$  (unexploited SSB at current conditions).

The ICES Advisory Committee (ACOM) accepted the new definition of  $B_{lim}$ . However, after the WKBBALTPEL was adjourned and during WGBFAS 2023, ACOM considered that it was more appropriate to adopt reference points derived from MSE without implementation error. ACOM will discuss how to handle implementation error and produce guidelines on this for both MSE in general and the estimation of reference points in particular. The selection of reference points based on MSE is not straightforward (several  $F_{brp}$ ,  $B_{urg}$  and  $B_{trigger}$  combinations can be selected according to stock-specific productivity and trade-offs) and therefore ACOM suggested that the decision on the new set of reference point should be taken at WGBFAS 2023.

The Table below includes the set of agreed reference points at WGBFAS 2023.



Framework	Reference point	Value	Technical basis	Source
MSY approach	MSY $B_{trigger}$	$B_{30\%}$	Relative value. Set at 30% of $B_0^*$ . Determined through management strategy evaluation with the objective to achieve high sustainable yields without exceeding a 5% probability of SSB falling below $B_{lim}$ in any single year.	ICES (2023a)
	$F_{MSY}$	$F_{B30\%}$	Relative value. Set as the F which will achieve 30% of $B_0$ . Determined through management strategy evaluation with the objective to achieve high sustainable yields without exceeding a 5% probability of SSB falling below $B_{lim}$ in any single year.	ICES (2023a)
Precautionary approach	$B_{lim}$	$0.15 \times B_0$	Relative value. Set at 15% of $B_0$ .	ICES (2023b)
	$B_{pa}=MSY B_{trigger}$	$B_{30\%}$	Relative value. Set at 30% of $B_0$ . Determined through management strategy evaluation with the objective to achieve high sustainable yields without exceeding a 5% probability of SSB falling below $B_{lim}$ in any single year.	ICES (2023a)
	$F_{pa}$	$F_{B25\%}^{**}$	$F_{P05}$ . Relative value. Determined through management strategy evaluation. The F that leads to $SSB \geq B_{lim}$ with 95% probability.	ICES (2023a)
Management plan	MAP MSY $B_{trigger}$	$B_{30\%}$	MSY $B_{trigger}$	ICES (2023a)
	MAP $B_{lim}$	$0.15 \times B_0$	$B_{lim}$	ICES (2023a)
	MAP $F_{MSY}$	$F_{B30\%}$	$F_{MSY}$	ICES (2023a)
	MAP target range $F_{lower}$	$F_{B40\%}$	Relative value. Determined through management strategy evaluation, consistent with the ranges which result in no more than a 5% reduction in long-term yield compared to MSY.	ICES (2023a)
	MAP target range $F_{upper}$	$F_{B25\%}^{**}$	Relative value. Determined through management strategy evaluation, consistent with the ranges which result in no more than a 5% reduction in long-term yield compared to MSY. Capped to $F_{P05}$ .	ICES (2023a)

\*  $B_0$  is the estimated unexploited spawning biomass at current conditions (average of the last 10 years in biology)

\*\* Determined from the management strategy evaluation, to be precautionary this reference point can only be used with the MSY  $B_{trigger}$

### References:

- ICES. 2022. Benchmark workshop on *Pandalus* stocks (WKPRAWN). ICES Scientific Reports. 4:20. 249 pp. <http://doi.org/10.17895/ices.pub.19714204>
- ICES. 2023a. Baltic Fisheries Assessment Working Group (WGBFAS). ICES Scientific Reports. 5:58. 606 pp. <https://doi.org/10.17895/ices.pub.23123768>
- ICES. 2023b. ICES. 2023. Benchmark Workshop on Baltic Pelagic stocks (WKBBALPEL). ICES Scientific Reports. 5:47. <https://doi.org/10.17895/ices.pub.23216492>

## Annex 5: Working documents

WD	Title	Page
1	Natural mortalities as estimated by SMS	147
2	Correction of the catch data (CATON, CANUM) for herring ( <i>Clupea harengus</i> ) in subdivisions 25-29 and 32, excluding the Gulf of Riga (central Baltic Sea)	167
3	Notes on sampling design, data storage and data edits underlying the Swedish data biological data on commercial caught herring (her.27.25-2932) provided to the 2023 ICES Benchmark of Central Baltic Herring (BWKBALTPEL)	175
4	SmartDots Summary Report for the 2022 exchange for the central Baltic herring stock her.27.25-2932 (event ID 449)	202
5	Comparing WEST (weight-at-age in the catch) and WECA31, (weight-at-age in the stock) for sprat ( <i>Sprattus sprattus</i> ) and central Baltic herring ( <i>Clupea harengus</i> ) in the Baltic Sea	207
6	Maturity-at-age analysis for herring ( <i>Clupea harengus</i> ) in subdivisions 25-29 and 32, excluding the Gulf of Riga (central Baltic Sea)	215
7	Natural Mortality estimated using Life History based methods Stock - Herring ( <i>Clupea harengus</i> ) in subdivisions 25-29 and 32, excluding the Gulf of Riga (central Baltic Sea)	229
8	Working Document on potential corrections of national catch data of Baltic sprat and central Baltic herring	240
9	A proposal for FMSY reference points for Baltic sprat SDs 22-32	285
10	Maturity-at-age analysis for Gulf of Riga herring (Subdivision 28.1)	291
11	Analysis of Gulf of Riga herring trap net tuning series	303
12	Gulf of Riga herring assessment input data and assessment model settings	314
13	Reference point calculation for Gulf of Riga herring in subdivision 28.1	325
14	Workshop on Ecosystem-Based Fisheries Advice for the Baltic II- (WKEBFABII) – preliminary conclusions and perspective.	343
15	Effects of herring and sprat misreporting on assessment of both stocks – simulations study	346

# Natural mortalities as estimated by SMS.

## Working document for the ICES WKBBALTPEL

Morten Vinther, DTU Aqua, Denmark

2022-11-24

### Summary

Estimates of natural mortality ( $M = M1 + M2$ ) of herring were obtained from a number of runs using the SMS models with alternative configurations of food intake by cod and alternative values of  $M1$  for herring estimated from growth. The key run made by ICES WGSAM in October 2022 is considered to provide the “best” estimate of  $M$ , as it has been through a review process. The alternative runs provide however quite similar results and provided in some cases a better model likelihood than the key run, even though the differences were small. The difference in total mortality ( $Z$ ) estimated by the various model configurations are small, especially for the age 3+. Key run  $Z$  for herring age 3+ seem higher than  $Z$  estimated from catch curves, which may indicate that the key run  $Z$  might be too high for herring, and that an alternative model configurations which produces a lower  $M$  are preferable for use in ICES single species assessments

### Introduction

ICES WGSAM (2022) made an updated, 1974-2021, key run for the Eastern Baltic Sea for the predator species cod and prey species herring and sprat. This key run estimates natural mortalities ( $M$ ) consisting of an assumed known and constant residual mortality ( $M1$ ) and a predation mortality ( $M2$ ) estimated by the SMS model. These natural mortalities are available for single species assessment.

ICES WKBBALTPEL has requested  $M$  values for herring and sprat as estimated by the key run. For herring, WKBBALTPEL will try an ensemble approach for the stock assessment which includes several model configurations using “likely”  $M$  values derived from the key run. ICES WGSAM considers that the most likely  $M$  values are those estimated by the key run. SMS produces uncertainties for the estimated  $M$  values, but these uncertainties seem to narrow, probably a result of the configuration where the stock size of the predator (cod) is assumed known without errors. This means that “likely” estimates of  $M$  cannot be obtained from the estimated values within a single SMS run.

### Methods

Sensitivity analyses within the SMS framework (see the Stock annex for the key run) show that the estimated  $M$  values are sensitive to the total eaten biomass (by cod) and the (input) value of  $M1$ . Alternative model configurations with likely estimates of cod consumption or  $M1$  values were made to estimate likely estimates of  $M$  for use in the single species assessments.

### Variable food consumption

The food consumption by cod will depend on the estimated stock size of cod, and the biomass consumed by the individual cod. Cod in SMS is treated as a so-called “other predator” for which the stock size is assumed

known without error. The ICES single stock assessment results are considered as the best available estimate of the stock size of cod and is used as input to SMS. The ICES cod assessment estimates uncertainties around the stock size, but the estimated confidence intervals seem unrealistically narrow, such that “populations” of cod drawn from the ICES results will probably not reflect the real uncertainties of the SMS cod population, and as such not provide realistic confidence intervals around the estimated M values.

The consumption rates (kg biomass per quarter per cod) used in the key run is based on the results from Neuenfeldt et al., 2020, which uses the same stomach data set as applied for estimation of diet for use in SMS. Food consumption is calculated for three periods based on the relationship between food intake and cod length (Food ration =  $a * \text{length}^b$ ) as shown below (Section 2.3.3 of the StockAnnex provides more details).

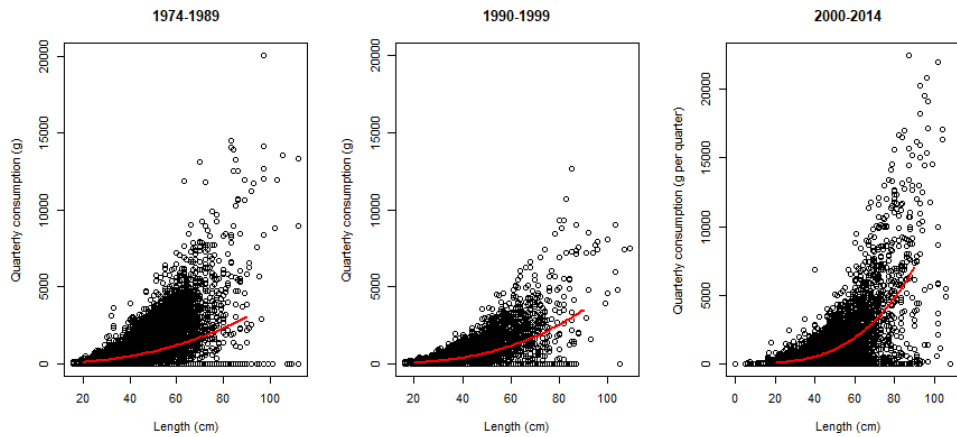


Figure 1: Scatterplots of cod total length and estimated quarterly consumption rate. The consumption rate has been estimated separately for 1974-1989, 1990-1999 and 2000-2014 in order to account for recent changes in cod consumption (Neuenfeldt et al., 2020).

Table 1: Estimated parameters for cod consumption based on the relation consumption =  $a * \text{length}^b$ .

Period	Parameter	Estimate	Std. Error
1974-1989	a	0.10367	0.01184
	b	2.28617	0.02834
1990-1999	a	0.01741	0.00397
	b	2.71370	0.05457
2000-2014	a	0.003230	0.000354
	b	3.243353	0.02550

A number of “likely” set of M values was produced from SMS configurations using different values of cod food consumptions, derived from the parameter estimate of the mean consumption. Individual SMS runs were made for following configurations of the a and b parameters for food consumptions.

- lim\_10: 10% quantile of the parameter a and b (ignoring correlation)
- lim\_25: 25% quantile
- key\_run: 50% quantile
- lim\_75: 75% quantile
- lim\_90: 90% quantile

The resulting average cod consumption rates for are shown in Figure 2 for the period 1974-1989.

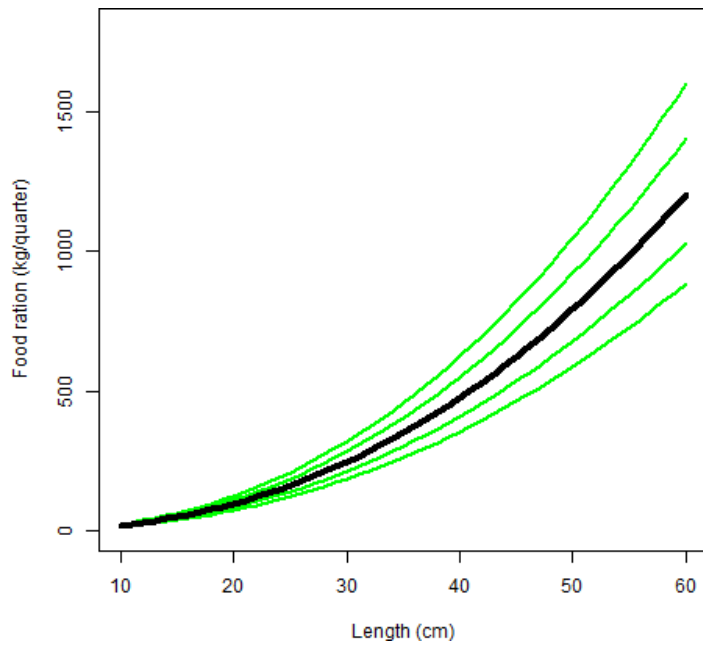


Figure 2: Average quarterly consumption rate by length and cod for the period 1974-1989. The lines show from the bottom to the top the 10, 25, 50, 75 and 90 percentile quantiles of the parameter estimates  $a$  and  $b$  used to calculate consumption:  $\text{Consum} = a * L^b$

## Variable M1 for herring

Values of M1 for **herring** was guided by analysis of growth, which shows a lower growth rate after year 2000 which is interpreted as a higher mortality. The M1 is not known, but given an assumed annual M1 and a change in mortality after 2000 the following scenarios for “likely” M1 were made

- M1\_005: average annual M1 = 0.05, Quarterly M1(1974-1999)=0.04/4 and M1(2000-2021)=0.06/4
- M1\_010: average annual M1 = 0.1, Quarterly M1(1974-1999)=0.08/4 and M1(2000-2021)=0.12/4
- M1\_020\_average annual M1 = 0.2, Quarterly M1(1974-1999)=0.17/4 and M1(2000-2021)=0.23/4
- key\_run: average annual M1=0.1, Quarterly M1(1974-2021)=0.1/4

M1 for **sprat** was kept constant at 0.2 (annual), as used in the key run, for all scenarios.

## Results

The likelihoods for each run is presented in Table 2, and shows that the best fit (lowest “neg.log.like”) is obtained from the “M1\_020” run, but the differences in likelihoods between runs are small. All runs are made with the same model configuration, while a run-specific configuration might change the performance of the individual runs.

The likelihoods from herring and sprat (Table 3) show that for herring, the best likelihood for catch observations is by the “M1\_020” run, while the best likelihood for survey cpue is for the run “lim\_010”. Sprat input are the same for all runs, but the indirect effect means that the best fit for sprat is obtained by the “lim\_090” run.

Table 2: Likelihood contribution from Catch, survey (CPUE), stock-recruitment relation (SSB.Rec) and stomach contents. The total weighted likelihood (neg.log.like) and AIC are also shown

Run	catch	CPUE	SSB.Rec	stomachs	neg.log.like	AIC
key_run	-1045	-263	-14	-338	-1646	-2635
lim_10	-1043	-264	-18	-346	-1654	-2649
lim_25	-1044	-263	-16	-342	-1651	-2643
lim_75	-1047	-262	-12	-332	-1642	-2626
lim_90	-1048	-262	-11	-327	-1637	-2617
M1_005	-1045	-261	-13	-335	-1642	-2627
M1_010	-1049	-262	-17	-347	-1659	-2661
M1_020	-1055	-261	-25	-368	-1686	-2713

Table 3: Likelihood contribution from Catch, survey (CPUE), stock-recruitment relation (SSB.Rec) and weighted sum (neg.log.like).

Run	Species	catch	CPUE	SSB.Rec	neg.log.like
key_run	Herring	-858	-144	-7	-1002
lim_10	Herring	-859	-145	-10	-1004
lim_25	Herring	-858	-145	-8	-1003
lim_75	Herring	-857	-144	-5	-1001
lim_90	Herring	-856	-144	-4	-1000
M1_005	Herring	-857	-143	-5	-1000
M1_010	Herring	-861	-144	-9	-1005
M1_020	Herring	-866	-142	-17	-1009
key_run	Sprat	-188	-118	-8	-307
lim_10	Sprat	-184	-119	-8	-303
lim_25	Sprat	-186	-119	-8	-305
lim_75	Sprat	-190	-118	-7	-309
lim_90	Sprat	-192	-118	-7	-311
M1_005	Sprat	-188	-119	-8	-307
M1_010	Sprat	-188	-119	-8	-308
M1_020	Sprat	-189	-119	-8	-309

### Variable food consumption

For both Herring (Figure 3) and Sprat (Figure 4), a higher food consumption by cod gives a higher mortality and a higher Recruitment and SSB. Fishing mortality decreases by increasing consumption as the stock size is estimated lower.

The natural Mortality ( $M=M1+M2$ ) increases for both Herring (Figure 5) and Sprat (Figure 6) with increasing food consumption by cod.

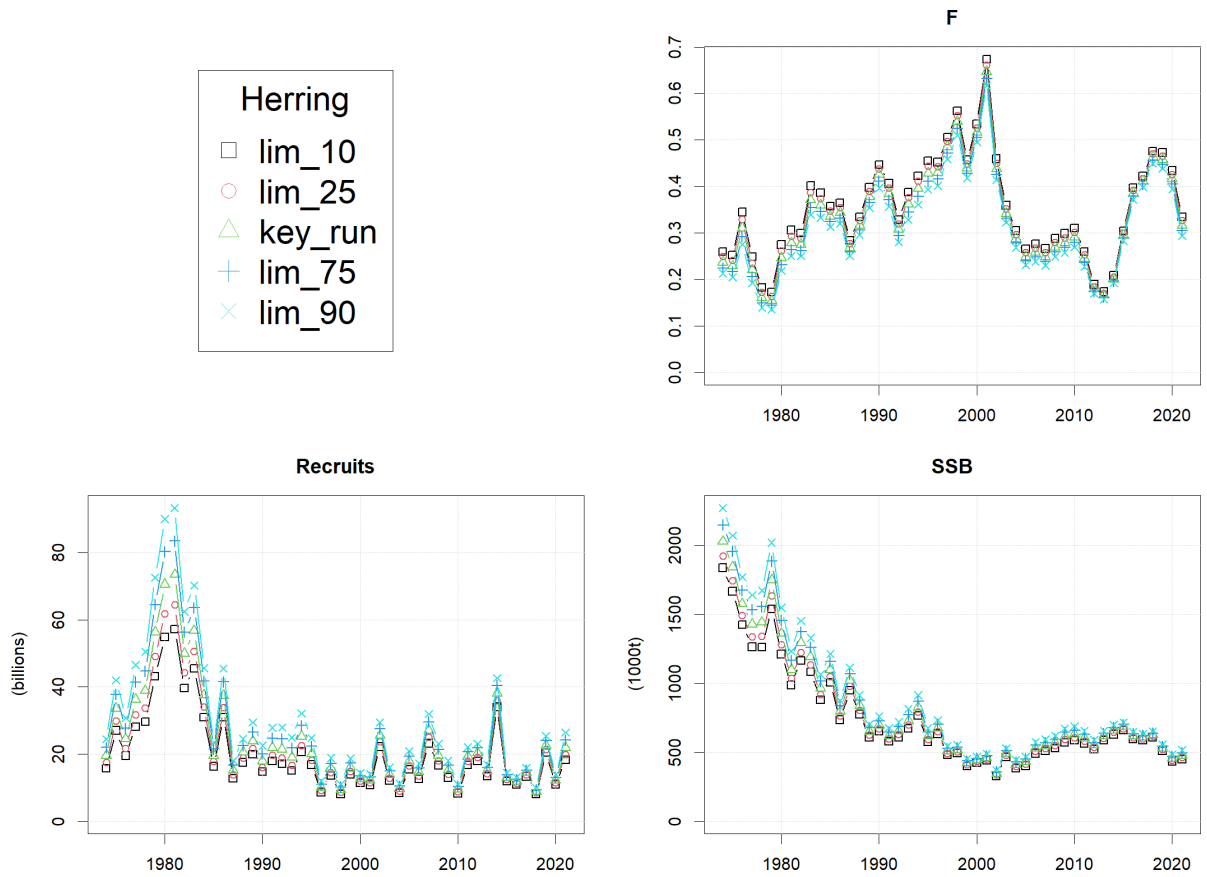


Figure 3: Assessment results for the five runs with variable cod consumptions.



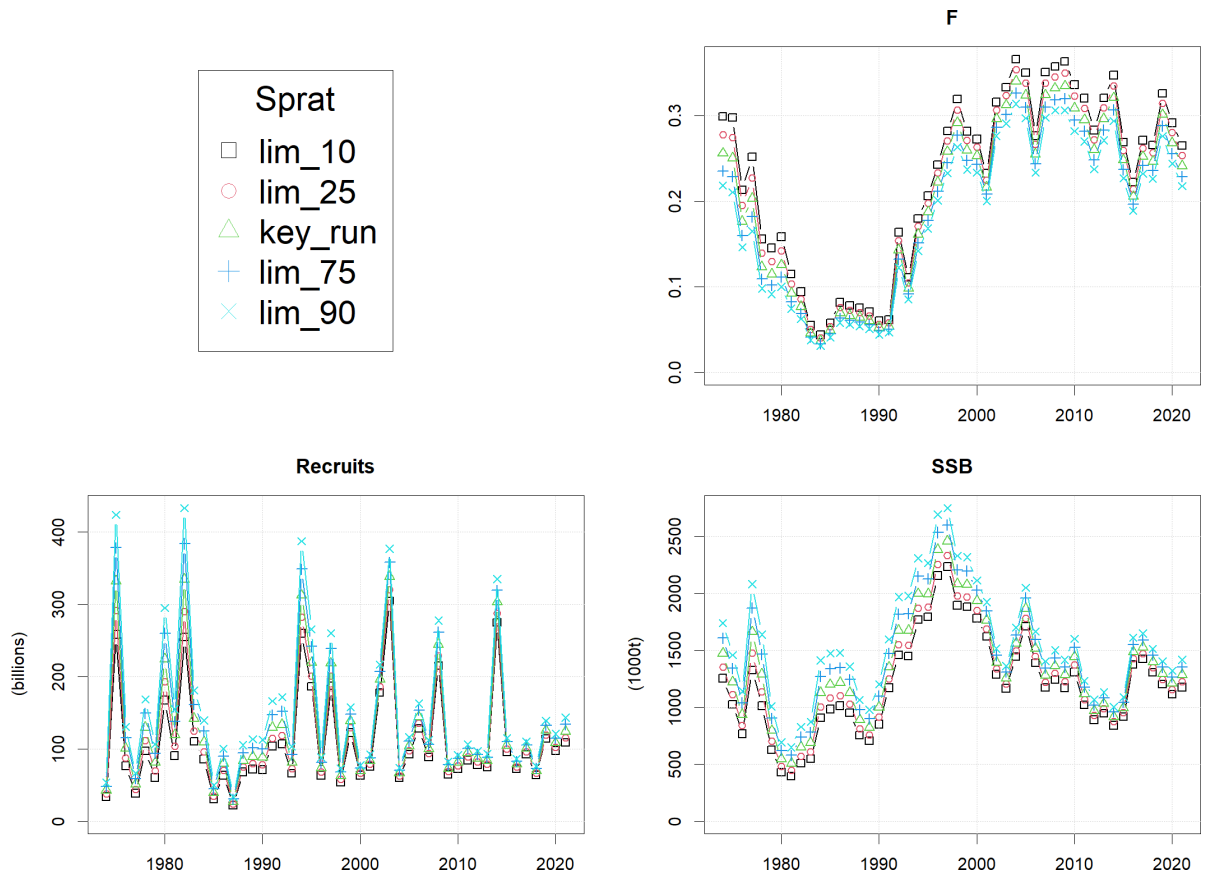


Figure 4: Assessment results for the five runs with variable cod consumptions.

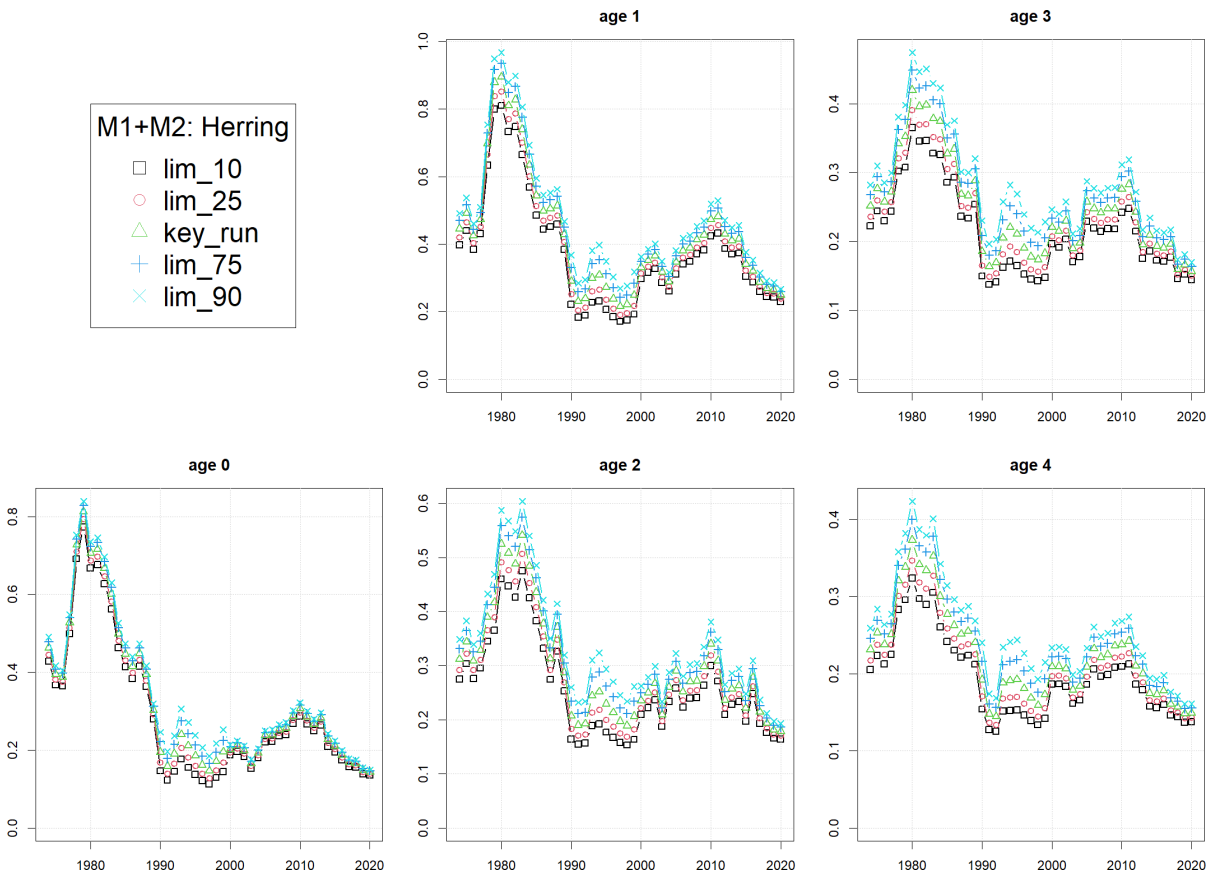


Figure 5: Natural mortality ( $M = M1 + M2$ ) for the five runs with variable cod consumptions.

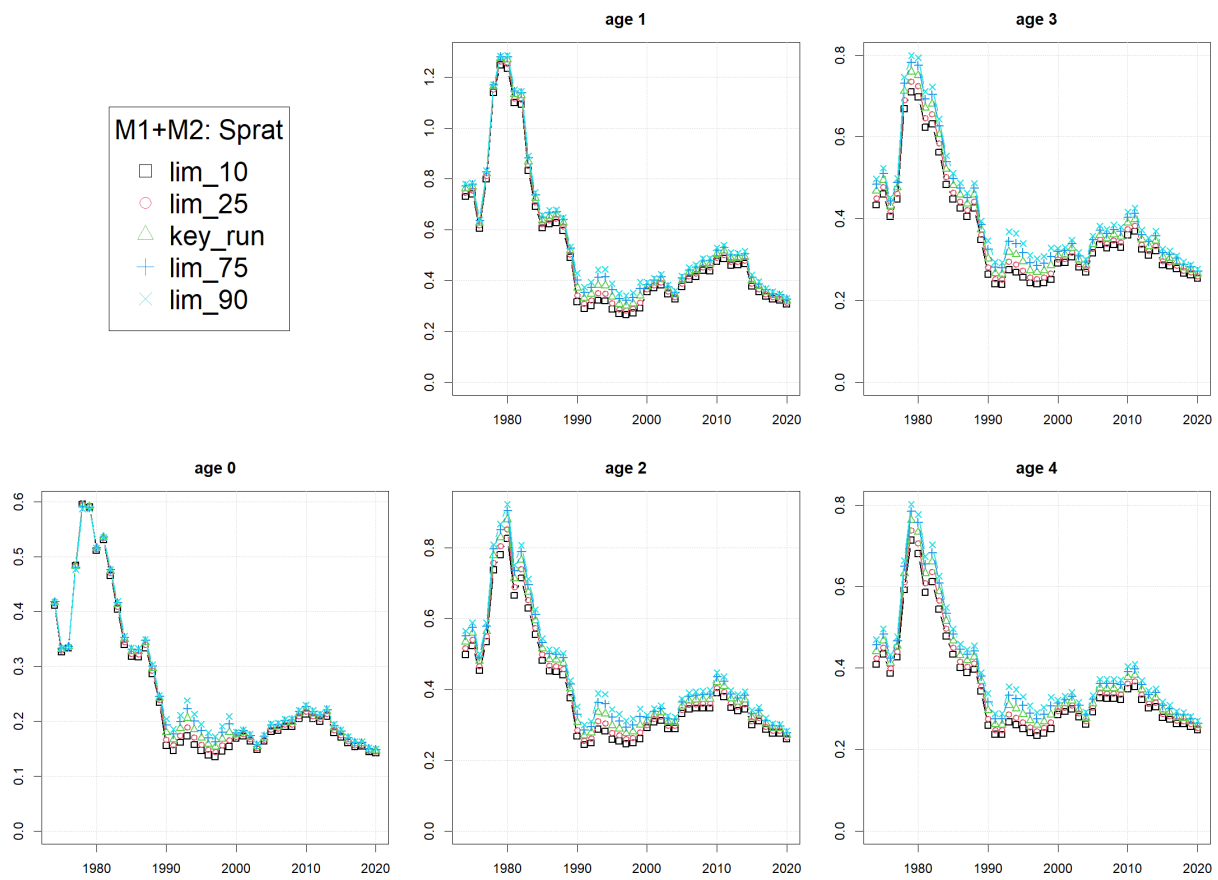


Figure 6: Natural mortality ( $M = M1 + M2$ ) for the five runs with variable cod consumptions.

### Variable M1 for herring

The effects of an increasing M1 for herring is a higher recruitment and SSB, and a lower F (Figure 7). Sprat has unchanged M1 values in the runs and the indirect effect of the changes for herring is very small (Figure 8).

A lower M1 for herring gives a lower M ( $M=M1+M2$ ) for herring (Figure 9) and an almost unchanged M for Sprat (Figure 10).

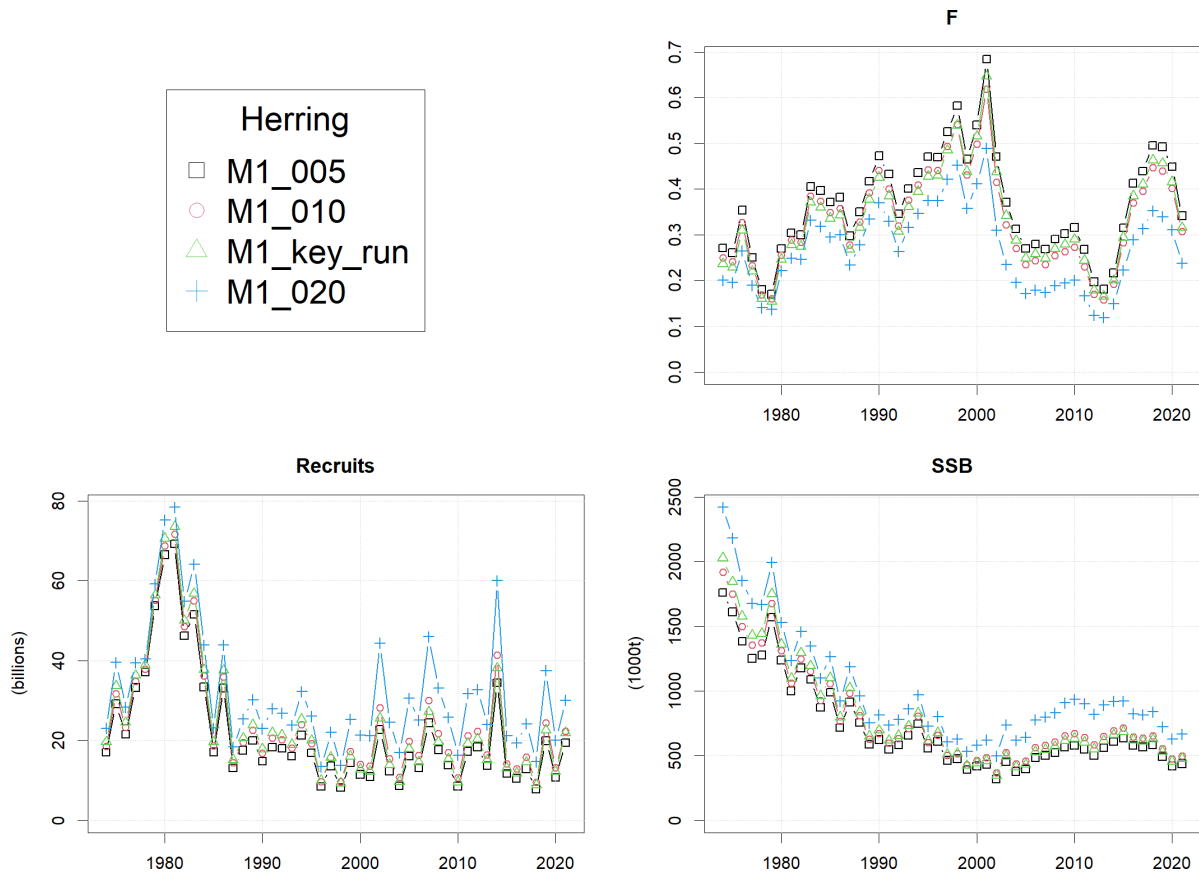


Figure 7: Assessment results for the four runs with variable M1 values.

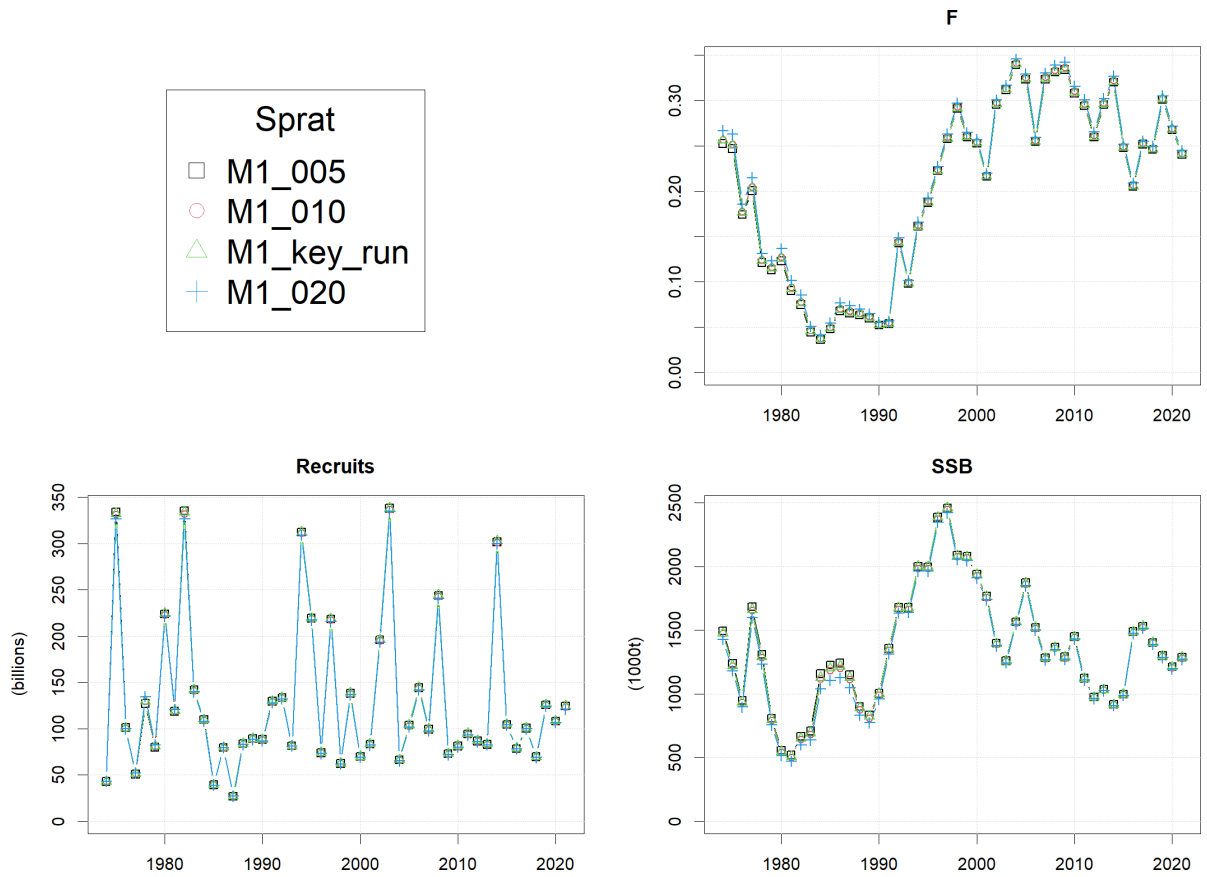


Figure 8: Assessment results for the four runs with variable M1 values.

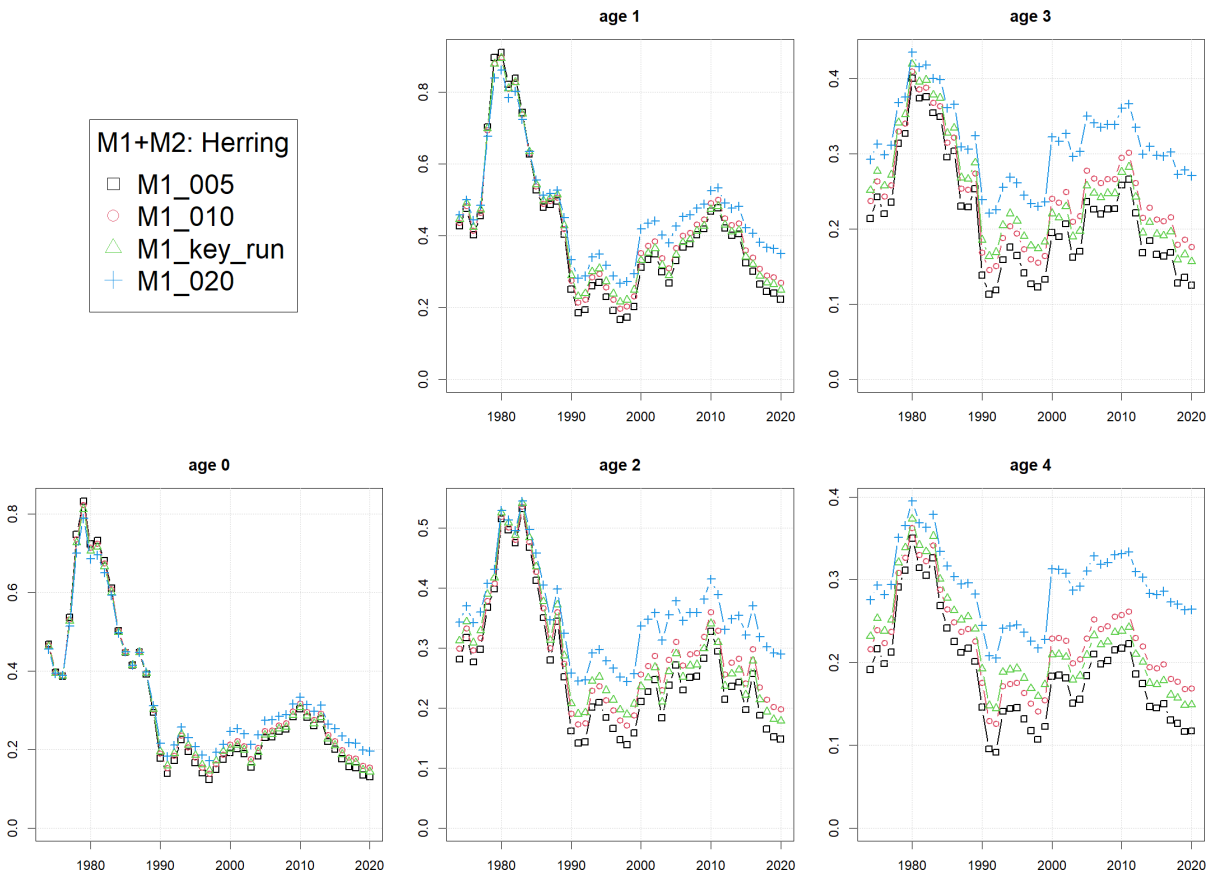


Figure 9: Natural mortality ( $M = M1 + M2$ ) for the four runs with variable M1 values.

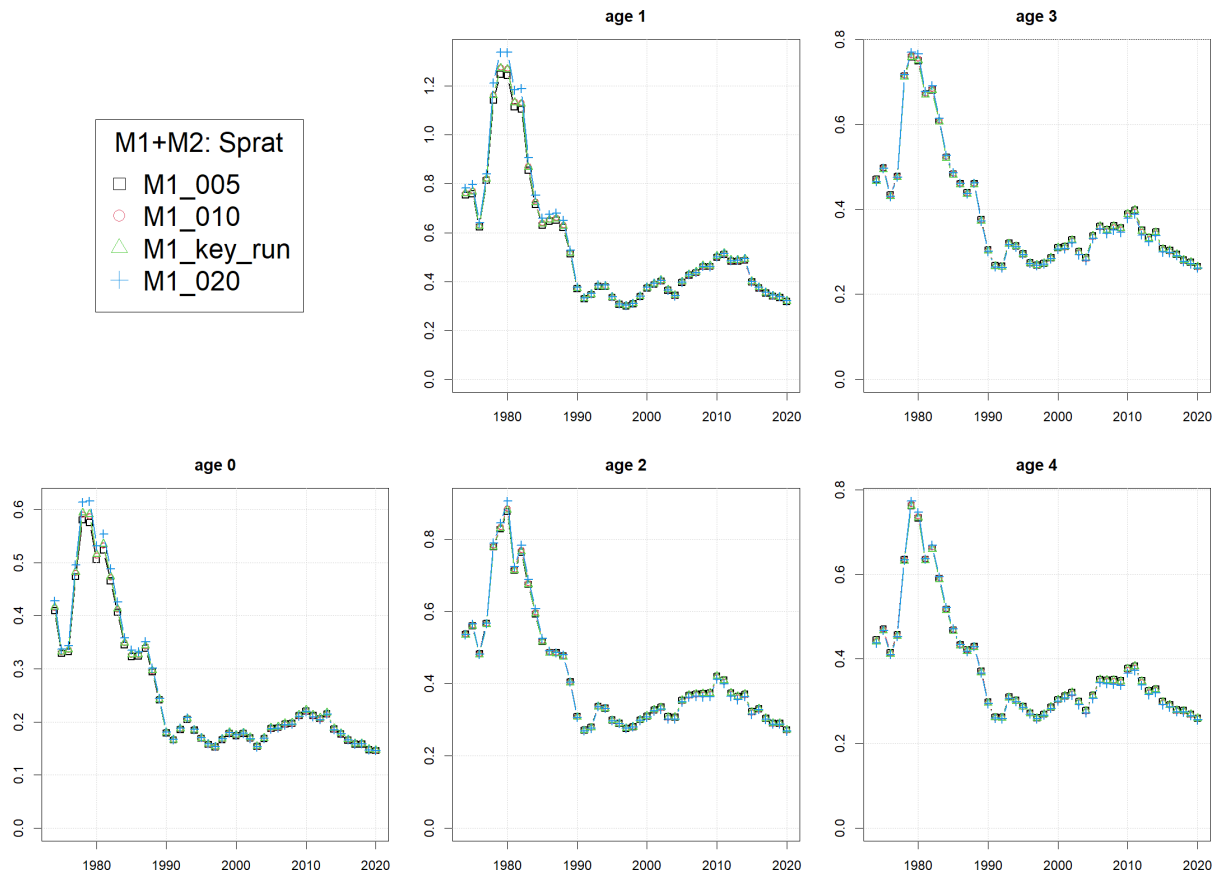


Figure 10: Natural mortality ( $M = M1 + M2$ ) for the four runs with variable M1 values.

### Likely values of M for herring

The model performance statistics (Table 2) show a limited difference in the likelihoods for the various model runs, and cannot in itself be used to classify the estimated M values as “likely” or not.

Total mortality (Z) is not sensitive to either total food consumption by cod or the M1 value for herring (Figure 11), but the highest Z are generally obtained by the highest food consumption (run “lim\_90”) and the lowest Z by the lowest consumption (run “lim\_10”) or lowest M1 (run “M1\_005”).

Catch curve analysis provides a very rough estimate of the total mortality (Z) as the assumptions for such analysis (that total fishing effort and catchability at age are constant over the years) are violated. The estimated Z values from catch curve analysis are presented in Figure 12 (Herring) and Figure 13 (Sprat). For both species a median Z around 0.6 is estimated for ages where the assumption of constant catchability might be met.

The Z estimated from the SMS runs are in the same range as the Z estimated from catch curve analysis. Comparing herring Z from the SMS key run and the Z from catch curve analysis (Figure 14) show that Z from catch curve analysis seems lower than the Z from the key run. This may be an effect of a not fully selected age 3 in the fishery. For ages 4 to 6 the two Z values are quite the same (if the obvious outliers of catch curve Z is ignored), however key run Z seems to be slightly higher than catch curve Z, indicating that runs with a lower Z (e.g. run “lim\_10” or “M1\_005” (Figure 16)) are the most appropriate. However the difference in model Z is small, especially for the age 4+ herring.

For sprat (Figure 15) the key run Z seems lower than key run Z for age 3. For age 4 and especially for age 5 the key run Z might be underestimated, if catch curve Z provides the real value.



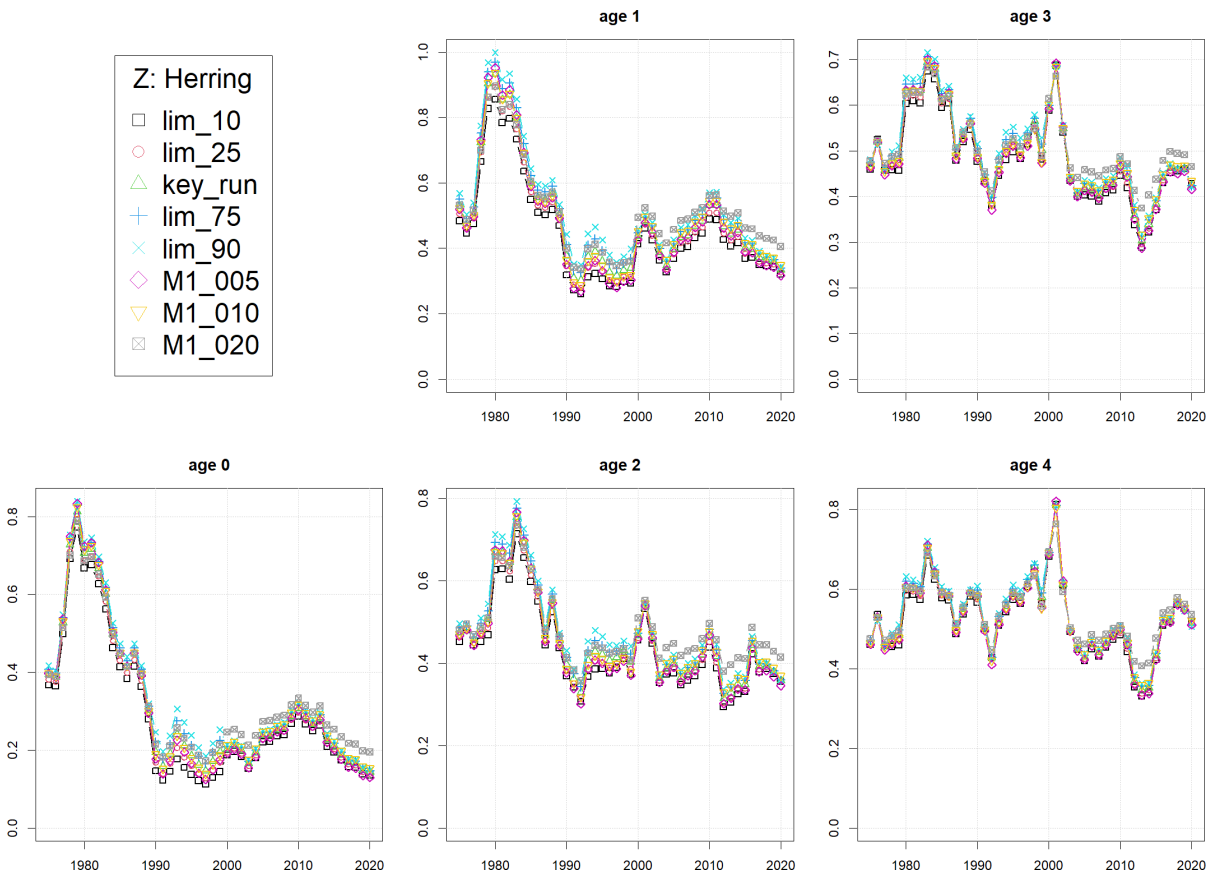


Figure 11: Total mortality ( $Z = M1 + M2 + F$ ) of herring by run.

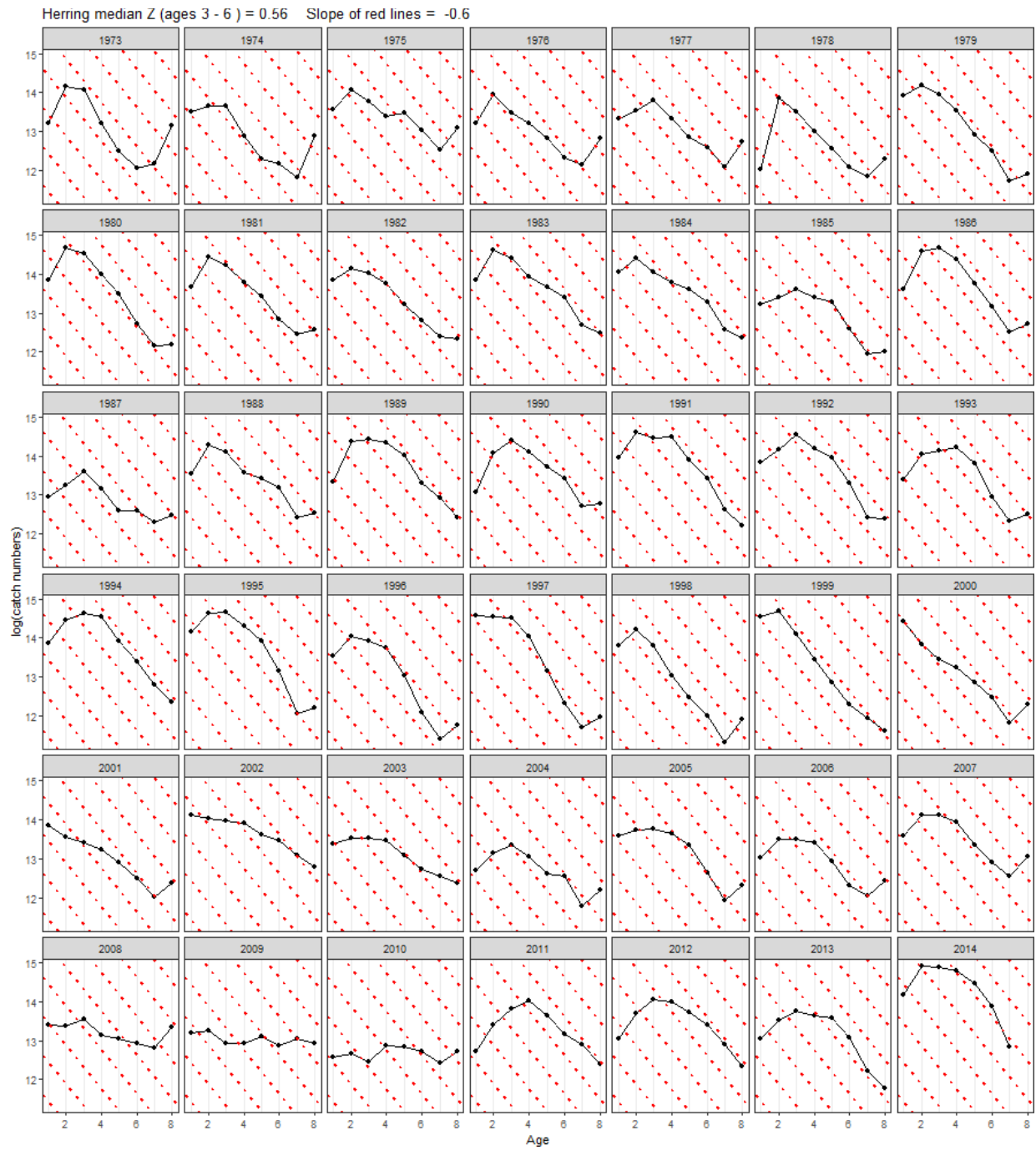


Figure 12: Total mortality ( $Z = M1 + M2 + F$ ) of Herring by year-class.

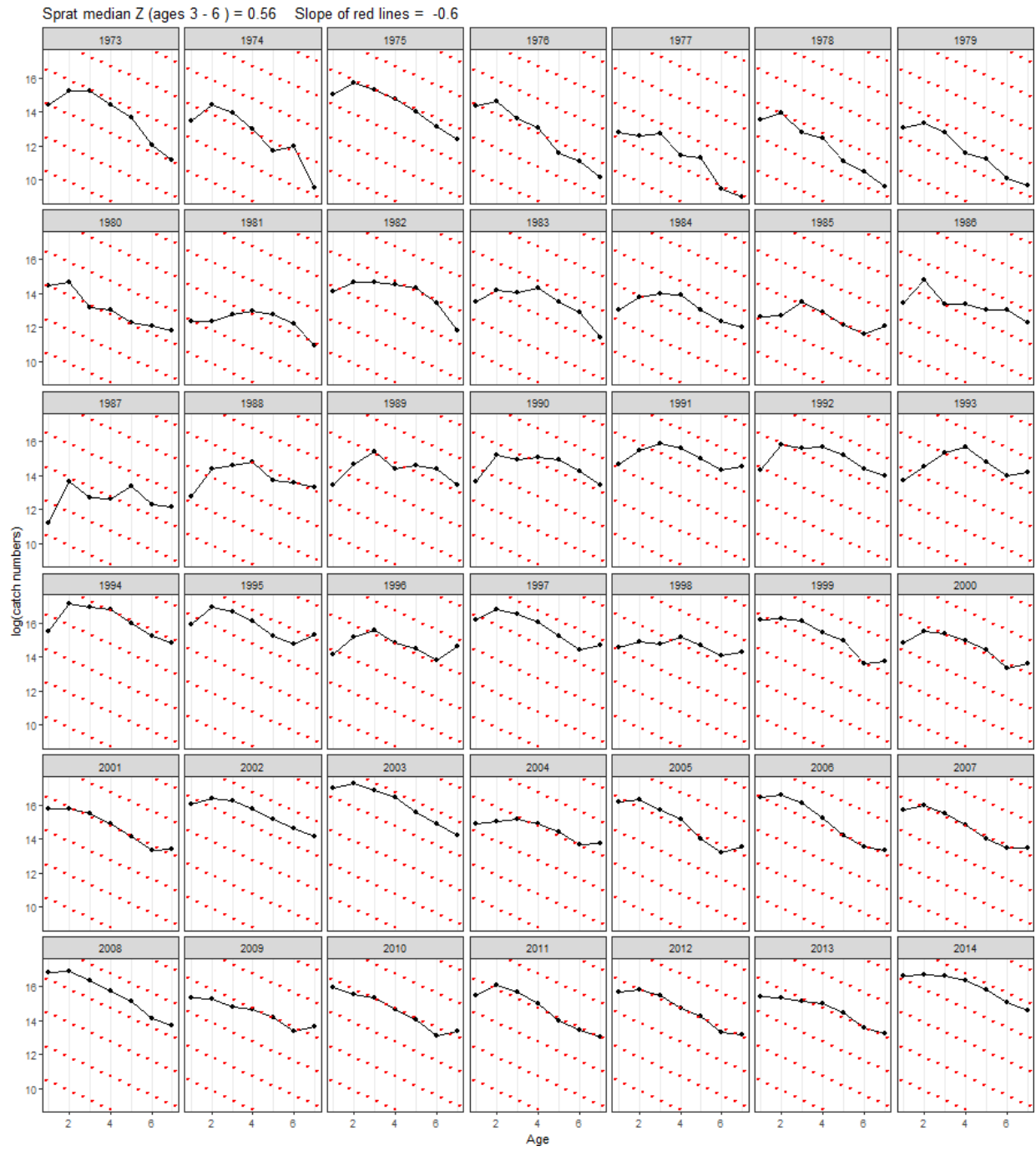


Figure 13: Total mortality ( $Z = M1 + M2 + F$ ) of Sprat by year-class.

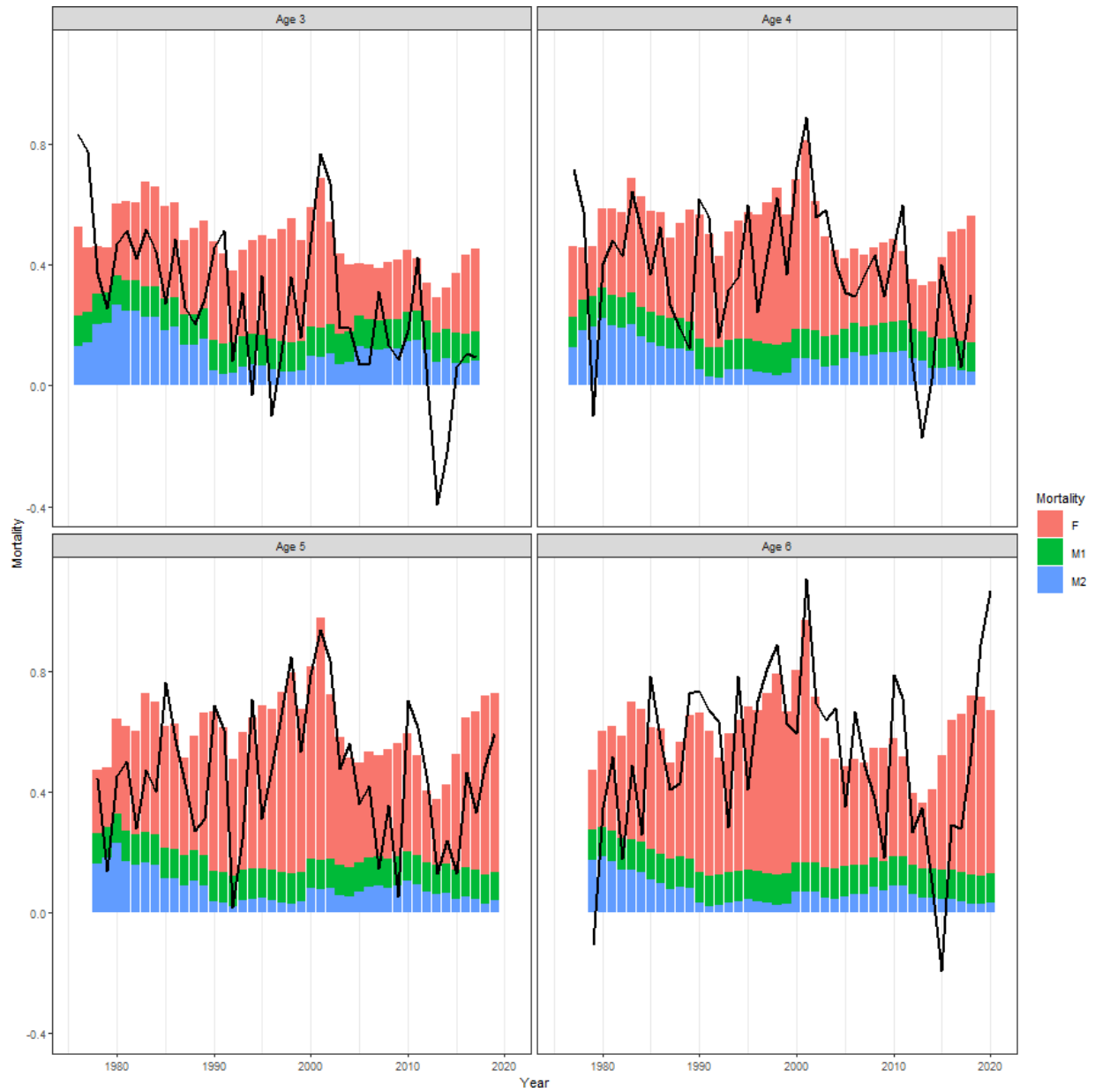


Figure 14: Total mortality ( $Z = M1 + M2 + F$ ) of Herring as estimated by the key run overlaid by Z (black line) estimated from catch curve analysis.

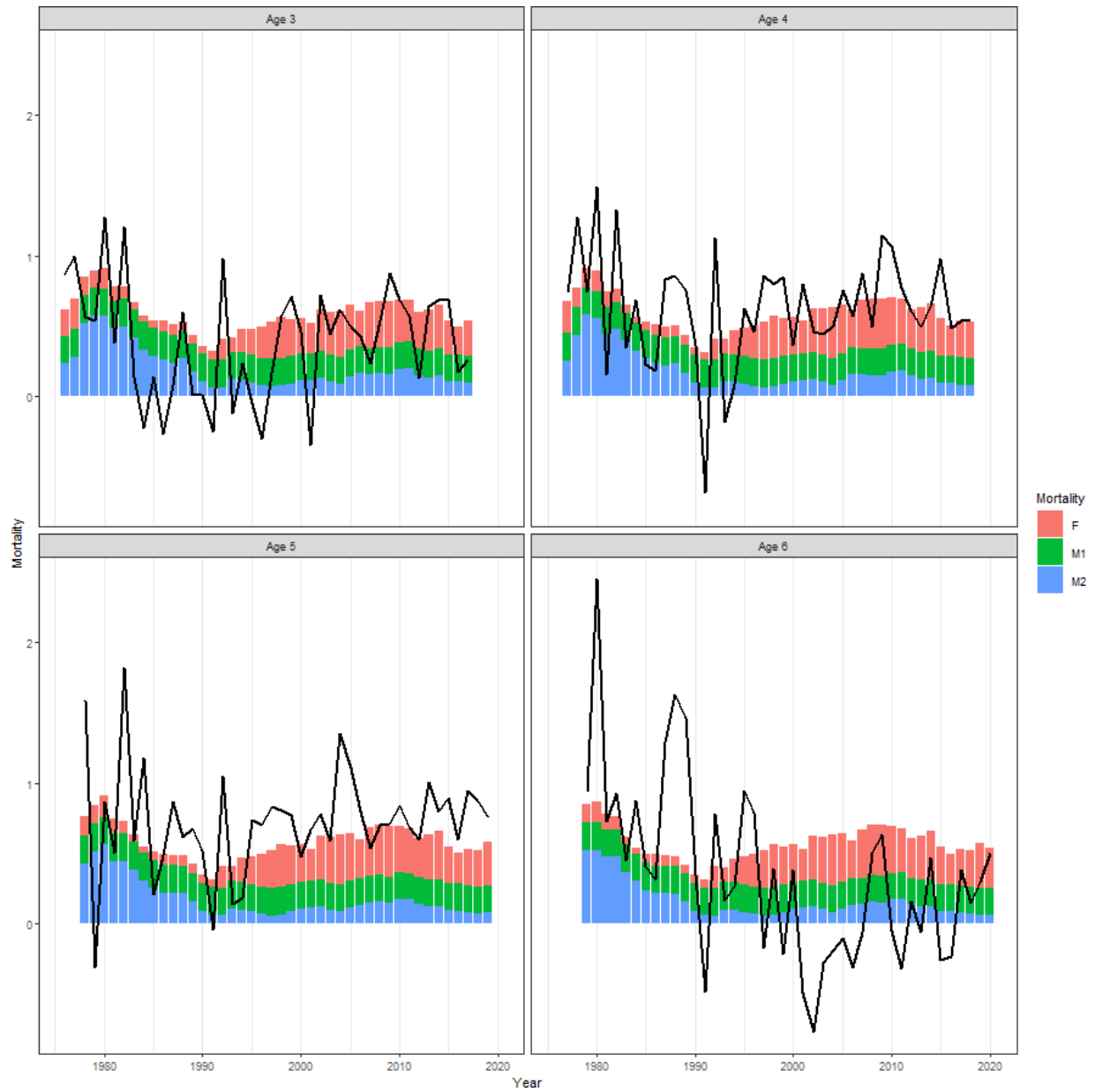


Figure 15: Total mortality ( $Z = M1 + M2 + F$ ) of Sprat as estimated by the key run overlaid by  $Z$  (black line) estimated from catch curve analysis.

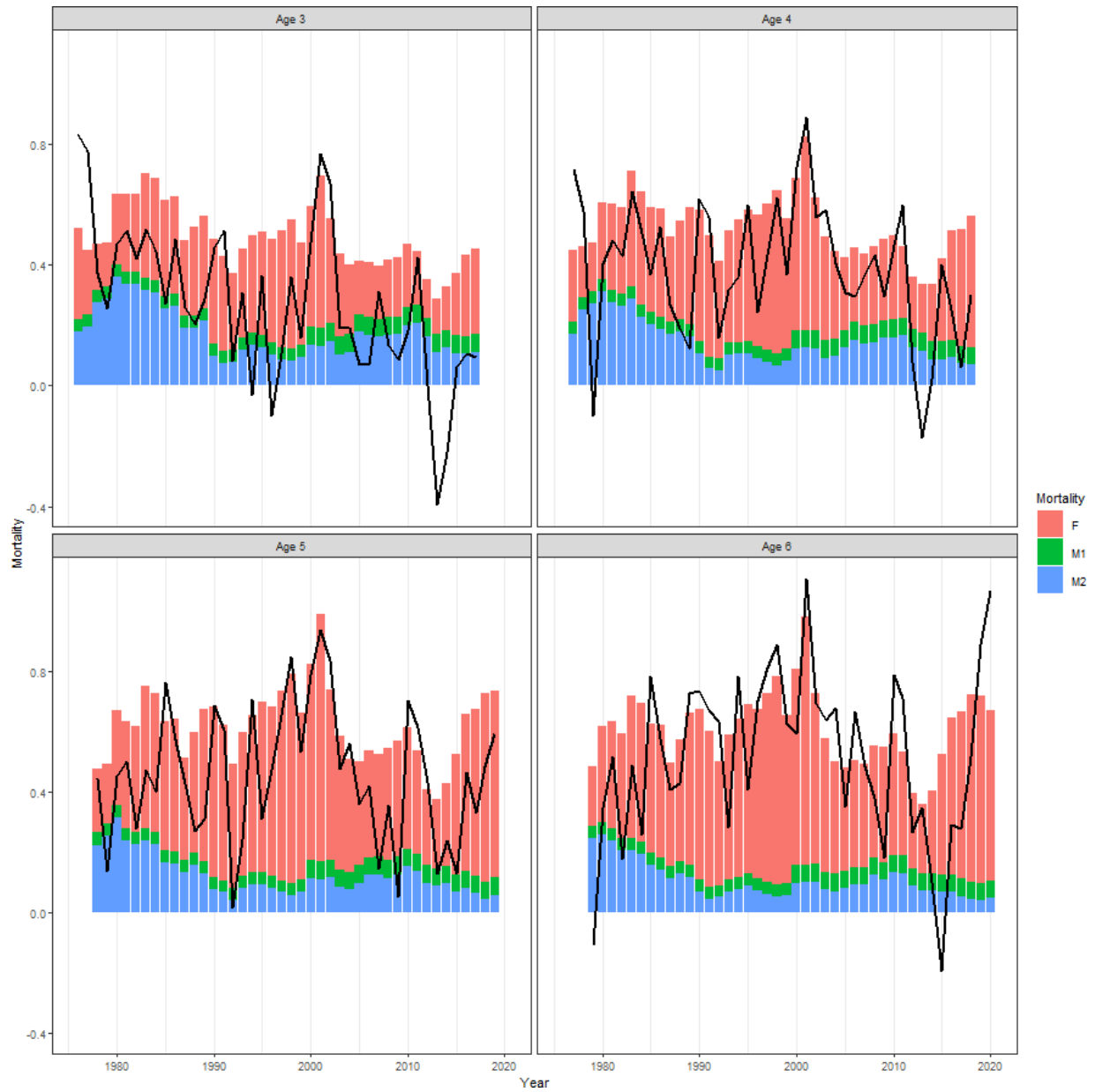


Figure 16: Total mortality ( $Z = M1 + M2 + F$ ) of Herring as estimated by the run "M1 - 005" overlaid by  $Z$  (black line) estimated from catch curve analysis.

## WD2: Correction of the catch data (CATON, CANUM) for herring (*Clupea harengus*) in subdivisions 25-29 and 32, excluding the Gulf of Riga (central Baltic Sea)

Author: Szymon Smoliński (National Marine Fisheries Research Institute, Gdynia, Poland)

BWKBALTPPEL 2023

Data provided by Denmark to the benchmark workshop represent old and corrected Danish catches from 1987 onwards (Table 1). Old XSA input files (CANUM, CATON from 2022 WGBFAS assessment; Table 2, Table 3) were used to recalculate CANUMs and CATON. Old CATONs were corrected using the difference between the old and corrected Danish catch time series. The ratio between the old and corrected CATON (including the corrected Danish catch time series) was used to up-scale or down-scale the CANUMs proportionally. For most years, the correction of old CATON (old XSA input) was within the range  $\pm 3.0\%$  (Fig. 1). Only in the year 2001, the CATON value was reduced by  $\sim 7.5\%$  in relation to the initial value. This year, Denmark reported initially 15786 tons, while the corrected value was 4462 tons, significantly lowering the contribution of Danish catches and lowering CBH CATON. With respect to the interannual variation of the time series, the corrected CANUMs were similar to the initial time series and differences were visible only in certain years (e.g., 1987, 1988, 2001, 2006) and particular age groups dominating in the catches (Fig. 2). Corrected CATON and CANUM are given in Table 3 and Table 4.

**Table 1.** Time series of old and corrected Danish catches [t].

Year	Old catches (WGBFAS)	Corrected catches
1987	4158	11003.25
1988	10794	17617.82
1989	7313	7877.817
1990	4596	3640.737
1991	6789	6722.852
1992	8091	8567.505
1993	8851	11857.88
1994	11250	11105.78
1995	11423	10650.91
1996	12148	10718.13
1997	9397	8451.445
1998	13876	12236.26
1999	6185	5979.503
2000	15786	14440.94
2001	15786	4461.929
2002	4557	3679.322
2003	5339	3872.561
2004	175	2319.818
2005	3053	2554.778
2006	136.5655	3300.921

2007	1351.554	1111.613
2008	1249.717	1457.892
2009	1463.009	2994.976
2010	5367.321	5850.785
2011	1848.187	3626.555
2012	1415.259	2048.701
2013	3419.226	2949.215
2014	2722.546	4504.731
2015	331.7174	844.155
2016	4040.493	2625.929
2017	9341.661	6252.979
2018	11367.51	7740.911
2019	8853	5371.08
2020	9275.421	6717.216
2021	6625.247	6625.247

**Table 2.** CANUM - catch in numbers by age groups (1-8+) – old XSA input (WGBFAS).

Year	1	2	3	4	5	6	7	8+
1974	2436300	1553800	1090600	1347900	483100	343500	619000	285100
1975	1861800	1229200	1405600	829900	870700	364000	274800	546800
1976	2093100	1114800	1034000	907300	476800	558500	246500	494400
1977	1258500	1825900	773600	608300	621700	365300	284000	545400
1978	1044000	1298700	1575100	436800	355100	370700	186800	478300
1979	405300	1195500	873200	1159500	338900	278700	281200	478500
1980	1037000	907100	977400	524600	654900	182500	204400	550500
1981	1325500	1523500	680000	615000	343600	436300	146600	527500
1982	867000	2277000	810100	334200	312000	188100	250500	420700
1983	744300	1698700	1875700	625300	233100	245700	162500	433400
1984	822000	1177900	1282900	1145700	374300	165500	166300	421100
1985	1237800	2124100	1076100	867300	707200	240300	131000	346900
1986	552824	1733617	1601914	838843	614707	320221	114772	208901
1987	920000	726000	1445000	1237000	607000	461000	238000	194000
1988	474000	2091300	746300	1009600	849400	354300	254200	210100
1989	792900	540600	1988300	580000	840700	695100	266500	336600
1990	643300	1194800	585500	1245900	419400	541100	370500	306000
1991	372900	1571700	1286100	512700	807700	278400	265900	238200
1992	1112600	1139400	1696900	702900	324100	422300	157700	218600
1993	826300	1852600	1503000	1473400	615700	274000	197500	140100
1994	486870	1138560	1559930	1068900	1057400	495520	213790	282450
1995	820500	960200	1742700	1555400	645700	440400	205200	212100
1996	985800	1441300	1095900	1216600	798100	492000	301100	223800
1997	549200	1350300	1738700	1173900	904800	492600	244200	186100
1998	1873286	947360	1810804	1781642	813071	481770	211361	186102

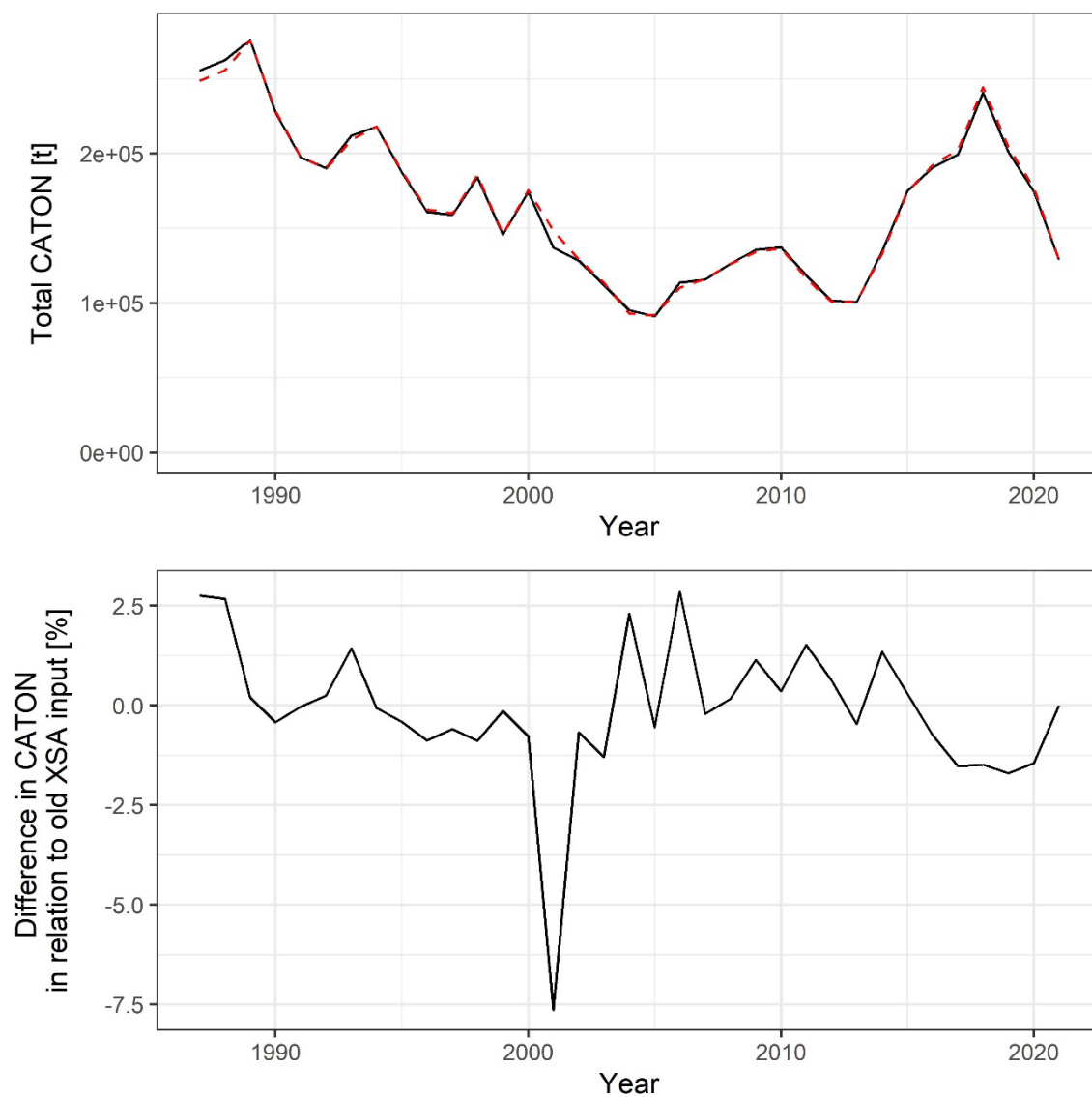


1999	628815	1660328	949293	1307772	950155	340256	185943	119952
2000	1842170	940000	1682170	818970	864530	567220	191280	185030
2001	1052466	1930067	605055	1010660	375834	391122	303247	199646
2002	1034640	1012975	1339851	456838	522442	179710	169851	230139
2003	1347364	782607	687478	686673	261252	226812	89925	202367
2004	656630	1242941	673629	568055	384598	162350	119700	129883
2005	326272	753498	1187077	557148	378447	219723	82530	159318
2006	808387	505592	754016	1104978	409059	264865	154493	147666
2007	457582	920291	630258	703185	823805	268661	135977	112019
2008	789388	735511	968418	461494	485798	711012	165897	215625
2009	653043	1395081	745935	855049	302486	340499	486075	239340
2010	546352	645269	1357314	661735	630229	283763	283721	362390
2011	293118	568892	770797	1130531	415505	312765	128881	235287
2012	333355	317009	416640	517743	642002	234424	160708	208441
2013	470327	655679	260040	410703	467439	403588	172879	224139
2014	470062	902642	1003705	385671	488077	409753	285297	250759
2015	1415576	745130	1264634	1252762	378036	384811	369954	473420
2016	602141	3014945	934748	1188734	838456	331740	465961	629002
2017	983743	823614	2898360	840730	923686	527598	248465	411819
2018	1737640	1280367	1174100	2637412	789008	663989	398905	335250
2019	416846	1561422	1127576	891782	1957135	485302	396557	239356
2020	1644919	781308	1423813	788676	662488	1080601	199821	228471
2021	691437	1805171	831906	867236	519655	377932	373009	129976

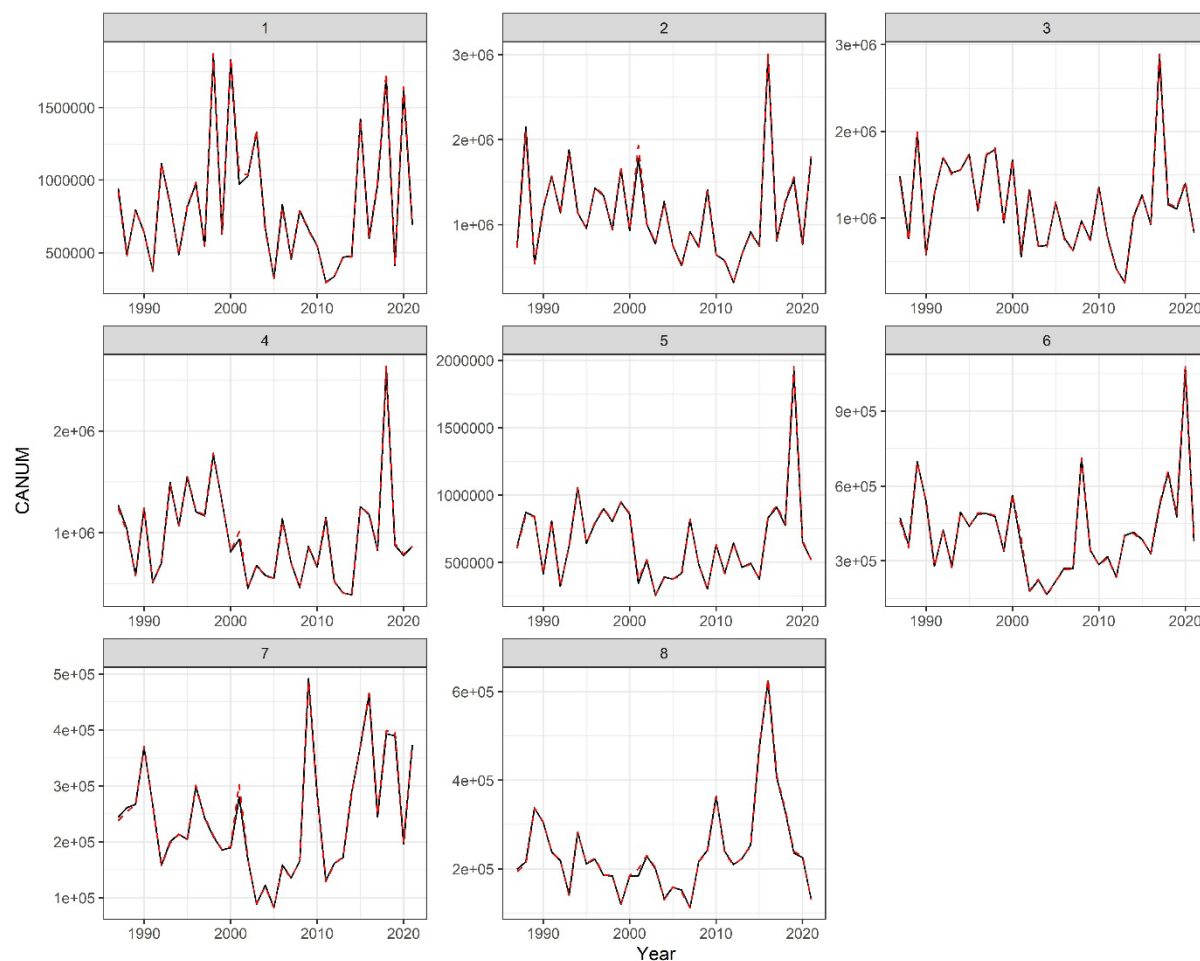
**Table 3.** CATON – catch in tons – old XSA input (WGBFAS).

Year	CATON
1974	368652
1975	354851
1976	305420
1977	301952
1978	278966
1979	278182
1980	270282
1981	293615
1982	273134
1983	307601
1984	277926
1985	275760
1986	240516
1987	248653
1988	255734
1989	275501
1990	228572

1991	197676
1992	189781
1993	209094
1994	218260
1995	188181
1996	162578
1997	160002
1998	185780
1999	145922
2000	175646
2001	148404
2002	129222
2003	113584
2004	93006
2005	91592
2006	110372
2007	116030
2008	126155
2009	134127
2010	136706
2011	116785
2012	100893
2013	100954
2014	132700
2015	174433
2016	192056
2017	202517
2018	244365
2019	204438
2020	177079
2021	128961



**Fig. 1.** Total catch in tons (CATON) of the Central Baltic herring (upper panel) with indicated old (red, dashed line) and corrected (black, solid line) values. Difference between old and recalculated CATON (lower panel).



**Fig. 2.** Total catch in numbers (CANUM) of the Central Baltic herring by age group with indicated old (red, dashed line) and corrected (black, solid line) values. Note that the lines are strongly overlapping.

Table 3. Corrected CATON – catch in tons.

Year	Caton
1974	368652
1975	354851
1976	305420
1977	301952
1978	278966
1979	278182
1980	270282
1981	293615
1982	273134
1983	307601
1984	277926
1985	275760
1986	240516

1987	255498
1988	262558
1989	276066
1990	227617
1991	197610
1992	190258
1993	212101
1994	218116
1995	187409
1996	161148
1997	159056
1998	184140
1999	145717
2000	174301
2001	137080
2002	128344
2003	112118
2004	95151
2005	91094
2006	113536
2007	115790
2008	126363
2009	135659
2010	137189
2011	118563
2012	101526
2013	100484
2014	134482
2015	174945
2016	190641
2017	199428
2018	240738
2019	200956
2020	174521
2021	128961

**Table 4.** Corrected CANUM – catch in numbers.

Year	1	2	3	4	5	6	7	8+
1974	2436300	1553800	1090600	1347900	483100	343500	619000	285100
1975	1861800	1229200	1405600	829900	870700	364000	274800	546800
1976	2093100	1114800	1034000	907300	476800	558500	246500	494400
1977	1258500	1825900	773600	608300	621700	365300	284000	545400
1978	1044000	1298700	1575100	436800	355100	370700	186800	478300

1979	405300	1195500	873200	1159500	338900	278700	281200	478500
1980	1037000	907100	977400	524600	654900	182500	204400	550500
1981	1325500	1523500	680000	615000	343600	436300	146600	527500
1982	867000	2277000	810100	334200	312000	188100	250500	420700
1983	744300	1698700	1875700	625300	233100	245700	162500	433400
1984	822000	1177900	1282900	1145700	374300	165500	166300	421100
1985	1237800	2124100	1076100	867300	707200	240300	131000	346900
1986	552824	1733617	1601914	838843	614707	320221	114772	208901
1987	945327	745986	1484780	1271054	623710	473691	244552	199341
1988	486648	2147103	766214	1036539	872065	363754	260983	215706
1989	794526	541708	1992376	581189	842424	696525	267046	337290
1990	640611	1189807	583053	1240693	417647	538839	368952	304721
1991	372775	1571174	1285670	512528	807430	278307	265811	238120
1992	1115394	1142261	1701161	704665	324914	423360	158096	219149
1993	838183	1879241	1524614	1494588	624554	277940	200340	142115
1994	486548	1137808	1558899	1068194	1056701	495193	213649	282263
1995	817134	956260	1735550	1549018	643051	438593	204358	211230
1996	977130	1428624	1086262	1205900	791081	487673	298452	221832
1997	545954	1342320	1728425	1166963	899453	489689	242757	185000
1998	1856752	938998	1794821	1765917	805895	477518	209495	184459
1999	627929	1657990	947956	1305930	948817	339777	185681	119783
2000	1828063	932802	1669288	812699	857910	562876	189815	183613
2001	972157	1782792	558886	933541	347156	361277	280108	184412
2002	1027613	1006095	1330751	453735	518894	178489	168697	228576
2003	1329969	772503	678602	677808	257879	223884	88764	199754
2004	671773	1271605	689164	581155	393467	166094	122460	132878
2005	324497	749399	1180620	554117	376388	218528	82081	158451
2006	831563	520087	775634	1136658	420787	272459	158922	151900
2007	456636	918388	628955	701731	822101	268105	135696	111787
2008	790691	736725	970016	462256	486600	712185	166171	215981
2009	660502	1411015	754455	864815	305941	344388	491627	242074
2010	548284	647551	1362114	664075	632458	284767	284724	363672
2011	297582	577555	782534	1147746	421832	317528	130844	238870
2012	335448	318999	419256	520994	646033	235896	161717	209750
2013	468137	652626	258829	408791	465263	401709	172074	223095
2014	476375	914765	1017185	390851	494632	415256	289129	254127
2015	1419735	747319	1268349	1256442	379147	385941	371041	474811
2016	597706	2992739	927863	1179979	832280	329297	462529	624369
2017	968739	811053	2854156	827908	909598	519551	244676	405538
2018	1711852	1261365	1156675	2598270	777298	654135	392985	330275
2019	409746	1534828	1108371	876593	1923802	477036	389803	235279
2020	1621155	770021	1403244	777282	652917	1064990	196934	225170
2021	691437	1805171	831906	867236	519655	377932	373009	129976

### **WD3: Notes on sampling design, data storage and data edits underlying the Swedish data biological data on commercial caught herring (her.27.25-2932) provided to the 2023 ICES Benchmark of Central Baltic Herring (BWKBALTPEL)**

Nuno Prista, Carina Jernberg, Katja Ringdahl

Swedish University of Agricultural Sciences, Department of Aquatic Resources,  
Institute of Marine Research, Turistgatan 5, SE-453 30 Lysekil, Sweden

#### **Summary**

This document provides supplementary notes and methodological details related to Swedish commercial data on Central Baltic Herring (her.27.25-2932) provided to the 2023 ICES Benchmark Workshop on Baltic Pelagic Stocks (BWKBALTPEL). Aspects such as changes through time in sampling designs used to select the samples and the methods used in biological analyses are detailed. Details are also given on data edits made prior to the submission, procedures used to fill missing values and a set of additional biological variables that were not requested but are available for exploration in future benchmarks. Finally, a set of analysis are highlighted for future improvement of biological estimation of commercial catches of the stock.

#### **Introduction**

In the context of the data preparation towards the 2023 ICES Benchmark Workshop on Baltic Pelagic Stocks (BWKBALTPEL), a data call was sent to countries requesting data on individual maturity at age and length, and weight at age and length from central Baltic herring (her.27.25-2932) by ICES statistical rectangle/subdivision for the period 1984 – 2021. The request was issued with the aim of carrying out analysis on time series of i) maturity at age and length and ii) weight at age and length with data being requested in a pre-specified format (see annex).

When fisheries data is requested over a long-time span on a format with a limited number of predefined mandatory/optional fields and code-lists, aspects related to the evolution of data collection over time tend to become implicit and not readily visible to end-users. Some of those aspects are worth considering in the context of data analysis. Examples of these are:

- National and EU data collection regulations evolving through time and sampling designs and objectives for data collection programmes changing accordingly;
- Lab methodologies used in age and maturity determination being updated through a combination of national and international (ICES) efforts;
- Increased digitalization also taking place: where initially sampling records were logged in lab notebooks and earlier versions of modern spreadsheets subjected to manual quality checking, presently data are entered directly into electronic databases and with inherent consistency checks and quality controls.

Furthermore, when preparing the final data set several decisions are made by data submitters that may condition both the quantity and quality of data available to end-users. Some of these decisions restrict or alter the original data potentially changing their interpretation. Examples of these decisions are:

- Decisions taken on the interpretation of historical codes (e.g., how convert them to the codes requested in the present format?);
- Decisions on how to handle atypical observations (e.g., how to identify them? keep them or remove them?)
- Decisions on how to deal with periods or sets of data that may have low degree of quality relative to other more consistent ones (e.g., which data are considered good-enough to be provided? which are not?)

Finally, it should be noted that over the years significant changes in the IT systems have taken place and that transitions between systems, changes in database structures, sometimes alter data themselves and/or change their interpretation. Over the requested time span, some of those changes took place. As an example, up to the 1980s several different files and databases existed and, at the end of 1980s, a first central database started to be built (FiskData). That database was developed in UNIQUE and modified through time, e.g., to incorporate the sampling of discards onboard commercial vessels (2002). In 2005 a decision was taken to create an ORACLE database (FiskData2) and migrate data from the original database to it. Several system updates and upgrades were made thereafter. Each of these system changes constitutes a potential source of error.

When combined across countries and data submission processes, the abovementioned aspects significantly influence the quality of data provided, generating biases (or in a best-case scenario, just variance) in final analysis. These biases and variance will be difficult to control for and inspect given the considerable number of data points being analyzed without at least some indications of where (*sensu* which variables) the potential issues may be. ICES is currently working towards improved documentation of the potentially problematic sampling details alongside the data themselves (RDBES). It is expected that will facilitate automation of data evaluations and bias considerations in the future. Until that system is available, it is the present authors hope that a structured supplementary document like the present document will help.

The present document summarizes the main supplementary information that, from the present authors' perspective, may be of interest considering when analyzing Swedish data on Central Baltic Herring submitted to WKBBALTPEL. Aspects covered include data availability, sampling design and methods of biological analysis, but also some details on the methods used to handle atypical and missing values. Information on availability of a few additional biological variables that were not requested but may be of exploratory interest in future benchmarks is also provided.



## Data availability, data storage and structure

Lab records indicate that central Baltic herring has been sampled in Sweden since at least 1908. Records up to 1952 have some gaps (some years missing) but from 1952 onwards, all years are represented in the archives. From the 1970s onwards, data collection is considered consistent and better documentation of the trips sampled exists. WGBFAS assessment of the central Baltic herring stock presently uses age structure information back to 1974 indicating that data as early as that year were considered good-enough for assessment by earlier colleagues. Individual samples readily available in present database go back to 1987. Some indications exist that data from 1985 and 1986 has also been digitized but failed migration to between the old and the present database system.

## Sampling design

Three sampling programmes have taken place during the period of BWKBBALTPEL request:

- a general sampling programme targeting commercial landings of herring and sprat (internally known as type 4 sampling),
- a supplementary sampling programme on landings of these species destined for industry use (internally known as type 8 sampling),
- more recently, a randomized haul-based sampling programme of catches (internally known as type 7 sampling).

## General sampling programme on commercial landings (1977-2021)

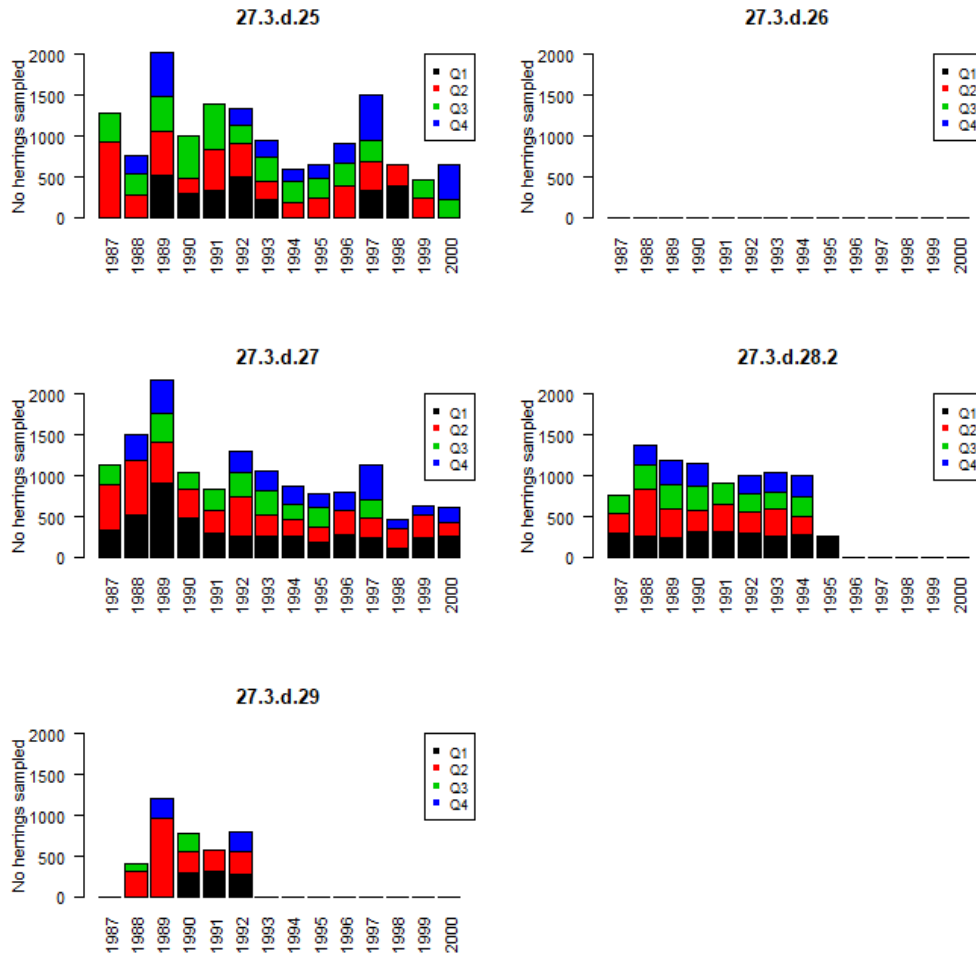
This sampling programme can be considered a continuation of the sampling started in the early 1900s. Before 1977 sampling designs and methods for lab analysis are sparsely documented or difficult to retrieve. From 1977 onwards, sampling and lab protocols are better known, and documentation is more readily available. Within the latter period, two different sampling strategies were carried out.

### *1977-2000: length stratified sampling of sorted landings*

Herring samples were collected one to two times each quarter at subdivision level (SD 24 to 31). In each sampling occasion fully sorted landings of herring destined (mostly) to human consumption were purchased. Landings sampled came mostly from the bottom trawl fishery (with some exceptions).

The size categories of each trip were treated as individual samples. The samples from largest sizes were fully processed; the samples from smaller sizes categories (size 4 and above) were subsampled (half the box) except when the proportion of sprat exceeded 70-80% in which case the entire box was sorted to maximize the number of herrings sampled.

The general aim was to measure 200 individuals per sample, but in the case of wide spread of the length groups a larger number of individuals were selected. From the 70s to mid-80s, individuals were biologically sampled in proportion to their representation in length distribution curve of the sample. From mid-80s onward, a fixed number of  $n = 5$  herrings per half-cm group was selected for biological analysis.



**Figure 1. Annual and quarterly distribution of herring sampled in general sampling programme on commercial landings (1977-2000).**

*2001-2021: random sampling from unsorted landings*

Samples were bought from unsorted landings from each subdivision (SD 23-29S). Samples comprised herring and/or sprat and were taken from the pelagic trawl fishery. Samples were

sorted to species and species fractions weighed. Individuals from each species were either censused or randomly sampled. The aim was to attain ca. 400 individuals per quarter and area.

Trips were considered as individual samples. 50-100 individuals were randomly selected from each species fraction for biological analysis

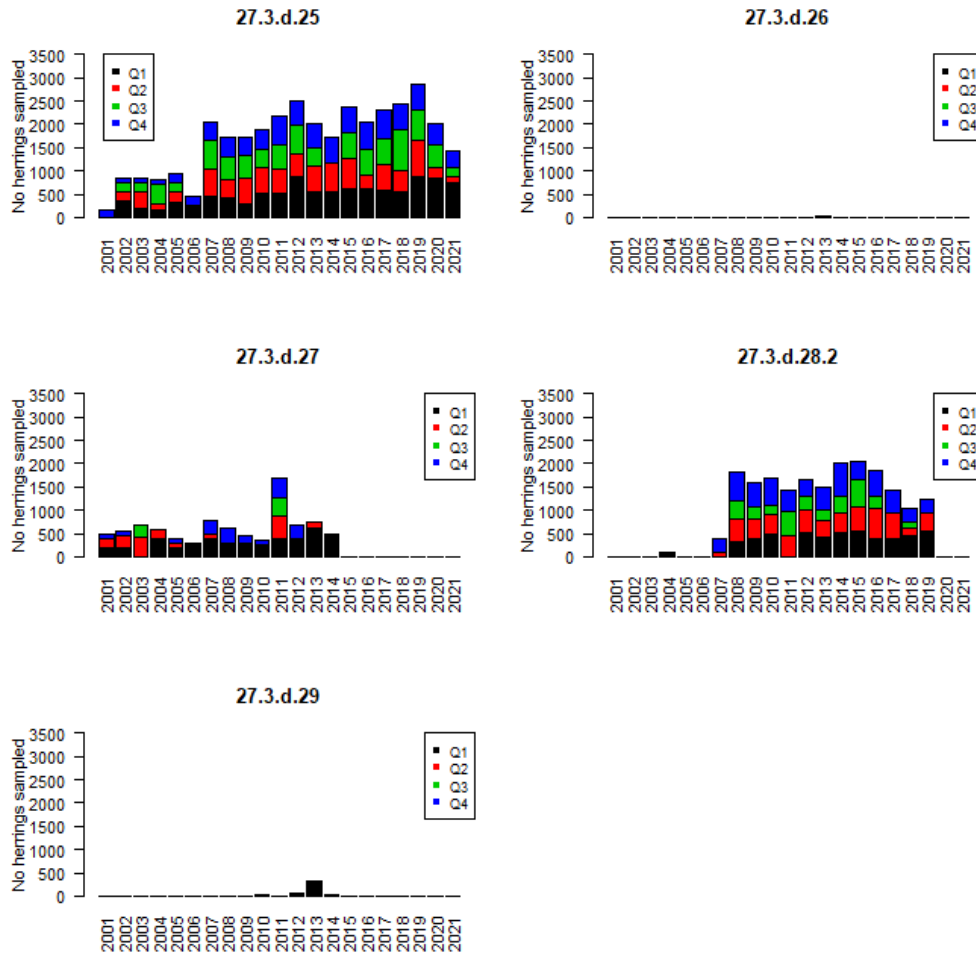


Figure 2. Annual and quarterly distribution of herring sampled in the general sampling programme on commercial landings (2001-2021).

**Supplementary sampling programme on landings destined for industry use (1993 to 2020)**

Herring samples collected from landings destined for industrial usage. Different methods were used through time. Between 1993 and 2001, industrial landings caught in the Baltic were sampled on an irregular basis either when landed at Rönning (Skagerrak coast) or after transport to a factory in Ängholm. When transported the cargo of multiple trips was mixed making it difficult if not impossible to associate e.g., individual fish with individual trips or subdivisions. Starting

in 1998 sampling increasingly took place in one port (Västervik, Baltic coast) and from 2002 onwards further extended to some other ports of the Baltic coast - Nordersund and Simrishamn (2002-2006) and Ronehamn (2004-2006) - and also to Denmark (Skagen, 2006-2019) where major industrial landings take place. To achieve this, various local actors were involved in the sampling (fishers, 1st hand buyers, and control authorities) making sampling hard to standardize. The Danish Technical University (DTU) sampled Swedish vessels landing in Skagen. In general, n = 4 boxes of landings were taken from a single landing at one time per month, with little possibility of reliably associating a particular subdivision to it given the multi-subdivision nature of some of the industrial trips. At the lab, the boxes were pooled together and treated as individual samples. After species sorting, sampling followed the same protocol as the in the “general sampling programme on commercial landings” over the same periods (length stratified to 2000; random from 2001 onwards). However, the poor condition of industrial samples made maturity frequently difficult to determine.

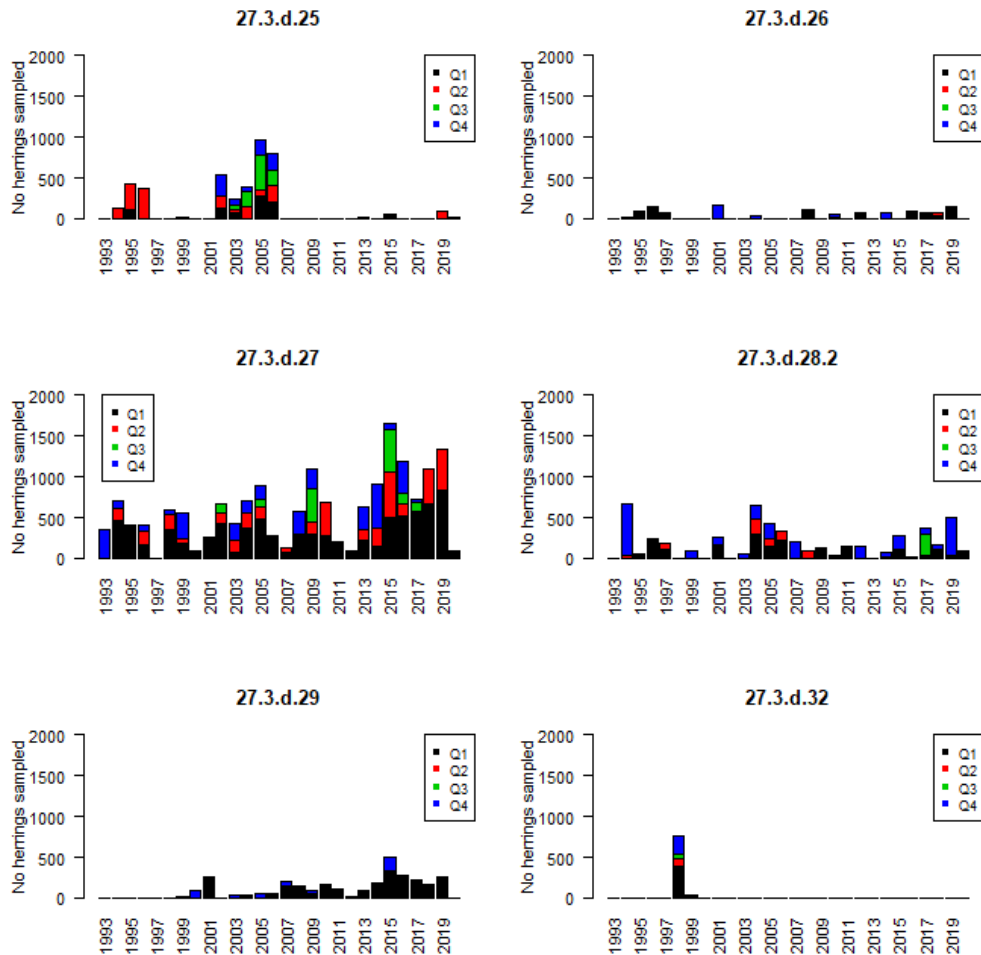


Figure 3. Annual and quarterly distribution of herring sampled in the supplementary sampling programme on landings destined for industry use (1993 to 2020).

### **Randomized haul-based sampling (2020-present)**

Following new requirements to increase probabilistic sampling of commercial fisheries<sup>1</sup>, RCG efforts on regional coordination of the sampling of small-pelagic fisheries at Baltic level (RCG NA NS&EA RCG Baltic, 2020), and new opportunities of cooperation with the industry, in 2020 Sweden implemented a new sampling scheme for the sampling of Baltic fisheries (catches for consumption and industrial use). The goal was to increase control over sample selection and improve spatial resolution of individual data (namely the certainty of subdivision of provenance when long fishing trips operated in multiple subdivisions) through the adoption a standardized protocol of weekly vessel selection and a close collaboration with the fishing industry that facilitated spatially precise haul-by-haul sampling. This effort was coupled to increased optimization of biological sampling and documentation of sample design (RCG NA NS&EA RCG Baltic, 2020). In brief, a set of vessels is now randomly selected each week and the skippers contacted to collect a box (3-5 kg) of unsorted catch from each haul in the next trip they make. The boxes are then shipped back to the lab along with their documentation (date, position, etc.) and biologically analyzed after potential subsampling by subdivision. Lab procedures involve the sorting of samples to species, weighing of each species fraction and then the measurement and biological analysis of max 50 individuals per species. The selection of individuals for biological sampling is not length-stratified nor dependent on pre-defined quotas: rather, successive quasi-random splits of the original sample are used to reach the approximately 50 individuals/sample desired for biological sampling.

<sup>1</sup> Regulation (EU) 2017/1004 of the European Parliament and of the Council of 17 May 2017 on the establishment of a Union framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common fisheries policy and repealing Council Regulation (EC) No 199/2008 (recast)

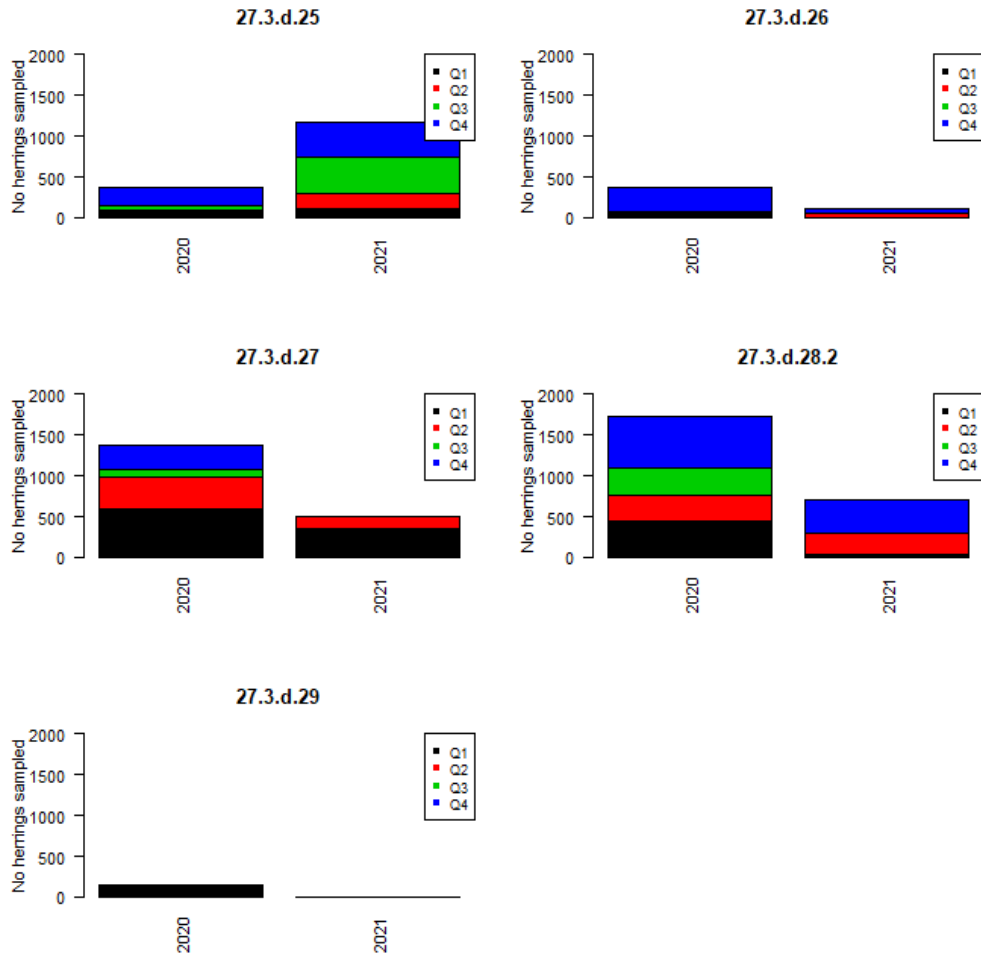


Figure 4. Annual and quarterly distribution of herring sampled in the randomized haul-based sampling programme (2020-2021).

### Biological analysis

Laboratory protocols for biological analysis were the same for all sampling programmes with minor evolution through time.

### # Individual length and weight

Individuals were measured in mm up to Q3-2000 and in semi-cm from Q4-2000 onwards. Individual weights were taken to the nearest 1g (before 2017), to the nearest 1g or 0.1g (2017) and to the nearest 0.01g (2018-2021).

### **# Sex and maturity**

Gender and maturity stage were recorded for all specimens. Up to 2000, the 8-point Hjort-scale was used to stage maturity of herring gonads except for the period between 1992-Q2 and 1993-Q4 when a 4-point scale was briefly used. From 2010 onwards an extended 8-point Hjort-scale was used: initially (2010-2011) the extended scale included an extra level "9" (abnormal gonad); from 2012 onwards a splitting of level 8 in two levels (regeneration and regression) was added. A formal maturity manual does not exist but standardized internal procedures and stage descriptions have been used. May Carlsson (among others) was the maturity reader from the 1960s to 1992. From 1992 onwards, Carina Jernberg assumed that responsibility.

### **# Age determination**

Until 1992, scales were collected from all individuals and age assessed based on microscopy analysis of those scales. In 1992, otoliths started to be collected and age assessed using them (whole otoliths observed under lens with reflected light against a black background). Scales kept being collected and used as support to the ageing of larger individuals. With time, the fish sizes where scales were collected were increased and in 2015, scale sampling was discontinued. From then onwards ages have been determined solely based only on otoliths. An internal qualitative scale for age quality has been used from 2013 onwards. From 2013 to 2020 the scale had four levels. From 2021 onwards WGBIOP age quality scores have been used. There is some internal documentation on scale reading and an internal manual on otolith reading. May Carlsson was the age reader from 1960s to 1992. From 1992 onwards, Carina Jernberg assumed that responsibility. Further details on the age determination are available from authors upon request.

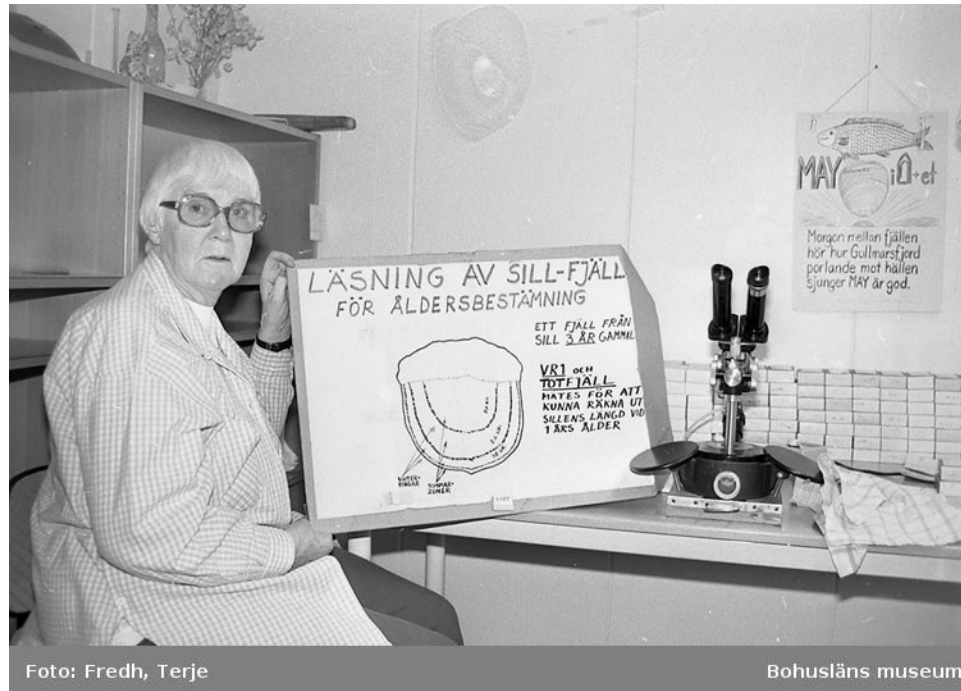


Figure 5. May Carlsson, main age reader of Baltic herring (photo: Terje Fredh, 1992; source: [Bohuslän Museum](#))

### # Other variables

Besides length, weight, age, sex, and maturity, other biological variables were collected through time:

- Fat content: data available in database from 1987-1993, and from 2012 to present (in five qualitative categories)
- *Ichthyophonus*: data available from dedicated study from 1992 to 1997 (presence or absence)
- Nematodes: data available in database all years (in four quantitative categories)
- Number of keeled scales and vertebrae: data available to the mid-90s (counts).

### Specific details on data preparation and interpretation

#### Time span and sampling programmes of data provided

Data provided to BWKBALTPPEL includes all readily available individual data from the General sampling programme on commercial landings (i.e., 1985-2021), the supplementary sampling



programme on landings destined for industry use (1993 to 2020), and the randomized haul-based sampling (2020-2021).

### **Variable by variable comments**

#### **Sampling\_type [M]**

Data available: variable does not explicitly exist in database

Procedure: set to "M"

Comments: "S" would be a better choice for some randomized data collected in 2020-2021.

#### **Landing\_country [M]**

Data available: n = 7467 records missing in year 1989 (n = 3165), 1990 (n = 3960) and 2008 (n = 342)

Procedure: where missing data was set to "SWE" in the general sampling programme (targets vessels that landed only in Sweden) or "UNK" in the industry sampling programme (as it is possible vessels landed in Denmark)

Comments: sampling designs do not secure that the distribution of samples is representative of the landing countries used by the fishery (e.g., industrial landings to Denmark are likely misrepresented in the data available).

#### **Vessel\_flag\_country [M]**

Data available: n = 1795 records missing

Procedure: imputed with "SWE" since the sampling programmes only targets Swedish vessels

Comments: In recent years, central Baltic herring landings from other flag countries to Sweden have been reduced, but it could be worth making an analysis back in time.

#### **Year [M]**

Data available: complete in database

Procedure: extracted from year of sampling

Comments: none.

**Project [M]**

Data available: variable does not explicitly exist in database

Procedure: set to "SWE-DCF"

Comments: "SWE-DCF" only really applies to data 2000 onwards. Before that other programmes were in place

**Trip code [M]**

Data available: inconsistencies present before 2004, likely related to migrations between database systems and/or variations in interpretation of sampling design

Procedure: Added a note "doubtful trip" or "dummy trip"

Comments: number of trip codes may not be a reliable proxy for number of trips sampled

**Station numbers [O]**

Data available: missing in years < 2004. Labeled with "999" with minor mislabeling here and there. "999" is a dummy value used to represent lack of knowledge of *de facto* fishing operation when landings are sampled. Information on individual fishing operations is only available from randomized haul-based sampling carried out in 2020-2021.

Procedure: set to "999" when missing from samples of landings.

Comments: only in the randomized haul-based programme (2020-2021) can station numbers be certain to correspond to fishing operations. In that programme up to 10 hauls were registered in some of the trips sampled.

**Quarter [M]**

Data available: variable does not explicitly exist in database. Arrival database includes dates of departure and arrival. Date of arrival should be interpreted as landings date in the case of general and industrial sampling programmes and arrival date in the case of the randomized haul-based programme (2020-2021).

Procedure: Based landing or arrival data, depending on the programme.

Comments: Arrival date and landing date rarely differ more than 1 day, but trip length can extend over one week.

**Month [M]**

Data available: variable does not explicitly exist in database

Data available: variable does not explicitly exist in database. Arrival database includes dates of departure and arrival. Date of arrival should be interpreted as landings date in the case of general and industrial sampling programmes and arrival date in the case of the randomized haul-based programme (2020-2021).

Procedure: Based landing or arrival date, depending on the programme.

Comments: Arrival date and landing date rarely differ more than 1 day, but trip length can extend over one week.

**Species [M]**

Data available: complete in database.

Procedure: none

Comments: Species identification problems are not known.

**Sex [M]**

Data available: "male", "female" and "not sexed" are used. In the general sampling programme n = 563 individuals were missing sex information somewhat spread over the years. A few individuals (n=6) were originally assigned as "transitional" or other codes. In the industrial programme approximately half of the individuals (n = 15064) were missing sex information. Nearly all the unsexed individuals were collected before 2007 (n= 14487) likely relating to which were collected and pre-analyzed in Denmark by DTU with only otoliths and length and weight information having arrived to Sweden. In industrial landings sex was labelled as transitional in n = 2 individuals. In the randomized haul-based programme n = 1 missing value and n = 3 transitional sexes were registered.

Procedure: all individuals labeled as "not sexed" were assigned code "I". The original records of "transitional" codes were investigated. Only n = 4 individuals were verified to be "transitional", others were set to "I". With respect to missing values, the maturity of individuals was evaluated. According to sampling protocol, all individuals should have been sexed and have a maturity level assigned so the presence of incipient maturity (national scale 1 and 2) likely indicates "I"; On the other hand, individuals classified as mature and missing sex were likely not sexed at all ("U").

All remaining missed maturity and sex. They were also spread over the length gradient. Accordingly, they were set to “U”.

Comments: The large proportion of missing values in samples from industrial landings may generate biases in analyses when these data are pooled together with data from other sampling programmes.

### **Catch category [M]**

Data available: variable does not explicitly exist in database

Procedure: set to “LAN” since all data comes from landings

Comments: “CATCH” is a more accurate depiction of haul-based randomized samples collected in 2020-2021

### **Landing category [M]**

Data available: variable does not explicitly exist in database

Procedure: In the general programme, the variable was set based on respective size categories based on personnel experience. In earlier years when samples were taken per size category, larger fish (size category  $\leq 4$ ) were set to “HUC” and smaller fish (6) were set to “IND”. Intermediate size 5 that is either “IND” or “HUC” was labeled as “HUC” and a note added (“Landing\_category HUC or IND”). Samples from the industrial and randomized programmes were all assigned to “IND”.

Comments: The exact category of landings is hard to determine from size category. Information on this variable is likely not reliable.

### **Commercial size category scale [O]**

Data available: variable does not explicitly exist in database

Procedure: none (i.e., kept as missing).

Comments: Herring sorting in the Baltic Sea is very site specific. In general EU standards are followed at least when applicable (after 1995), but there are countless variants over time, e.g., Västervik sorting. It should also be noted that, concerning herring, the EU standards themselves have changed through time. From 1983 to August 1993 a system with four size categories (1, 2, 3a, 3b) was in place (CR (EEC) 3166/82) with size category 3b amended from September 1993 to 1996 (CR (EEC) 1935/93). From 1997 to 2021 a system with six size categories (1, 2, 3, 4a, 4b and 5)

was set in place (CR 2406/96). That system was later amended to eight categories (1, 2, 3, 4a, 4b, 4c, 5 and 6) in 2002 and kept until present (CR 2495/2001). It is not known how fast and consistently such changes permeated into the national practices and databases. The absence of codes terminating in "a", "b" or "c" from both the official sales data and the sampling data appears to indicate size categorization reported to BWKBALPEL may be incomplete or poor beyond trip level. It is therefore advised the data are not used for any size-category-based analyses.

### **Commercial size category [O]**

Data available: Up to 2002 samples came from sorted landings but code lists used in the sampling database appear to reflect more of a national or local size categorization than an EU standard. Consequently, linking the reported size categories to standard size categories should not be done. From 2022 onwards, the sampling of unsorted catches was put in place and that is visible in the data with size "9".

Procedure: all size "9" were set to "". Remaining were kept they were in the sampling database. Exploratory analysis indicates that size "0" registered in 1987 to 1992 is the largest, with an order of increased sizes taking place between 0 and 6; and to confirm that size 9 samples, even when registered in earlier years, correspond to unsorted samples. Still, considerable inconsistencies are likely to take place across trips (e.g., with landing site) and across years (with changing regulations) so the variable should not be used in cross trip analysis.

Comments: see "Commercial size category scale"

### **Stock [O]**

Data available: variable does not explicitly exist in database

Procedure: set to "her.27.25-2932" based on current subdivisions of the target stock.

Comments: none.

### **Area [M]**

Data available: complete records. Information is for the most based on most fished subdivision in the trip (information collected from logbooks, at landing site, from control authorities, etc.) except for the randomized haul-based samples collected in 2020-2021 where it comes from fishers.

Procedure: simple matching and code conversion.

Comments: all individuals have an associated subdivision (based on most fished subdivision or deduced from other information) but in the provenance of individuals from sometimes multi-area or pooled trips cannot be reliably ascertained. Spatial information can only be assumed to be *de facto* reliable in individuals collected from randomized haul-based sampling in 2020-2021. Furthermore, the non-probabilistic and stratified nature of much of the sampling is prone to induce misrepresentations and imbalances in the data (e.g., between gears). In-depth investigation of potential biases and evaluation of bias-minimization strategies is recommended.

### **Rectangles [M/O]**

Data available: most records complete (exception are n = 812 fish from the industrial programme). Information is for the most based on most fished rectangle in trip (information collected from logbooks, at landing site, from control authorities, etc.) except for the randomized haul-based samples collected in 2020-2021 where it comes from fishers.

Procedure: simple matching.

Comments all individuals have an associated rectangle (based on most fished rectangle or deduced from other information) but in the provenance of individuals from sometimes multi-area or pooled trips cannot be reliably ascertained. Spatial information can only be assumed to be *de facto* reliable in individuals collected from randomized haul-based sampling in 2020-2021.

### **Subpolygon [O]**

Data available: variable does not explicitly exist in database

Procedure: none (I.e., kept as missing).

Comments: none.

### **Length class [M]**

Data available: complete, data in mm. In the general sampling programme individuals were measured in mm up to Q3-2000 and in semi-cm (scm) from Q4-2000 onwards. For the most part, in the industrial sampling programme individuals were measured in scm, with some exceptions (some individuals in 1993, 1994, 1998, 1999, 2006, 2007, 2008 and 2009). All individuals collected under the randomized haul-based sampling programme (2020-2021) were measured in in scm.

Procedure: all lengths in mm were transformed to semi-centimeter. One individual with 19mm without weight or age was removed after checking of original records.

Comments: resolution of original data is standardized to semi-cm. For earlier years, better resolution is available in original data source.

### **Age [O]**

Data available: ages 0 to 19, with n = 1654 missing values. Missing values spread over the time series.

Procedure: simple matching with ages with low age quality ("AQ3") being set to missing.

Comments: depending on the period, ages were determined only from scales (previous 1992), from otoliths with support of scales (1993-2014), or only otoliths (2015-2021). Only from 2013 is there information on readability of age structures (see "Age Quality").

### **Single fish number [O]**

Data available: all individuals have a unique number associated to them (FD2 variable SPMEN\_ID)

Procedure: simple matching

Comments: individuals can be traced back to original data source if at some point that is needed

### **Length code [M]**

Data available: not present in FD2

Procedure: set to "scm"

Comments: Code used does not always correspond to the resolution of data collection but rather resolution of data provision (see comments in "Length class").

### **Aging method [M]**

Data available: not present in FD2

Procedure: set to "SCA" to 1992 and "OWR" from 1993 onwards based on lab procedures.

Comments: From 1992 to 2015, scales were still used as support in the ageing of larger individuals.

### **Age-plus-group [M]**

Data available: not present in FD2

Procedure: set to “-” based on lab protocols.

Comments: Plus-group is not used in Sweden during age determination of central Baltic herring

### **Otolith weight [O] and Otolith side [O]**

Data available: no records exist in database

Procedure: none (i.e., kept as missing).

Comments: none

### **Weight [O]**

Data available: for the most complete with only a few missing values (n=88). Resolution of the weight data collected varied between 1g (before 2017), a mix of 1g and 0.1g (2017) and 0.01g (2018-2021)

Procedure: data floored to lowest integer (g).

Comments: see “outlier analysis”

### **Maturity staging method [O]**

Data available: variable does not explicitly exist in database

Procedure: set as “visual” in agreement with lab protocol

Comments: none

### **National Maturity scale [M]**

Data available: variable does not explicitly exist in database

Procedure: set according to protocol (but see comments): “1-8” (<2010), “1-9w8” (2010-2011), “1-9w81&82” (2012 onwards). These levels reflect the successive modifications of the original eight level Hjort-scale: before 2010, anomalies were not registered; in 2010-2011, they started being registered (maturity “9”); from 2012, onwards stage 8 was split in two: “81” and “82”.



Comments: a more detailed description of the different maturity scales is available from the authors upon request. Due to poor documentation National Maturity Scale between 2012-Q2 to 2013-Q4 was incorrectly set to “1-8” when during this period a scale “1-4” was used<sup>2</sup>.

### **National Maturity stage [M]**

Data available: In the general sampling programme maturities information was for the most complete with a few exceptions. Regarding the industrial sampling programme, few maturity determinations were made before 2007. After 2010, data are nearly complete. Only n = 1 fish lacks maturity stage in the samples of the current randomized haul-based programme.

Procedure: In the general sampling the n = 16244 missing values were kept and n = 3 individuals with erroneous code “0” were set to missing values. The maturity stages of industrial samples were all set to missing values.

Comments: The maturity of individuals collected within the industrial sampling programme were set to missing values because the maturity staging of these samples is considered unreliable due to preservation issues. A detailed description of the different maturity stages is available from the authors upon request.

### **Sampler [M]**

Data available: variable does not explicitly exist in database

Procedure: set as “SelfSampling” in approximation of the dominant box-selection protocol.

Comments: lab personnel always selected individual specimens from boxes.

### **Age quality [M]**

Data available: only available from 2013 onwards. Few missing values.

Procedure: direct correspondence with Code “4” set to “AQ3”. Ages with “AQ3” were set to NA.

Comments: Prior to 2013 some information on readability of otoliths exists in the database but only as “notes” resulting difficult to compile in the time available. As such for purposes of current data compilation it was assumed that ages determined before 2013 were good-enough for purpose. In support of this decision is the fact that the annual percentage of AQ3 readings in

<sup>2</sup> This aspect only became apparent to data submitters a couple of weeks before the benchmark. Personal recollections of this period indicate that the reason for the scale change was the suggestion for the use of a 4-level scale at ICES level. The suggestion appears to have been reversed soon afterwards. Some investigation into original data indicates that some notes on likely conversion of this scale to the original 1-8 scale characteristic of the 1977-2009 period exist on individual specimens. With sufficient time and dedication, it should be possible to recover those notes and attempt a conversion.

years when quality was determined was lower than 2%. Note however that in earlier years, scales were used and the certainty of age determinations in that earlier period may be significantly underestimated).

### SMSF Maturity stage and SMSF Maturity conversion [M]

Data available: variable does not explicitly exist in database

Procedure: set according to the algorithm in Table 1.

Comments: a more detailed description of this correspondence is available from the authors upon request. Note that the conversion of data collected between 2012-Q2 and 2013-Q4 featured incorrectly done in the original data sent to BWKBALPEL due to a point change in the maturity scale used during that period (see more details under 'National Maturity scale').

**Table 1.** Conversion of national (Swedish) maturity stages to SMSF maturity stages (herring).

National Maturity stage	SMSF Maturity stage	SMSF Maturity conversion
1	A	"Yes&full"
2	Ba	"Yes&full"
3	Bb	"Yes&full"
4	Bb	"Yes&full"
5	Cb	"Yes&full"
6	Ca	"Yes&full"
7	Da	"Yes&full"
8 with National_Maturity_scale "1-8" or "1-9w8"	D	"Yes&Partial"
81 with National_Maturity_scale "1-9w81&82"	Da	"Yes&full"
82 with National_Maturity_scale "1-9w81&82"	Db	"Yes&full"
9	F	"Yes&full"

### Length Stratified [M]

Data available: variable does not explicitly exist in database

Procedure: set according to lab protocol (I.e., "yes" in years <2001 and "no" thereafter)

Comments: auxiliary information on the number of individuals per length class is needed to produce unbiased trip-level length frequencies from length-stratified data.

### Sample id [M]

Data available: can be compounded from information available

Procedure: simple concatenation of FD2 fields TRIP\_NUMBER and CATCH\_SPECSIZE\_SIZESORT

Comment: none

### **Gear type [M]**

Data available: complete

Procedure: simple matching.

Comments: the non-probabilistic and stratified nature of the sampling programmes is likely to have led to misrepresentations and imbalances of fish size, age and maturity across gears, subdivisions, and other domains. In-depth investigation of potential biases and evaluation of bias-minimization strategies is recommended.

### **Mesh size [M]**

Data available: missing in most records before 2003-2004, close to full availability thereafter. Information is complete in the samples of the randomized haul-based programme.

Procedure: set to "999" where missing

Comments: the non-probabilistic and stratified nature of the sampling programmes is likely to have led to misrepresentations and imbalances of fish size, age and maturity across gears, subdivisions, and other domains. In-depth investigation of potential biases and evaluation of bias-minimization strategies is recommended.

### **Fishing activity category European level 6 [M]**

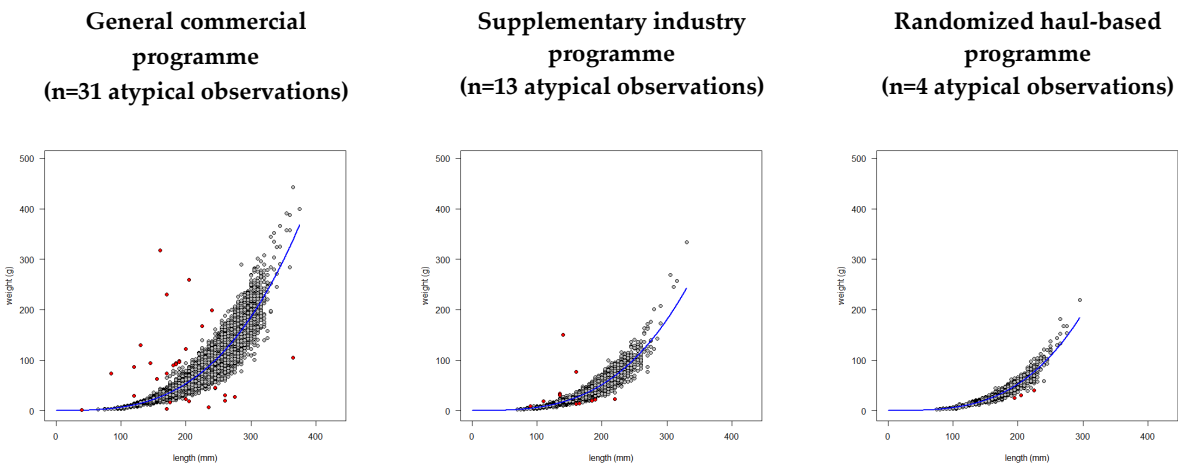
Data available: level 6 gear classification is only available (with no gaps) from 2012 onwards but is fully missing from samples of all earlier years. Information is complete in the samples of the randomized haul-based programme.

Procedure: level 6 gear was imputed based on a correspondence table between concatenations of level 3 and mesh size and level 6 gears.

Comments: level 6 gear prior to 2012 are likely less reliable than more recent ones.

### **Outlier analysis**

Atypical values of weight and length were determined based on automated test (function *outlierTest.lm* of R package *car* with  $\alpha=0.001$ ). This analysis indicated  $n = 31$  atypical fish in the data from the general commercial sampling programme,  $n=13$  fish in the data from the industrial programme and  $n = 4$  from in the data from the randomized haul-based programme (Figure 6). These fish were removed from the final dataset.



**Figure 6. Atypical observations of length-weight per sampling programme.**

## Discussion

The present document resulted from an extensive compilation of historical sampling and laboratory protocols coupled with an investigation of the structure and content of data available in the present national database. The work highlights the importance of requesting full documentation of sampling design, laboratorial procedures, and data storage systems, associated to data collected over extended periods in the context of ICES benchmarks. The aspects documented respect to the Swedish data, but similar evolutions and changes took place, in a multiplicity of different forms, in other countries involved in the herring Baltic fisheries, influencing their data submissions. Undocumented changes in countries' sampling and storage processes over time, variability in the quality of underlying data, and variable data editing practices across national data providers, not only generate variability in data but also, most importantly, affect the representativeness of the final data relatively to the population they aimed to portrait<sup>3</sup>. If not properly accounted for the combination of these aspects across multiple countries will likely lead to misinterpretation of available data (obscuring or over-interpretation of existing patterns), biasing the conclusions drawn and the inputs to assessment.

<sup>3</sup> We assume the population of interest to be the commercial catches of central Baltic herring (in the present case, the catches from Sweden) and the aim of the sampling to be their biological characterization.

The documentation and the exploratory analyses carried out in preparation of the Swedish commercial sampling data provided to BWKBALPEL highlight several aspects that, from a data preparation and submission point of view, may be worth further exploration in future work on this stock. We discuss these national aspects next but underscore to readers that the international nature of the fishing fleets and sampling programmes acting on this stock and the long-time span involved, require more than national corrections. A joint national and international analysis effort will be needed involving sufficient time for iterative communication between national labs, data analysts and stock assessors to obtain final improved datasets and analysis.

### **Quality and evolution of sample design**

Documented sampling designs involving some kind of probabilistic sampling are required to secure that samples effectively represent the population of interest, lowering the level of assumptions involved in estimations associated to biological parameters. Present day randomized haul-based sampling has a clearly identified stratified multi-stage structure and uses probabilistic methods in the sampling of small-pelagic catches in the Baltic. Some parts of the target population are not included in the probabilistic sampling frame<sup>4</sup> but, for the most, the samples should be reasonably representative of the most relevant Swedish catches of this stock with aspects like time, space and landing country/location of the most significant fleet components being reasonably represented (RCG NA NS&EA RCG Baltic, 2020).

The present design presents some significant qualitative improvements over earlier sampling of the stock. It finds improved support on survey sampling theory, improves spatial and temporal resolution of the data (see below), and opens way for bias and uncertainty calculations. Up to 2020, however the sampling of Baltic herring in Sweden was essentially non-probabilistic. Among other, there was little control over the sampling frame and quota sampling was used to obtain the samples. Such methodology works well in producing large-enough samples from each combination of time (quarter), spatial (subdivision) and fishery (industrial/human-consumption) (RCG NA NS&EA RCG Baltic, 2020). Still, it has been long advised against in statistical literature (e.g., Cochran, 1977) and its phase out has been recommended by ICES and STECF expert groups (e.g., ICES 2012, STECF 2017) and instituted in European legislation<sup>5</sup>. Furthermore, sampling was initially done by size category and length-stratified, having changed to unsorted and “random” after. Length stratified samples should be pondered by strata-weights before analysis to avoid biased results (e.g., ICES 2007, 2009). It is not clear how it will be possible to achieve that ponderation with the present BWKBALPEL format. To minimize biases a careful investigation into what was effectively sampled, what effective/potential strata those samples could have

<sup>4</sup> the lesser component of smaller vessels and smaller catches is sampled ad-hoc

<sup>5</sup> Regulation (EU) 2017/1004 of the European Parliament and of the Council of 17 May 2017 on the establishment of a Union framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common fisheries policy and repealing Council Regulation (EC) No 199/2008 (recast)

resulted from (*sensu* what vessels and trips, what gears/métiers, what landings countries/ports were sampled), and what size those strata were/might-have-been would be required.

Proper documentation of the data with all necessary information for investigation of bias and their possible correction is time consuming, going well beyond the time available for data preparation of the current benchmark. Concerning more recent times where fisheries and sampling are better documented such documentation should be possible to obtain. If that turns out to be the case, after some strong assumptions, it should be possible, e.g., to match the samples to “hypothetical” sampling strata that can then be used to weigh the original data and attempt to reduce bias (e.g., that created by the reduced or unrepresentative sampling of large components of the industrial landings that happened abroad). Concerning more historical periods, however, it is less clear if such documentation and bias corrections will be possible because present knowledge of fisheries and their sampling in those periods may be quite scarce. In such situations, it may very well be that there is never enough confidence in present day data-driven interpretations of sampling design and implementation to attempt a re-estimation of biological parameters or to use the older data for purposes other than those they were already used for.

### **Quality and evolution of laboratory analyses**

Laboratory procedures have also changed significantly through time, with possible discontinuities in methodologies used for age determination being worth highlighting. To our knowledge, at least in Sweden, the methodology changed from using scales (up to 1992), to using otoliths with support of scales (from 1993 to 2014), to using only otoliths (from 2015 onwards). Given the importance of age determination in the context of herring stock assessment and the likelihood of potential differences in fish ages that are determined from different calcified structures (Campana 1992) it will be important to check if calibration exercises have been done and account for the possibility of temporal biases in ages reported on this stock<sup>6</sup>. Of additional interest will also be the investigation of some discontinuities in maturity determination methods. This is particularly the case of the use of a 4-point scale during 1992-Q2 to 1993-Q4 and the conversion algorithms used to convert national scales into the final standard scale used in analysis.

### **Changes in the spatial and temporal resolution of data**

In 2020, a new sampling programme was put in place whereby samples started being collected by the fishers at fishing operation level instead of being collected from the landings by a diversity of actors. Those procedures alongside the improved catch records of current times (namely existence of logbooks) allow the direct association of individual fish to the haul they came from and consequently to the precise subdivision, rectangle, and time of fishing. In contrast, most of

<sup>6</sup> Marginal annuli are easier to identify in otoliths than in scales leading to higher annuli counts in the latter, an aspect that could be worth checking and correcting in analysis back in time (Carina Jernberg, pers. obs.).

the historical herring data comes from samples collected during landings or after transportation of the landed catch to factories. That sampling poses a significant challenge to the precision of the spatial and temporal characterization of the individuals sampled. It is not infrequent, especially in the industrial fisheries of more recent times, that fishing trips take several days and fish in a multiplicity of rectangles and subdivisions, making it impossible to certify the rectangle and subdivision (and in some point cases also the month or quarter) that samples came from. The detection and flagging of these situations should be possible (at least in more recent years after logbooks were introduced) allowing for improved results in spatial-temporal analysis.

### **Data availability and quality checking**

Maturity data from specimens collected in the supplementary sampling programme on industrial landings carried out between 1993 and 2020 were not included in the BWKBALPEL data submission because samples were frequently badly preserved, and their maturity staging was not considered reliable enough for end-use. Still, with sufficient time a detailed investigation can be conducted to salvage at least the better conserved of those samples<sup>7</sup>. Additional investigations are also warranted to the possibility that some individual records from this programme appear duplicated in the final dataset because they have been submitted by both Sweden (the vessel flag country) and Denmark (the sampling country). Under EU regulations data final uploads to international databases are responsibility of the flag country but situations of duplicates stemming from ambiguities in bilateral agreements meant to facilitate the sampling of landings abroad have been identified during previous work (ICES 2018, 2019). With additional time and analysis, it should also be possible to recover earlier biological data not yet present in the current database, that may help to clarify Baltic herring growth and maturation patterns in earlier periods of the time series. That is particularly the case of data collected in 1985 and 1986 (which appears to have already been digitized but failed migration to the present database) but also of data collected back to 1977 (or even earlier). According to information collected during the present BWKBALPEL data submission, is likely those earlier periods have at least some age and maturity information that has not yet been digitized<sup>8</sup>. Finally, it is worth noticing the existence in the present database of additional information on biological variables not included in BWKBALPEL request but that may be of interest to future work. That is the case of data on the amount of visceral fat and on the prevalence/abundance of nematodes in the abdominal cavity of Baltic herring individuals.

### **Final considerations**

The full compilation and analysis of information existing on sampling design, data storage and data edits and their potential biases, over a long-time span, across many countries, is time consuming and impossible to achieve within the present time limitations of benchmarks like

<sup>7</sup> Unfortunately, there is no variable in the national database that records specifically the good/bad preservation state of sampled specimens but it may be possible to evaluate other variables, e.g., weight-length relationships of the individuals and, with some assumptions, quality assure some of the existing maturity data.

<sup>8</sup>

BWKBALPEL. For that to happen, more time will need to be available for data preparation and data screening before *de facto* data submission. RCGs, WGCATCH and several other *fora* have, over the years, pointed out that issue, emphasizing that national laboratories need more than one-year ahead notification of upcoming benchmarks to appropriately plan for, and have a minimum chance at, that data compilation and evaluation (RCG NA NS&EA 2017). An extension of this period will be needed for particularly long or complex data submissions and submissions that may require joint work between countries or a more continuous collaboration between national data submitters and those analyzing the data at international level.

The organization of preliminary data scoping meetings ahead of the definition of benchmark data call and/or delegation of the work to expert groups responsible for sampling design and analysis at national level (e.g., WGCATCH, RCGs) are examples of how it may be possible to achieve the required documentation, analysis, and data quality prior to benchmark analysis. The upcoming implementation of the ICES Regional Database and Estimation System (RDBES) will contribute to the process by making available to end-users a comprehensive set of documentation on sampling design and sample quality, in standardized format, alongside the data. Still, it should be noted that, similar to the presented in this document, many quite substantial data (re-)interpretations and decisions happen on the national sphere, ahead of uploads to international systems. To make these visible and possible to scrutinize and evaluate during analysis, documents such as the present one will still be needed.

### **Acknowledgements**

This present work was financed by the Swedish National Work Plan for data collection in the fisheries and aquaculture sectors 2022-2024, in relation to EU regulations (EU) 2017/1004, (EU) 2021/1167 and EU 2021/1168. The Swedish National Work Plan is co-funded by the EMFF (Regulation (EU) No 508/2014) and by Swedish national funds through Swedish Agency for Marine and Water Management.

. Thanks are due to all those involved in data collection over the many years

### **References**

Campana SE 2001 Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of fish biology* 59 (2), 197-242. <https://doi.org/10.1111/j.1095-8649.2001.tb00127.x>

Cochran WG 1977. *Sampling techniques*, 3<sup>rd</sup> ed. New York:Wiley.



ICES 2007. Report of the Workshop on Sexual Maturity Sampling (WKMAT), 15–19 January 2007, Lisbon, Portugal. ICES CM 2007/ACFM:03. 85 pp. <https://doi.org/10.17895/ices.pub.19280243>

ICES 2009. PGCCDBS Guidelines for collecting maturity data and maturity ogive estimation for stock assessment purpose, 2 pp. <https://doi.org/10.17895/ices.pub.19266857>

ICES 2012. Report of the Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS 2012), 30 January–3 February 2012, Rome, Italy. ICES CM 2012/ACOM:50. 163 pp. <https://doi.org/10.17895/ices.pub.19281164>

ICES 2018: Report of the Working Group on Commercial Catches (WGCATCH). ICES Expert Group reports. Report. <https://doi.org/10.17895/ices.pub.8684>

ICES 2019. Report of the Working Group on Commercial Catches (WGCATCH). ICES Expert Group reports. Report. <https://doi.org/10.17895/ices.pub.8158>

RCG NA NS&EA 2017. Report of the Regional Co-ordination Group for the North Sea and Eastern Arctic (RCG NS&EA), 121 pp. <https://datacollection.jrc.ec.europa.eu/docs/rcg>

RCG NA NS&EA RCG Baltic 2020. Regional Coordination Group North Atlantic, North Sea & Eastern Arctic and Regional Coordination Group Baltic. 2020. Part III, Intersessional Subgroup (ISSG) 2019-2020 Reports, 154 pp. <https://datacollection.jrc.ec.europa.eu/docs/rcg>

STECF 2017. Quality Assurance for DCF data (STECF 17-11). Publications Office of the European Union, 78 pp. <https://doi.org/10.2760/680253>

# SmartDots Summary Report for the 2022 exchange for the central Baltic herring stock *her.27.25-2932* (event ID 449)

Working Document for BWKBALPEL (ICES Benchmark Workshop on Baltic Pelagic Stocks) 2023  
Coordination and analysis: Julie Coad Davies, National Institute of Aquatic Resources, DTU Aqua, Denmark

## 1 Summary

This summary gives the results presented to the BWKBALPEL (ICES Benchmark Workshop on Baltic Pelagic Stocks) 2023 data compilation meeting in November 2022. Age error matrices (AEM's) were provided following a request from the group. A single matrix per ICES SubDivision (SD) was provided as well as a combined AEM for all SD's.

The full report can be found <https://smartdots.ices.dk/ViewEvent?key=449>

The 2022 exchange for the central Baltic herring stock *her.27.25-2932* took place via the SmartDots platform between May and October 2022. The exchange was organised following a request from WGBFAS and in preparation for the 2023 benchmark of the stock. Fifteen readers from nine countries took part (Denmark, Poland, Sweden, Germany, Latvia, Lithuania, Estonia and Finland); twelve "advanced" readers (providing age data for assessment) and 3 "basic" readers (do not provide age data for assessment). 163 otoliths images, covering ICES SD25, 26, 29 and 32 were provided by Poland and Finland and uploaded to the SmartDots platform. The aim was to include samples from all SD's included in the stock assessment but the otoliths from SD27 were not included due to lack of resources within the lab photographing the otoliths. Images of whole otoliths from SD's 25 (n=27) and 26 (n=30) were provided by Poland. For SD 29, images of sectioned and stained and whole otoliths from the same fish (n = 24) plus additional images (n=18) of sectioned and stained otoliths were provided by Finland. For SD32, images of sectioned and stained otoliths (n=40) were provided by Finland. The aim was to cover all areas, quarters and age groups for each ICES SD's used in the stock assessment but this aim was not reached.

This summary report presents the results based on advanced readers only (those who provide age date for stock assessment purposes); for SD 25 overall PA was 93%, CV was 8% and relative bias -0.04; for SD 26, overall PA was 85%, CV was 9% and relative bias -0.01; for ICES SD 29 overall PA was 89%, CV was 12% and relative bias 0.06; for ICES SD 32, (based on the ATAQCS analysis) overall PA was 70%, CV was 7% and relative bias 0.38. The analysis was carried out by ICES SD as not all readers are experienced in reading otoliths from all areas and the growth patterns observed in the otoliths vary greatly from north to south, meaning a correct interpretation by readers not experienced with samples from another SD would introduce bias in the results.

## 2 Overview of samples and advanced readers

**Table 2.1:** Overview of samples for the 2022 exchange for the central Baltic herring stock. The modal age range is 0-10.

Year	ICES area	Quarte r	Number of samples	Modal age range	Length range
2021	27.3.d.25	3	8	1-7	10-20 mm
2021	27.3.d.25	4	5	1-10	15-20 mm
2022	27.3.d.25	1	6	1-7	15-20 mm
2022	27.3.d.25	2	8	1-8	10-20 mm

2021	27.3.d.26	3	6	1-8	15-20 mm
2021	27.3.d.26	4	8	1-8	15-20 mm
2022	27.3.d.26	1	8	1-8	10-25 mm
2022	27.3.d.26	2	8	1-8	10-20 mm
2022	27.3.d.29	1	24	0-6	95-180 mm
2021	27.3.d.32	1	20	1-8	90-175 mm
2021	27.3.d.32	2	20	1-8	95-175 mm

**Table 2.2:** Overview of advanced readers.

Reader code	Expertise
R01 DK	Advanced
R02 SE	Advanced
R03 EE	Advanced
R04 LT	Advanced
R05 FI	Advanced
R06 FI	Advanced
R07 PL	Advanced
R09 LV	Advanced
R10 EE	Advanced
R11 DE	Advanced
R12 DE	Advanced

### 3 Results

This section shows overall results from the SmartDots output for ICES SD 25, 26 and 29. A full description of the methods applied are available in the full report. Only two advanced readers read the samples from ICES SD 32, thus the SmartDots output is not available, results from SD 32 are based on a separate analysis.

#### 3.1 Age error matrix (AEM) for SD25

**Table 3.1:** Age error matrix (AEM) for ICES SD 25. The AEM shows the proportional distribution of age readings for each modal age. Only advanced readers are used for calculating the AEM.

Read age	1	2	3	4	5	6	7	8	10	Total
<b>Modal age</b>										
Age 1	<b>0,89</b>	0,11	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00
Age 2	0,07	<b>0,93</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00
Age 3	0,00	0,00	<b>1,00</b>	0,00	0,00	0,00	0,00	0,00	0,00	1,00
Age 4	0,00	0,00	0,00	<b>1,00</b>	0,00	0,00	0,00	0,00	0,00	1,00
Age 5	0,00	0,00	0,00	0,11	<b>0,89</b>	0,00	0,00	0,00	0,00	1,00
Age 6	0,00	0,00	0,00	0,00	0,00	<b>0,89</b>	0,11	0,00	0,00	1,00
Age 7	0,00	0,00	0,00	0,00	0,00	0,00	<b>1,00</b>	0,00	0,00	1,00
Age 8	0,00	0,00	0,00	0,00	0,00	0,00	0,17	<b>0,83</b>	0,00	1,00
Age 10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,33	<b>0,67</b>	1,00
Total	0,96	1,04	1,00	1,11	0,89	0,89	1,28	1,17	0,67	

#### 3.2 Age error matrix (AEM) for SD26

**Table 3.2:** Age error matrix (AEM) for ICES SD 26. The AEM shows the proportional distribution of age readings for each modal age. Only advanced readers are used for calculating the AEM.

Read age	1	2	3	4	5	6	7	8	9	Total
----------	---	---	---	---	---	---	---	---	---	-------

Modal age										
Age 1	<b>0,95</b>	0,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00
Age 2	0,00	<b>0,90</b>	0,10	0,00	0,00	0,00	0,00	0,00	0,00	1,00
Age 3	0,00	0,04	<b>0,96</b>	0,00	0,00	0,00	0,00	0,00	0,00	1,00
Age 4	0,00	0,00	0,00	<b>1,00</b>	0,00	0,00	0,00	0,00	0,00	1,00
Age 5	0,00	0,00	0,00	0,00	<b>0,93</b>	0,07	0,00	0,00	0,00	1,00
Age 6	0,00	0,00	0,00	0,00	0,16	<b>0,68</b>	0,16	0,00	0,00	1,00
Age 7	0,00	0,00	0,00	0,00	0,00	0,20	<b>0,65</b>	0,15	0,00	1,00
Age 8	0,00	0,00	0,00	0,00	0,00	0,05	0,05	<b>0,85</b>	0,05	1,00
Age 9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	<b>0,00</b>	0,00
Total	0,95	0,99	1,06	1,00	1,09	1,00	0,86	1,00	0,05	

### 3.3 Age error matrix (AEM) for SD29

**Table 3.3:** Age error matrix (AEM) for ICES SD 29. The AEM shows the proportional distribution of age readings for each modal age. Only advanced readers are used for calculating the AEM.

Read age	0	1	2	3	4	5	6	No age	Total
Modal age									
Age 0	<b>0,95</b>	0,05	0,00	0,00	0,00	0,00	0,00	0,00	1,00
Age 1	0,05	<b>0,90</b>	0,00	0,00	0,00	0,00	0,00	0,05	1,00
Age 2	0,00	0,00	<b>0,85</b>	0,15	0,00	0,00	0,00	0,00	1,00
Age 3	0,00	0,00	0,00	<b>0,81</b>	0,19	0,00	0,00	0,00	1,00
Age 4	0,00	0,00	0,00	0,13	<b>0,63</b>	0,19	0,06	0,00	1,00
Age 5	0,00	0,00	0,00	0,00	0,00	<b>0,00</b>	0,00	0,00	0,00
Age 6	0,00	0,00	0,00	0,00	0,00	0,00	<b>0,00</b>	0,00	0,00
Age 7	0,95	0,05	0,00	0,00	0,00	0,00	0,00	<b>0,00</b>	1,00
Total	1,00	0,95	0,85	1,09	0,81	0,19	0,06	0,05	

### 3.4 Age error matrix (AEM) for SD25, 26 and 29 combined

**Table 3.4:** Age error matrix (AEM) for ICES SD 25, 26 and 29. The AEM shows the proportional distribution of age readings for each modal age. Only advanced readers are used for calculating the AEM.

Read age	1	2	3	4	5	6	7	8	9	10	No age	Total
Modal age												
Age 1	<b>0,94</b>	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1
Age 2	0,04	<b>0,91</b>	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	1
Age 3	0,00	0,02	<b>0,93</b>	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1
Age 4	0,00	0,00	0,00	<b>0,93</b>	0,08	0,00	0,00	0,00	0,00	0,00	0,00	1
Age 5	0,00	0,00	0,00	0,08	<b>0,80</b>	0,10	0,03	0,00	0,00	0,00	0,00	1
Age 6	0,00	0,00	0,00	0,00	0,12	<b>0,74</b>	0,15	0,00	0,00	0,00	0,00	1
Age 7	0,00	0,00	0,00	0,00	0,00	0,13	<b>0,78</b>	0,09	0,00	0,00	0,00	1
Age 8	0,00	0,00	0,00	0,00	0,00	0,04	0,08	<b>0,85</b>	0,04	0,00	0,00	1
Age 9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	<b>0,00</b>	0,00	0,00	0
Age 10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,33	0,00	<b>0,67</b>	0,00	1
Total	0,98	0,99	0,95	1,06	0,99	1,00	1,03	1,27	0,04	0,67	0,02	

### 3.5 Coefficient of Variation by ICES SD

**Table 3.5:** Coefficient of Variation (CV) per ICES SD shows the CV of all advanced readers combined per modal age and a weighted mean of the CV per SD.

Modal age	27.3.d.25	27.3.d.26	27.3.d.29
0	-	-	-
1	30 %	21 %	25 %
2	13 %	15 %	0 %
3	0 %	7 %	10 %
4	0 %	0 %	14 %
5	7 %	5 %	8 %
6	5 %	10 %	8 %
7	0 %	9 %	-
8	5 %	7 %	-
9	-	-	-
10	12 %	-	-
11	-	-	-
<b>Weighted Mean</b>	<b>8 %</b>	<b>9 %</b>	<b>12%</b>

### 3.6 Percentage Agreement by ICES SD

**Table 3.6:** Percentage Agreement (PA) per ICES SD shows the PA of all advanced readers combined per modal age and a weighted mean of the PA per SD.

Modal age	27.3.d.25	27.3.d.26	27.3.d.29
0	-	-	100 %
1	89 %	95 %	93 %
2	93 %	90 %	100 %
3	100 %	96 %	91 %
4	100 %	100 %	75 %
5	89 %	93 %	87 %
6	89 %	68 %	75 %
7	100 %	65 %	-
8	83 %	85 %	-
9	-	-	-
10	67 %	-	-
11	-	-	-
<b>Weighted Mean</b>	<b>93 %</b>	<b>85 %</b>	<b>89 %</b>

### 3.7 Relative Bias by ICES SD

**Table 3.7:** Relative Bias by per ICES SD shows the relative bias of all advanced readers combined per modal age and a weighted mean of the relative bias per SD.

Modal age	27.3.d.25	27.3.d.26	27.3.d.29
0	-	-	-
1	0.11	0.05	0.07
2	-0.07	0.10	0.00
3	0.00	-0.04	0.06
4	0.00	0.00	0.14
5	-0.11	0.07	0.00
6	0.11	0.00	-
7	0.00	-0.05	-

8	-0.17	-0.10	-
9	-	-	-
10	-0.67	-	-
11	-	-	-
<b>Weighted Mean</b>	<b>-0.04</b>	<b>-0.01</b>	<b>0.06</b>

### 3.8 Results from SD32

**Table 3.8.1:** Overview of results from the ATAQCS workbook showing per modal age; number of age readings per age reader, number of agreed ages, PA (percentage agreement), coefficient of variation (CV) and bias.

Age	R06	R05	No. Agreed	PA %	CV	Bias
0	-	-	-	-	-	-
1	5	5	5	100,0	0,000	0,00
2	12	8	8	66,7	0,106	0,33
3	5	8	4	80,0	0,070	0,20
4	8	5	4	50,0	0,093	0,75
5	3	5	3	100,0	0,000	0,00
6	3	3	1	33,3	0,043	0,67
7	2	3	1	50,0	0,088	1,00
8	2	2	2	100,0	0,000	0,00
9	0	1	0			
<b>Totals</b>	<b>0</b>	<b>40</b>	<b>40</b>	<b>70,00</b>	<b>0.067</b>	<b>0.38</b>

**Table 3.8.2:** Reader comparison matrix. Green area is agreement, red area is overestimation by R05 FI.

R06 FI Age	R05 FI Age									
	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	5	0	0	0	0	0	0	0	0
2	0	0	8	4	0	0	0	0	0	0
3	0	0	0	4	1	0	0	0	0	0
4	0	0	0	0	4	2	2	0	0	0
5	0	0	0	0	0	3	0	0	0	0
6	0	0	0	0	0	0	1	2	0	0
7	0	0	0	0	0	0	0	1	0	1
8	0	0	0	0	0	0	0	0	2	0
9	0	0	0	0	0	0	0	0	0	1

**WD5:** Comparing WEST (weight-at-age in the catch) and WECA31, (weight-at-age in the stock) for sprat (*Sprattus sprattus*) and central Baltic herring (*Clupea harengus*) in the Baltic Sea

Authors: Steffanie Haase (Thünen Institute of Baltic Sea Fisheries, Rostock, Germany), Szymon Smoliński (National Marine Fisheries Research Institute, Gdynia, Poland)

BWKBALTPEL 2023

## Background

The average weight-at-age in the catch (WECA) and the average weight-at-age in the stock (WEST) are used to estimate the SSB and CATON (catch in tonnes). WECA is annually estimated from the commercial catch data covering all seasons. At the moment, WEST=WECA is assumed in the stock assessment models as survey data are only available during the annual Baltic International Acoustic Survey (BIAS) in quarter 4.

In order to test if the commercial fishery is selecting for individuals with higher mean weight at age compared to the weight-at-age in the stock and thus overestimates WEST by using WECA, we compare weight data in quarter 4 from commercial samples with weight data from the BIAS survey. Survey catches cover a larger part of the distribution area of the stocks and are taken using a smaller mesh size in the codend.

## Method to compare WECA and WEST

WECA (weight-at-age) files per ICES SD were extracted from Intercatch (IC) and selected for quarter 4 to ensure comparability between WECA and WEST (<https://intercatch.ices.dk/login.aspx>).

WEST was calculated based on the survey catches of the BIAS which were uploaded at the Acoustic Trawl Database (<https://www.ices.dk/data/data-portals/Pages/acoustic.aspx>). Individual fish, which are sampled length-stratified in the BIAS, were used to calculate an age-length key (ALK) per SD and year. Further, a length-weight relationship (LWR) was estimated per SD and year. The number of fish per length class in the catches was standardized to the 30-minute hauls and total length distributions were calculated using standardized values. Length were transformed to weights by the LWR and an ALK was applied to get weight-at-ages per SD, year and species.

WECA and WEST were compared for ICES SDs 25-32 and the years 2015-2021 as these years were available at the Acoustic Trawl Database.

## Comparison WECA and WEST for herring

Weight-at-age in the stock (WEST) and in the catch (WECA) generally shows a high correlation for herring (Fig. 1 and 2). Differences between WEST and WECA are larger for younger age groups.

Fig. 3 shows the difference between WECA and WEST per age group and SDs over the years. Differences are particularly high in SDs 25 and 26 and for age group 1. In most years, WECA is larger than WEST in SD 25 and 26. In contrast, WEST is larger in most years for age group 1 and 2 in SDs 27-32.

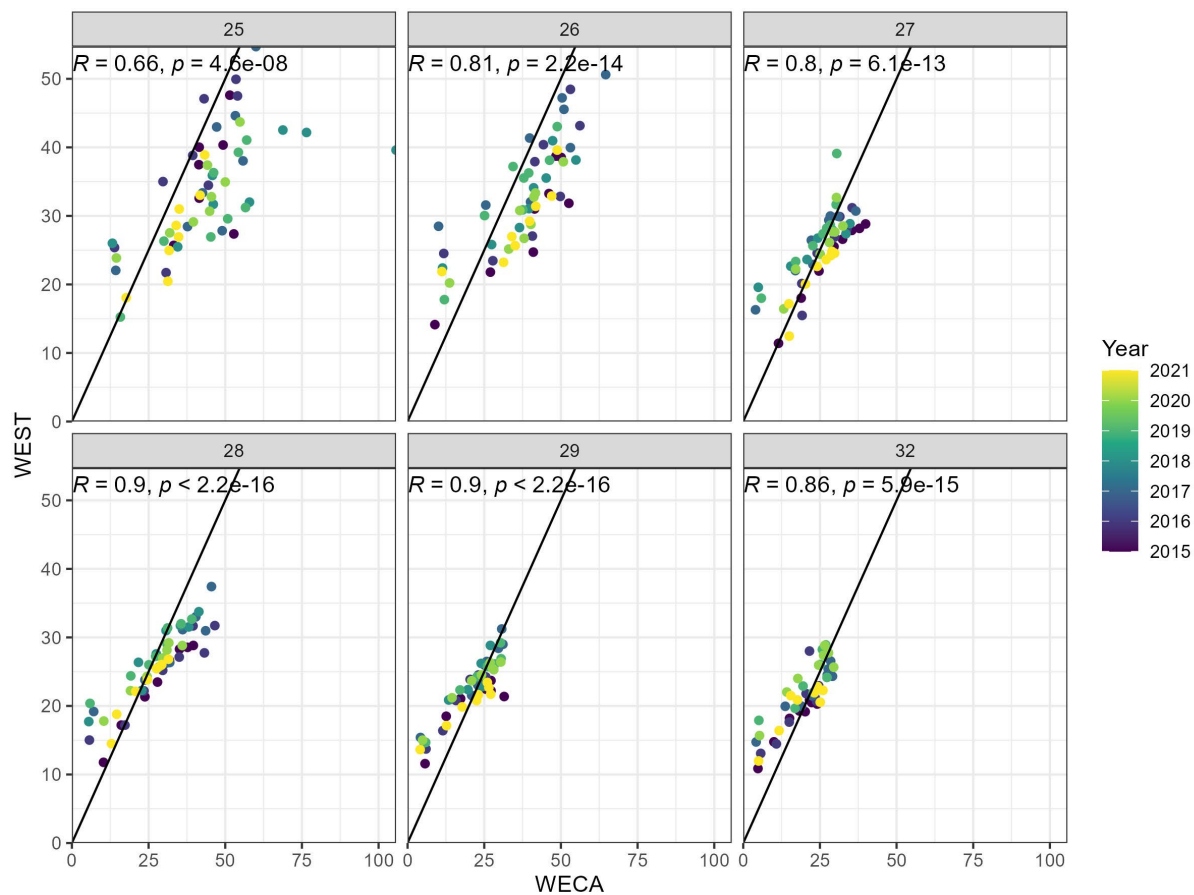


Figure 1: Comparison of weight-at-age in stock (WEST) and catch (WECA) of herring per ICES SD. The black line refers to the line through the origin.



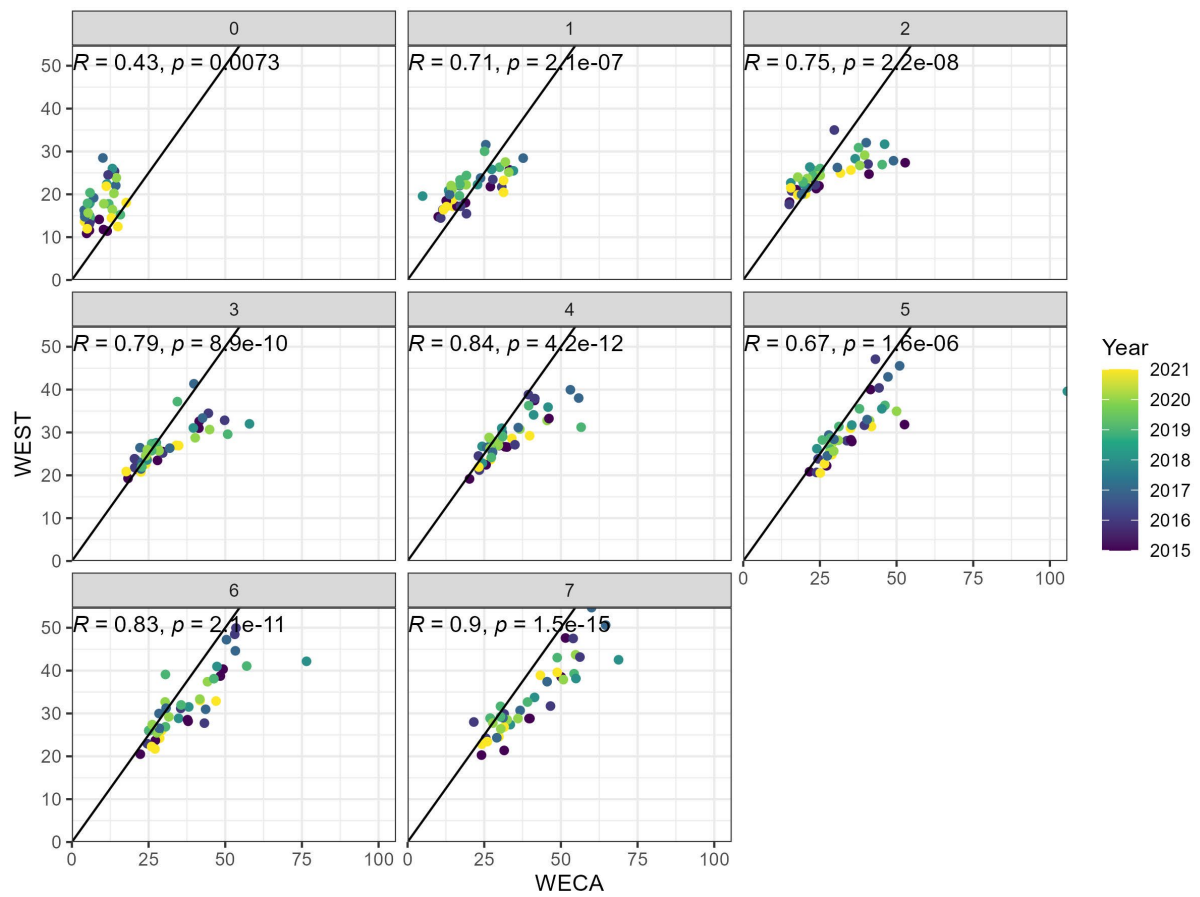


Figure 2: Comparison of weight-at-age in stock (WEST) and catch (WECA) of herring per age group. The black line refers to the line through the origin.

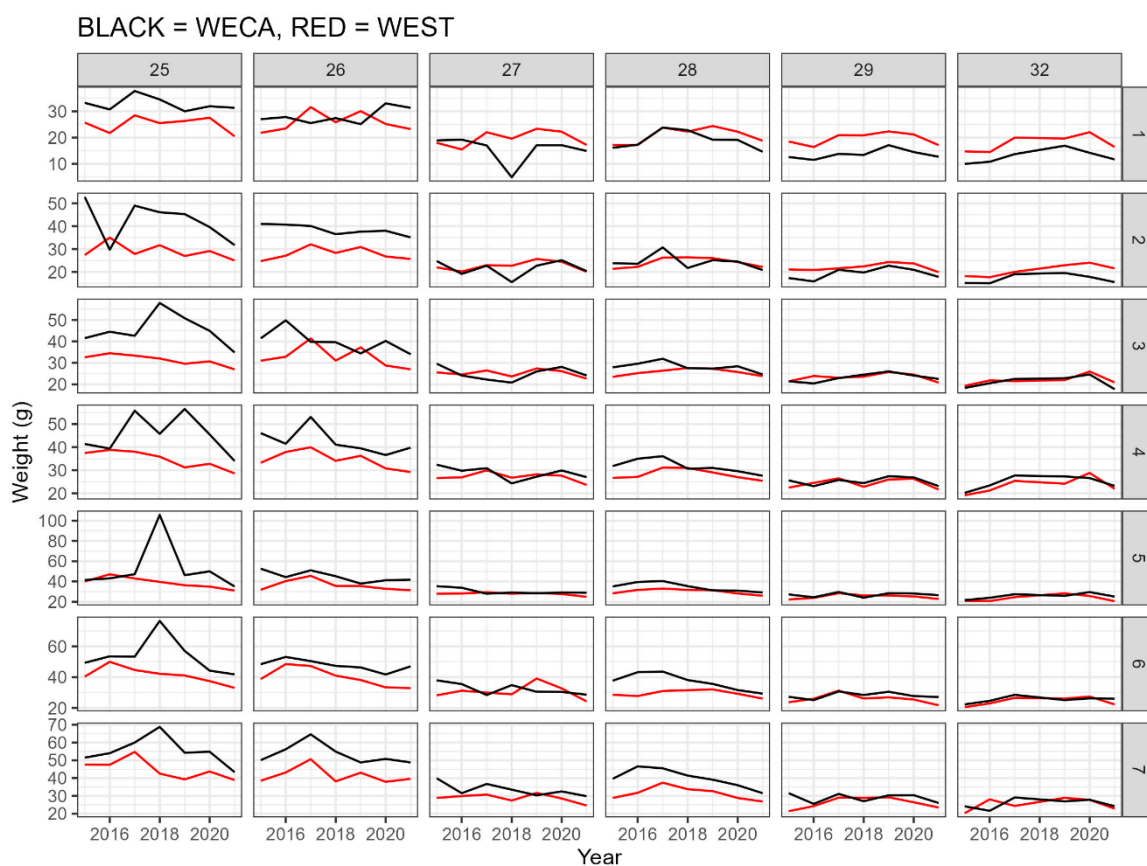


Figure 3: Comparison of weight-at-age in catch (WECA; black) and weight-at-age in the survey (WEST; red) per ICES SD (columns), age (rows), and year for herring.

### Comparison WECA and WEST for sprat

Also for sprat, weight-at-age in the stock (WEST) and in the catch (WECA) shows a high correlation (Fig. 4 and 5), but the correlation is compared to herring slightly lower. Again, differences in WEST and WECA are larger for younger age groups.

Fig. 6 shows the difference between WECA and WEST per age group and SDs over the years for sprat. Differences are particularly high for age group 1 and 2. In most years, WEST is larger than WECA in SD 27-32 independent of the age class. WECA is larger than WEST in SD 25 and 26 in age class two and older. The weight-at-age shows a distinct variation between years in some age groups and SDs in both WECA and WEST (e.g. SD 30).

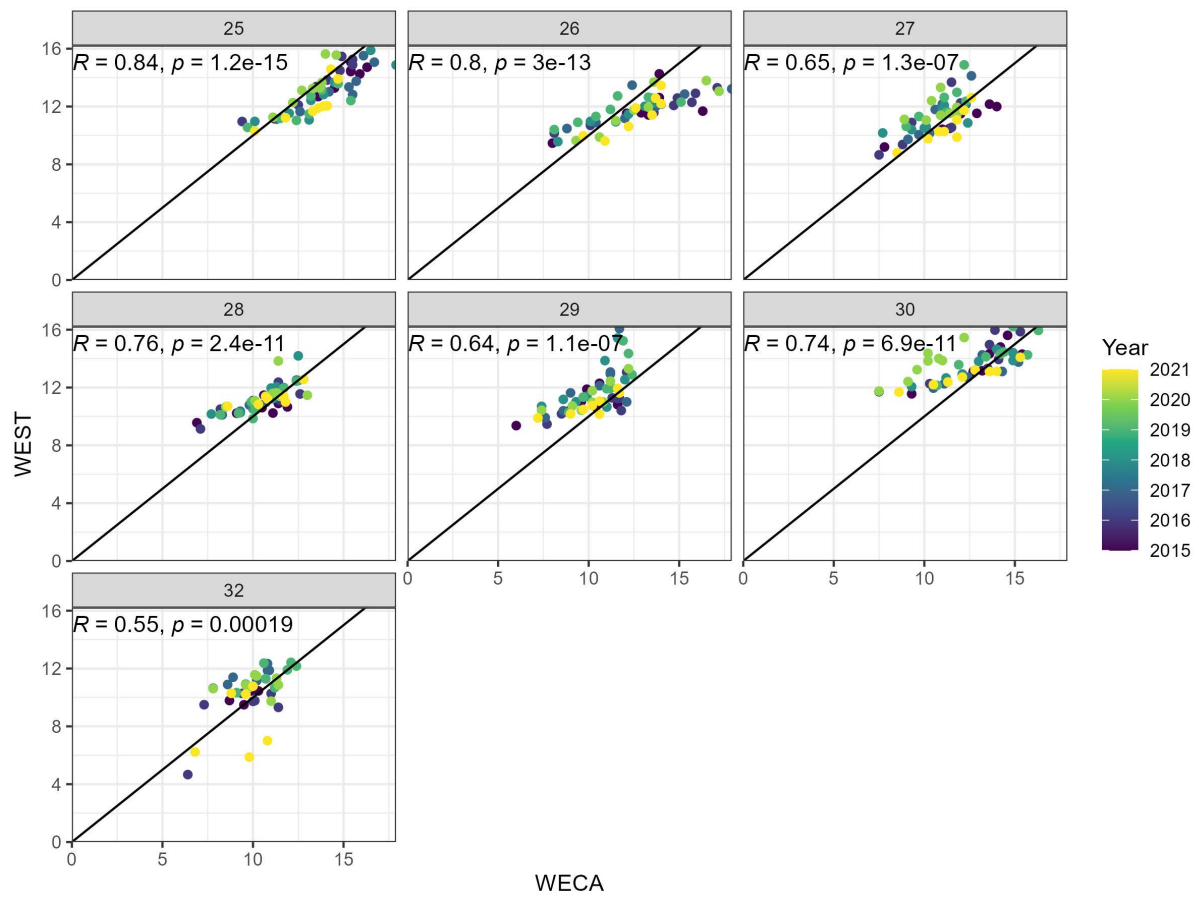


Figure 4: Comparison of weight-at-age in stock (WEST) and catch (WECA) of sprat per ICES SD. The black line refers to the line through the origin.

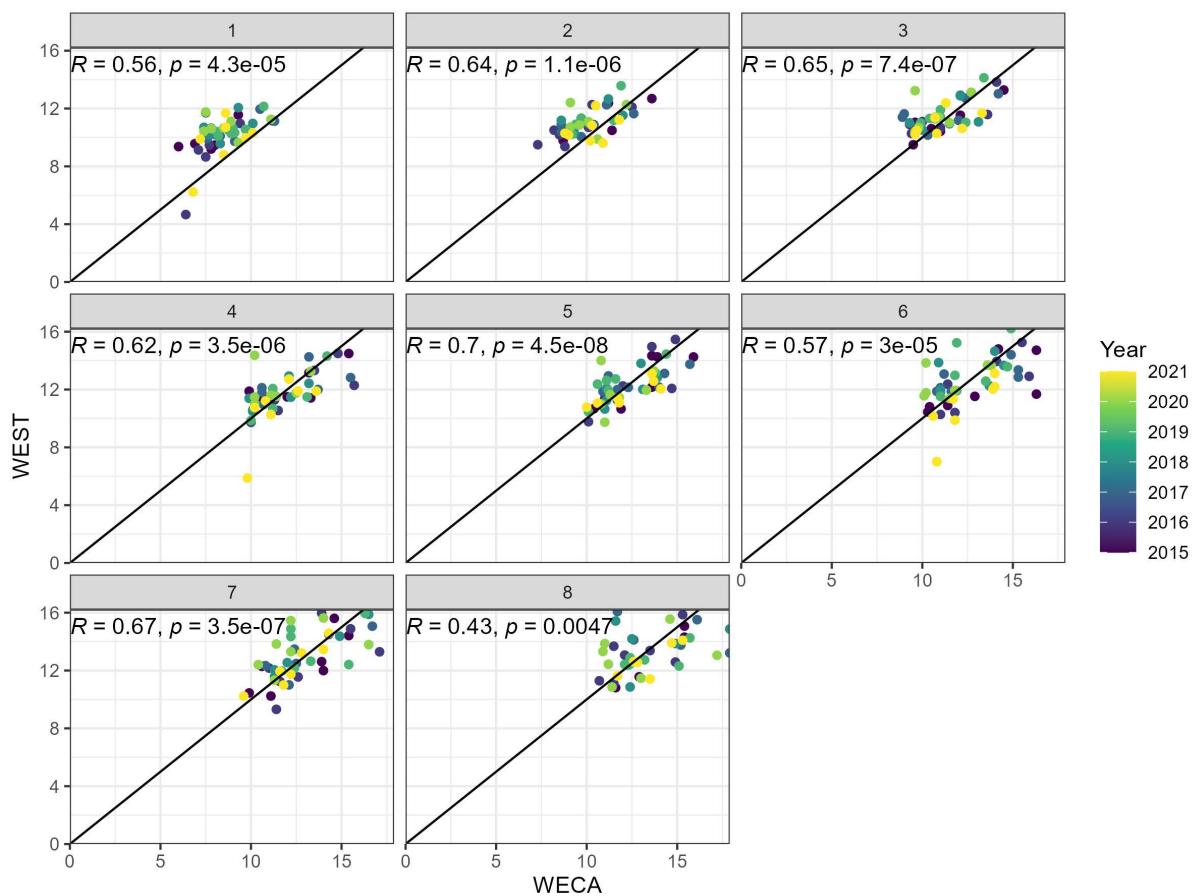


Figure 5: Comparison of weight-at-age in stock (WEST) and catch (WECA) of sprat per age group. The black line refers to the line through the origin.

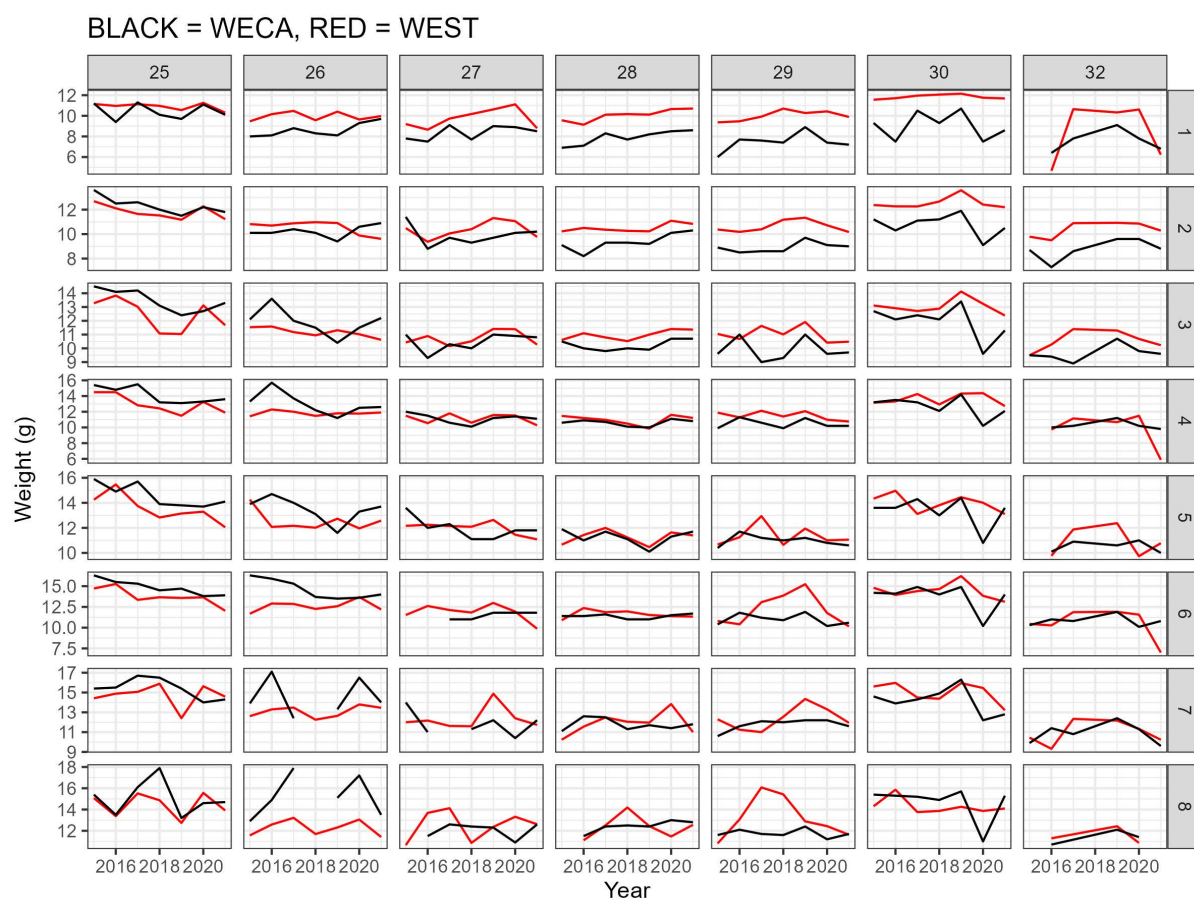


Figure 6: Comparison of weight-at-age in catch (WECA; black) and weight-at-age in the survey (WEST; red) per ICES SD (columns), age (rows), and year for sprat.

## Conclusion

WECA and WEST in quarter 4 generally show a high correlation, differences, however, occur especially in ages 1 and 2 and for SDs in the southern Baltic Sea for herring and northern Baltic Sea for sprat. Weights in the catch are compared to weights in the survey only distinctively larger for herring in all age groups in SD 25 and 26.

Differences in weight between the survey and commercial catches might occur due to the different selectivity of the gears and unequal geographical coverage of fishing effort and surveys. One further explanation for the difference in weight could be that the BIAS samples are weighted and measured fresh while commercial samples are often frozen before they are measured and weighted. Clupeids shrink when they are frozen, and this effect can be higher for smaller fish than for larger specimens (e.g. Santos *et al.*, 2009).

A quarterly coverage of survey weights is currently unavailable. Therefore, we only compared the weight-at-age from quarter 4. As SSB is calculated in the beginning of the year, weights from the quarter 4 BIAS could be used as WEST estimates. There is, however, a large difference in weights-at-age between the different SDs in both WEST and WECA. As there is currently no spatially resolved stock assessment in place, one WECA value is used for all SDs. Further investigation is needed on how to raise weight-at-ages from the survey based

on the spatial distribution of the stocks to implement weight-at-age from the survey as WEST.

## References

Santos, J. N. S., Araújo, F. G., and Silva, D. S. 2009. Length correction for early-juvenile Brazilian herring *Sardinella janeiro* (Eigenmann, 1894) after preservation in formalin, ethanol and freezing. *Neotropical Ichthyology*, 7: 87–92. Sociedade Brasileira de Ictiologia.

## WD6: Maturity-at-age analysis for herring (*Clupea harengus*) in subdivisions 25-29 and 32, excluding the Gulf of Riga (central Baltic Sea)

Authors: Francesco Masnadi, Francesca Vitale, Massimiliano Cardinale, Szymon Smoliński, Mikaela Bergenius Nord

BWKBALYPEL 2023

### 1. Purpose

In this WD maturity-at-age data from 1984 to 2021 were used to derive the proportion of mature at age (using GLMM) to provide an updated maturity ogive for Central Baltic herring (*Clupea harengus*) stock (Subdivisions 25 to 27, 29, and 32 excluding the Gulf of Riga). Moreover, according to evidence of a spatial-temporal trend in the life histories of the stock (ICES, 2022), maturity analyses were conducted taking into account Subdivision and year effect. A50 (age-at-50%-maturity) was used as a proxy to detect possible spatial-temporal trends in maturation.

### 2. Data exploration – Official corrections

The data used for the following analyses were requested by the ICES benchmark workshop on Baltic Pelagic stocks (BWKBALYPEL) in preparation for the assessments model that will be used as a basis for advice for the Central Baltic herring stock. The data contains biological information for herring in the Central Baltic Sea, ICES Division 3d (Subdivisions 25 to 29, and 32). For the data call, countries were asked to convert their national maturity scales into the internationally agreed maturity key (SMSF, Sexual Maturity Scale for all Fish) as approved by WKASMSF 2018 (ICES, 2018) as mandatory for the ICES database. This was the scale considered for all the analyses in the present WD. Hereafter a brief summary of data exploration and applied corrections to the data obtained from the official data call.

Before using the SMSF maturity scale's stages some data manipulation was deemed necessary. Some countries did not use the SMSF but they convert their national maturity scale into the previously used ICES scale M6 (<https://vocab.ices.dk/?ref=1625>). This required an additional manual conversion before data analysis.

Stages from 61 to 66 included in M6 were thus converted into SMSF, according to the table in Annex 6 of WKASMSF 2018 (ICES, 2018), as below:

M	SMSF
61	A
62	Bb
63	C
64	D
66	F

\*note that individual fish in Stage 65 were not found in the data, thus never recorded.

Moreover, a closer exploration of the dataset revealed some possible errors in the raw data. Therefore, the experts dedicated some extra time to identify and solve major problems. Listed below are the corrections made to the data, sometime after feedback received from the country specific experts of the data in question.

- Finland:

Individuals classified as stage 3 in the national 8-point scale and translated as Ba (immature) in the SMSF were corrected to Bb (mature);

Individuals classified as SI (immature) and SM (mature) in the national dataset (from 1984 to 1997) were converted to SI=A and SM=B of SMSF scale to be used for the purpose of the analyses;

- Lithuania:

From 2009 to 2018 LTU did not translate national scale to SMSF, so individuals were translated into the SMSF according to the table below:

LTU National scale	SMSF
1	A
2	Bb
3	C
4	D
5	D
6	F

- Latvia:

National's scale stage 2 (scale 1-6) and stage 5 (in scales 1-5 and 1-5AndAbnormal) according to Latvian maturity scaling refer to gonads without any signs of development (except juveniles). This stage may refer to resting after the spawning season or resting just before the spawning season, thus omitting spawning, i.e. a fish that has reproduced the year before but is going to jump over the spawning season during the current year. Both stages had initially been translated to stage E of the SFMF scale but probably stage E is not appropriate for both cases. As a solution, as suggested by the Latvian expert, stage E from the 1st quarter was left as stage E while for the 2<sup>nd</sup> quarter was translated into stage D.

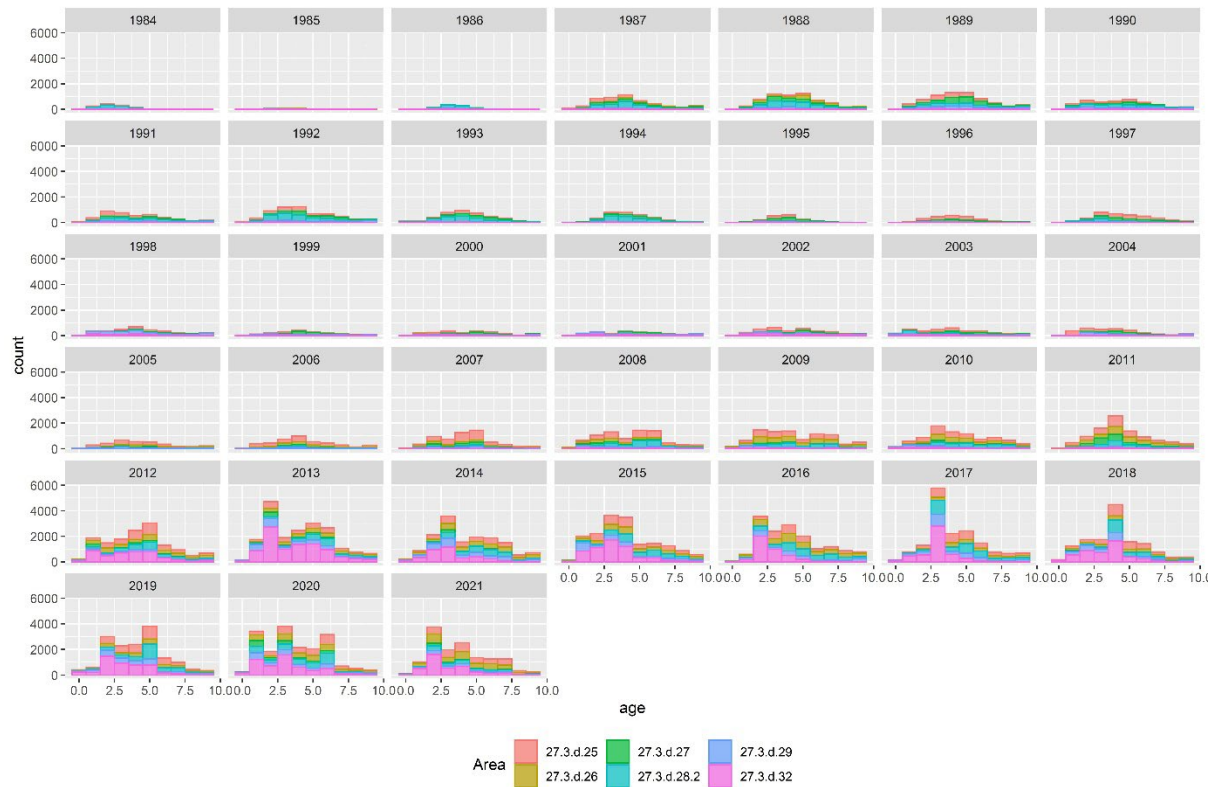
- Sweden:

Looking at individuals in stage Ba (immature, first spawner) in Swedish data multiple individuals at older ages were noticed. This phenomenon was particularly evident for some year, i.e. 1992 & 1993. After consultation with Swedish experts, it was ascertained that the national scale was erroneously reported as an 8 points instead of a 4 point scale that had been used during those two years (WD-XX by Nuno). Consequently, the conversion from the national to international scale was incorrect. After correcting the data, those individuals previously recorded as Ba stage ended up being Bb for quarters 1,2 and 3 in 1992 and for all quarters in 1993.

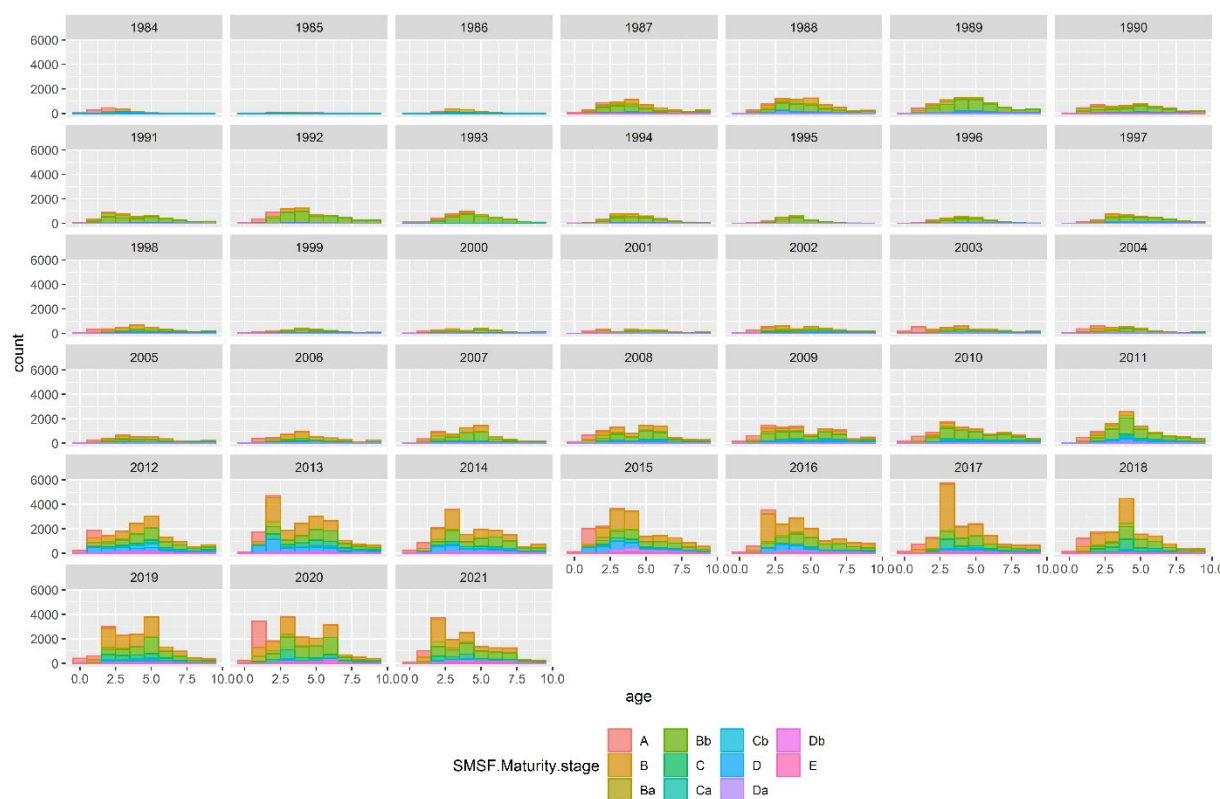
Finally, all abnormal individuals (stage F) were removed from the analyses. For various reasons these fish have hinders with their reproduction and do not participate in the current spawning season. Biologically, these fish are mature as they shown a certain gonad development, but as they do not participate in the current spawning season, they should not be included in the maturity ogive of the current year.



Figure 1 and 2 shows the availability of maturity-at-age data from 1984 to 2021 divided by Subdivision and maturity stage (SMSF scale) after the above-reported corrections in the dataset. Samples age  $\geq 9$  were considered a plus group (9+).



**Figure 1.** Sample size of maturity-at-age data (9+ represents a plus group) from 1984 to 2021 by Subdivision after official corrections to the dataset.



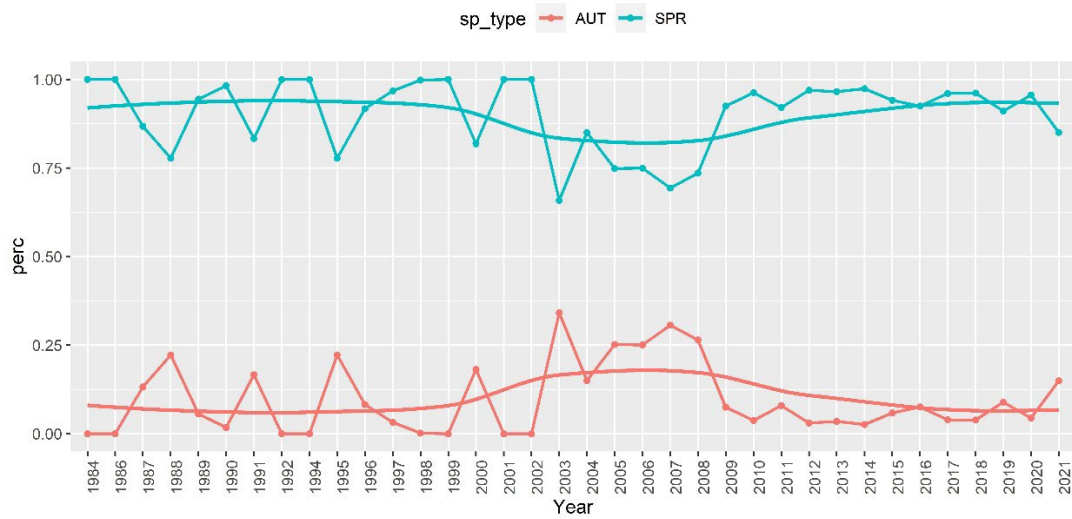
**Figure 2.** Sample size of maturity-at-age data (9+ represents a plus group) from 1984 to 2021 by maturity stage (SMSF scale) after official corrections to the dataset.

### 3. Proportion of autumn spawners

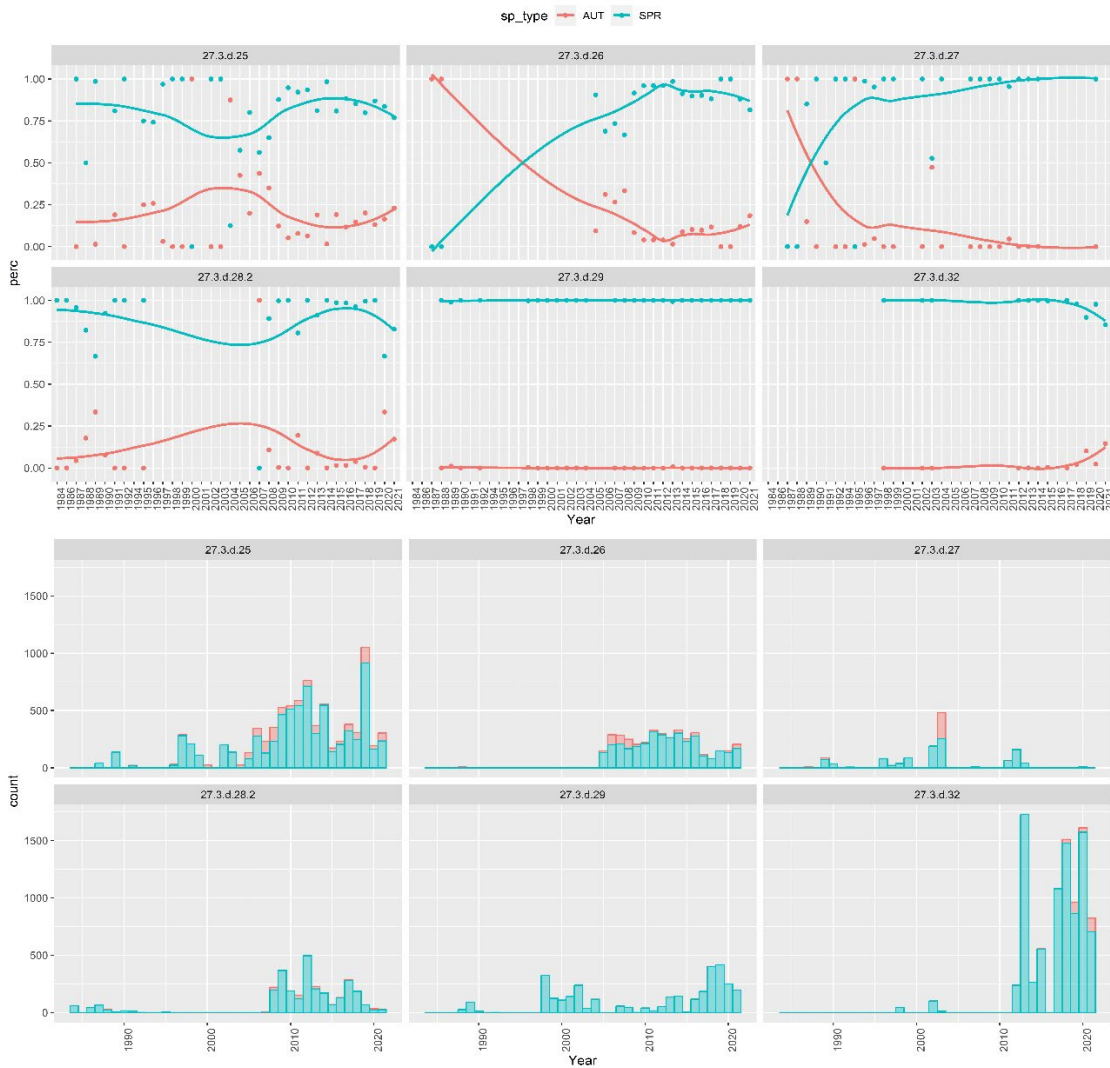
To investigate if the proportion of each of the two spawning types (autumn or spring) has changed over time in the study area, stage C individuals (Ca=actively spawning, Cb=spawning capable) were selected to check how they distribute overall through the time series (aggregated data; Figure 3) and by Subdivisions (Figure 4). Individuals in Stage C are close to or currently spawning.

The first plot shows that in general autumn spawners make up a small percentage of the total compared to the spring spawners (11% vs 89% on average) and no particular trend were detected over time series, apart from a small increase of autumn spawners from 2000 to 2010 ( Figure 3). Looking at the plots by subdivision (Figure 4, upper panels), the first impression is a greater variability in some SDs in which the trends seem to be sharply decreasing throughout time (i.e. SD26 and 27). However, given the limited sample size of the first part of the time series (Figure 4, lower panels), those trends are not considered reliable . Nevertheless, a small increase seems to have occurred in recent years in SD32.

Based on these observations, the experts deemed that there was not enough evidence of a substantial increase in autumn spawners in the area. In line with previous analyses (ICES, 2013), it was decided to produce a maturity ogive based on only the spring spawning part of stock (i.e. from Q1 & 2).



**Figure 3.** Percentage of autumn and spring spawners in the study area (Subdivisions 25 to 27, 29, and 32 excluding the Gulf of Riga)



**Figure 4.** Percentage (upper panels) and numbers (sample size, bottom panels) of autumn and spring spawners by Subdivision in the study area (Subdivisions 25 to 27, 29, and 32 excluding the Gulf of Riga).

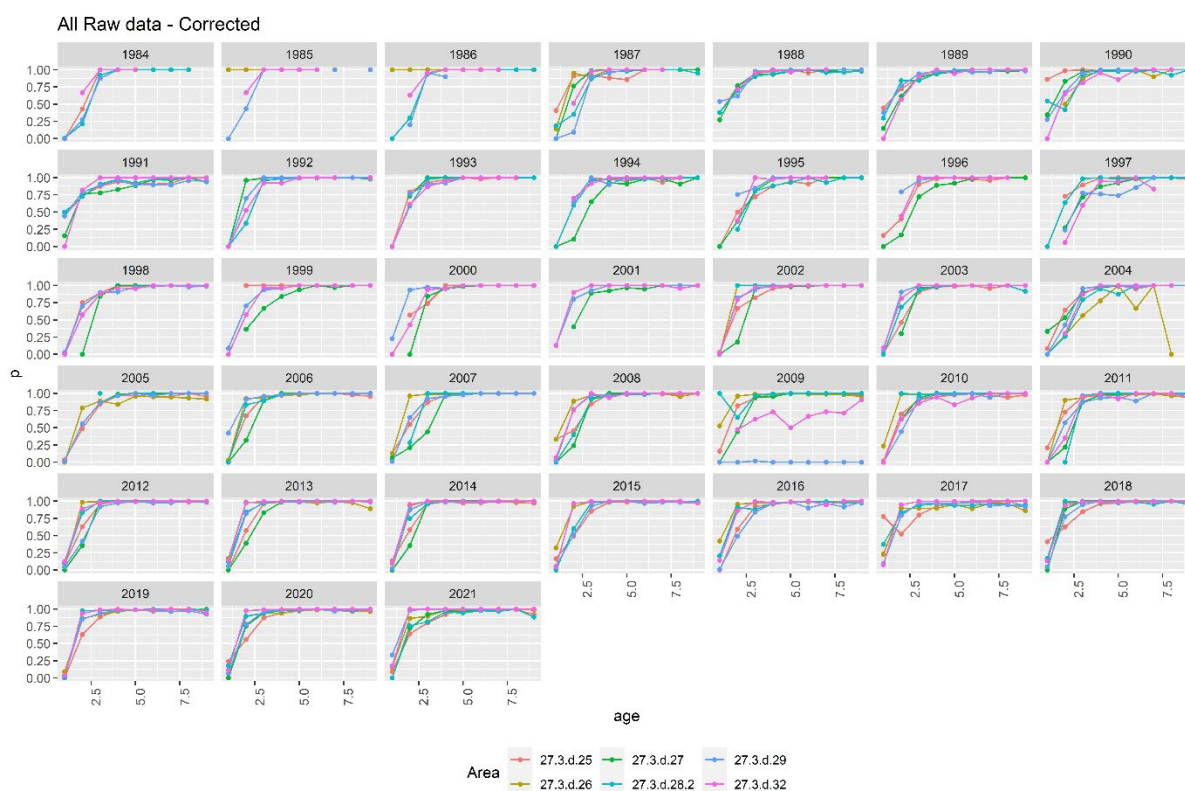
#### 4. Maturity ogive

##### 4.1. Proportion of mature at age

According to standard practice approved by WKMOG (ICES, 2008), the maturity ogive was produced by modelling maturity data as a binomial GLM, where maturity stages were reassigned to a binary response variable: immature (0) and mature (1) (Table 1). Derived proportions of mature individuals at age (0 to 100%) are shown in figure 5.

**Table 1.** Re-assignment of maturity stages from SMSF scale to binary response variable.

SMSF code	Description	New code
A	Juvenile/immature	0
Ba	Developing but functionally immature (first-time developer)	0
Bb	Developing and functionally mature	1
Ca	Actively spawning	1
Cb	Spawning capable	1
Da	Regressing	1
Db	Regenerating	1
E	Omitted spawning	0



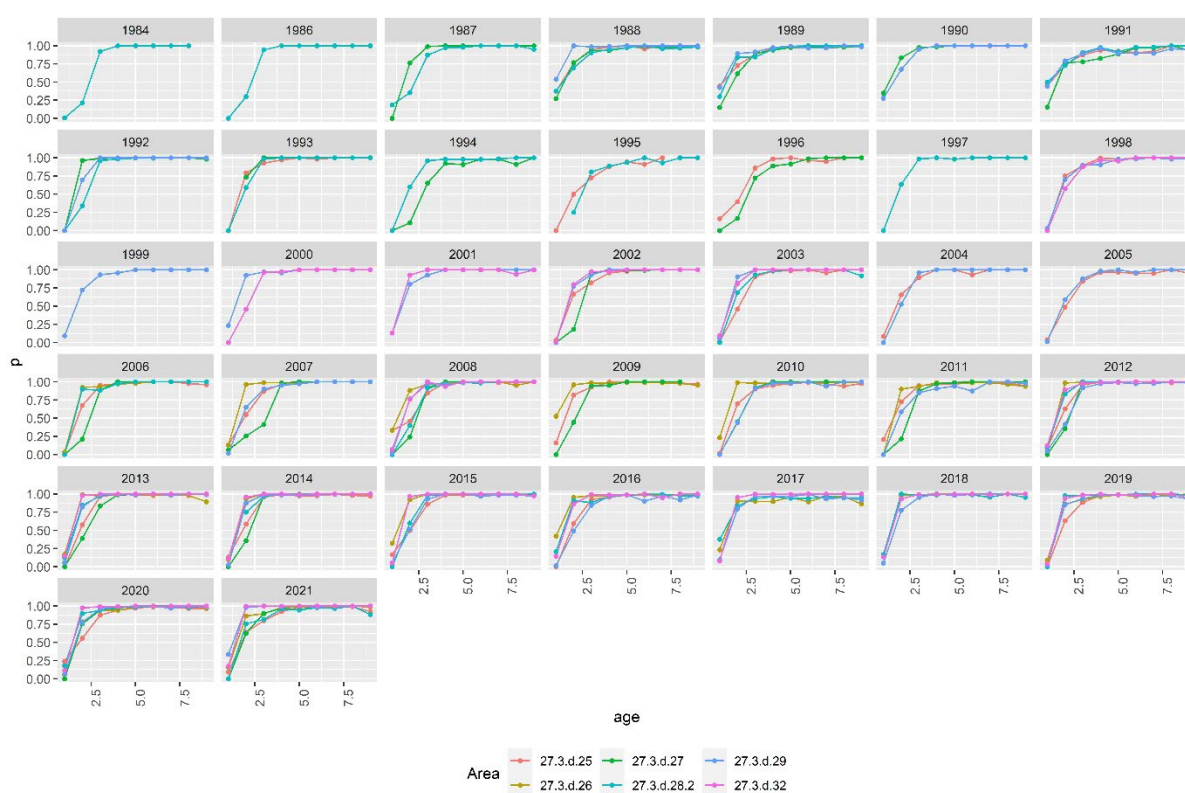
**Figure 5.** Proportion of mature fish ages 1-9+ (9+ represents a plus group) by year and SDs.

Looking at the proportion of mature by year and SD, it is clear that some issues are still unresolved in the data. As an example, SDs 29 & 32 series in 2009 can be easily considered biologically implausible. Nevertheless, given the limited time available, these issues were not adequately explored and addressed during the previous data correction phase (section 2). At this stage of the analyses, the experts decided to remove any implausible data to avoid introducing possible bias in subsequent

analyses. Outlined below is the data “cleaning” process that resulted in the final dataset used for the maturity analyses:

- use the threshold of at least 200 observations (based on the first quantile of the distribution) for the combination of Year\*Quarter\*SD to avoid under-sampling of certain SD/quarters;
- remove as much as possible truncated or incomplete series (especially when maturity data only partially cover the first ages);
- remove remaining outlier values one by one (e.g. in 2017 SD25 proportion of mature in age 1 is higher than age 2).

Figure 6 shows the proportion of mature fish in each age group by subdivision and year after the cleaning process.



**Figure 6.** Proportion of mature fish ages 1-9+ (9+ represents a plus group) by year and SDs after the cleaning process (final dataset used in the GLM analyses).

## 4.2. Modelling

According to evidence of a spatial-temporal trend in the life histories of the stock (ICES, 2022), maturity ogive analyses were conducted on proportion of mature data from 1984 to 2021 using binomial (link = logit) generalized linear model (GLM; generic *glm* function in R), and mixed effect extension (GLMM; *glmmTMB* package in R) taking into account Subdivision and year effect. Sample size has been used as weight in the GLMs.

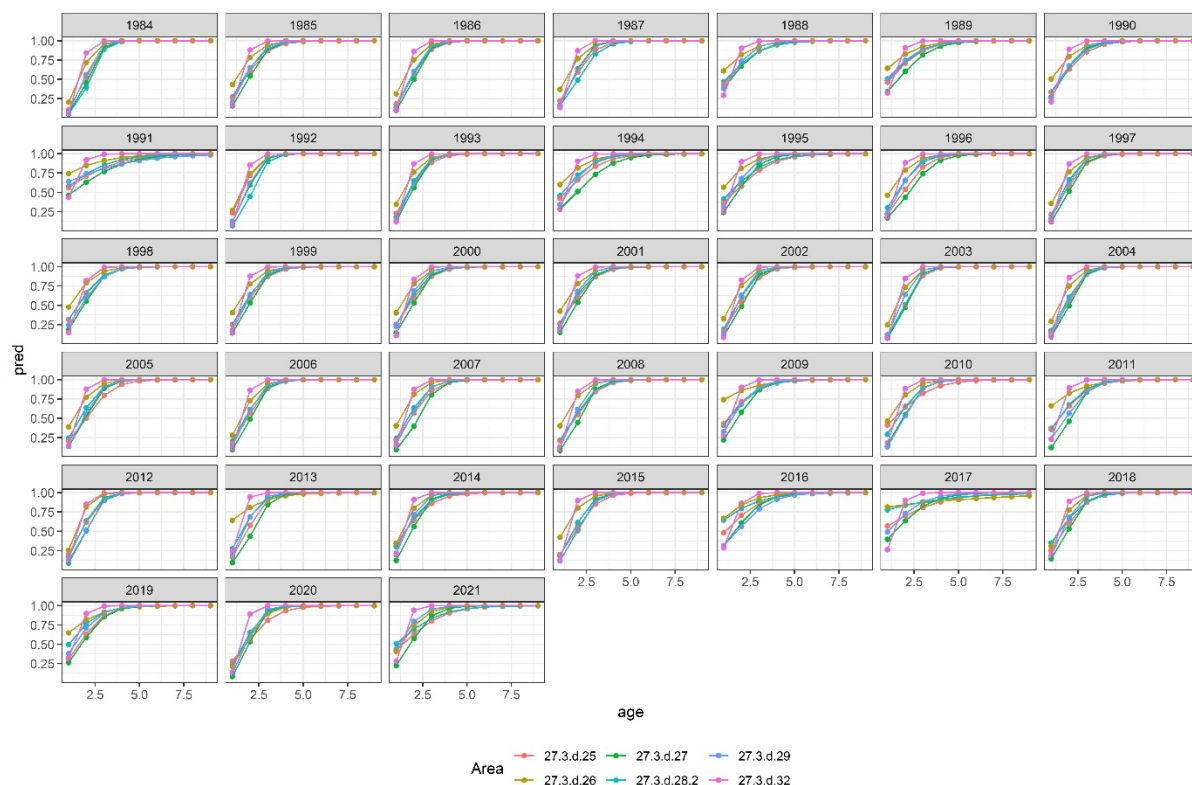
Nine different models have been fitted and compared using Akaike Information Criterion (AIC). A summary of model's structure and selection is provided in the table 2. First, we tested a simple model using only age effect (1). Then Subdivision and Year effects was added as single term both in a fixed effect (2 & 3) and mixed effect (6 & 7) formulation. Subdivision and Year effects was tested together in model 4 (fixed) and 8 (mixed) while interaction between effects was tested in model 5 (fixed) and 9 (mixed).

**Table 2.** Summary table of GLM models structure and model selection statistics (AIC and R<sup>2</sup>).

Model	AIC	R <sup>2</sup>
1) M ~ age	3123	0.827
2) M ~ age + SD	2791	0.843
3) M ~ age + Year	3050	0.833
4) M ~ age + SD + Year	2793	0.843
5) M ~ age + SD * Year	2766	0.844
6) M ~ age + (age   SD)	2476	0.918
7) M ~ age + (age   Year)	2721	0.884
8) M ~ age + (age   SD + Year)	2309	0.927
<b>9) M ~ age + (age   SD * Year)</b>	<b>2248</b>	<b>0.932</b>

\*In bold the selected model

The inclusion of a spatial-temporal structure is justified as the AIC of the model with mixed effect interaction is substantially lower than the AIC of the other models. Model 9 was retained as the best one. Good fitting is confirmed by residual plots below (by Year and SD). Additionally, model 9 is the one with higher coefficient of determination R<sup>2</sup>. Model prediction and is shown in Figure 7.



**Figure 7.** Proportion of mature fish ages 1-9+ (9+ represents a plus group) by year and SDs estimated by the GLMM model.

### 4.3. Spatial-temporal trends in maturation

A50 (age-at-50%-maturity), the midpoint of the modelled ogive, was calculated as follow:

1) on average:  $A50 = -\alpha / \beta$

where  $\alpha$  and  $\beta$  are the intercept and the slope estimated by the model (fixed term);

2) by year to detect possible temporal variation:  $A50_i = -\alpha_i / \beta_i$

where  $\alpha$  and  $\beta$  are the intercept and the slope estimated by the model and  $i$  represent the Year effect and obtained extracting random effect coefficients from the model;

3) by year and SD to detect possible spatial-temporal (3) trends:  $A50_{ij} = -\alpha_{ij} / \beta_{ij}$

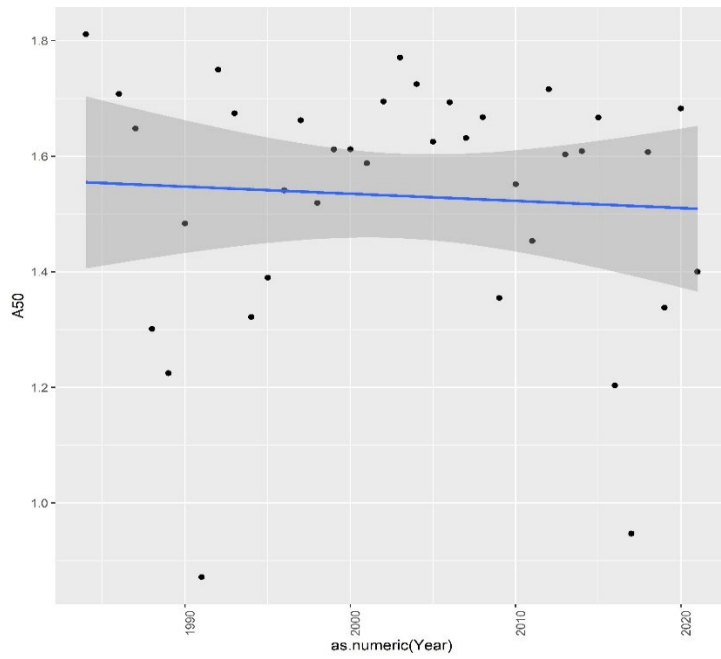
where  $i$  represent the Year effect and  $j$  the SD effect obtained extracting random effect coefficients from the model.

Once point 2 and 3 timeseries were obtained (overall and by SDs), Spearman's rank correlation coefficient ( $\rho$ ) was used to detect possible trends in A50. Here the result in terms of  $\rho$  value and p-value:

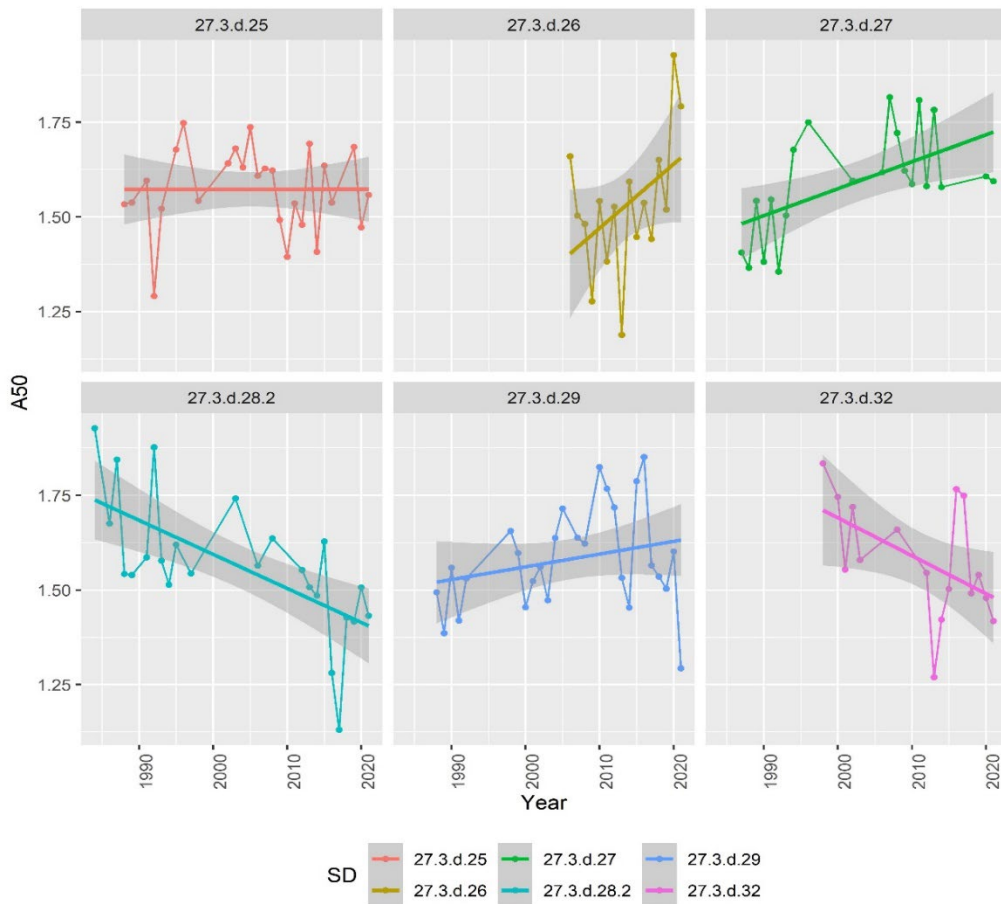
Area	$\rho$	p-value
Overall	-0.11	0.52
SD25	-0.06	0.78
SD26	+0.35	0.18
<b>SD27</b>	<b>+0.55</b>	<b>&lt; 0.05</b>
<b>SD28.2</b>	<b>-0.72</b>	<b>&lt; 0.001</b>
SD29	+0.24	0.22
<b>SD32</b>	<b>-0.54</b>	<b>&lt; 0.05</b>

\*In bold rows denote statistically significant trend

On average, the age at 50% maturity was close to 1.57 years. Figure 8 showed that, despite the high variability in A50, no clear increasing or decreasing trend was seen for time period 1984-2021 (p-value = 0.52). However, looking at the result by SD (Figure 9), a clear decreasing trend was present in SD 28.2 (-0.72; p-value < 0.001) and 32 (-0.54; p-value < 0.05) while an increasing in A50 was detected in SD 27 (+0.55; p-value < 0.05).



**Figure 8.** Age at 50% maturity (A50) by year obtained extracting random effect coefficients from the GLMM model. A smoother has been added to shown trend in time.



**Figure 9.** Age at 50% maturity (A50) by year and SD obtained extracting random effect coefficients from the GLMM model. Smoothers have been added to shown trends in time.



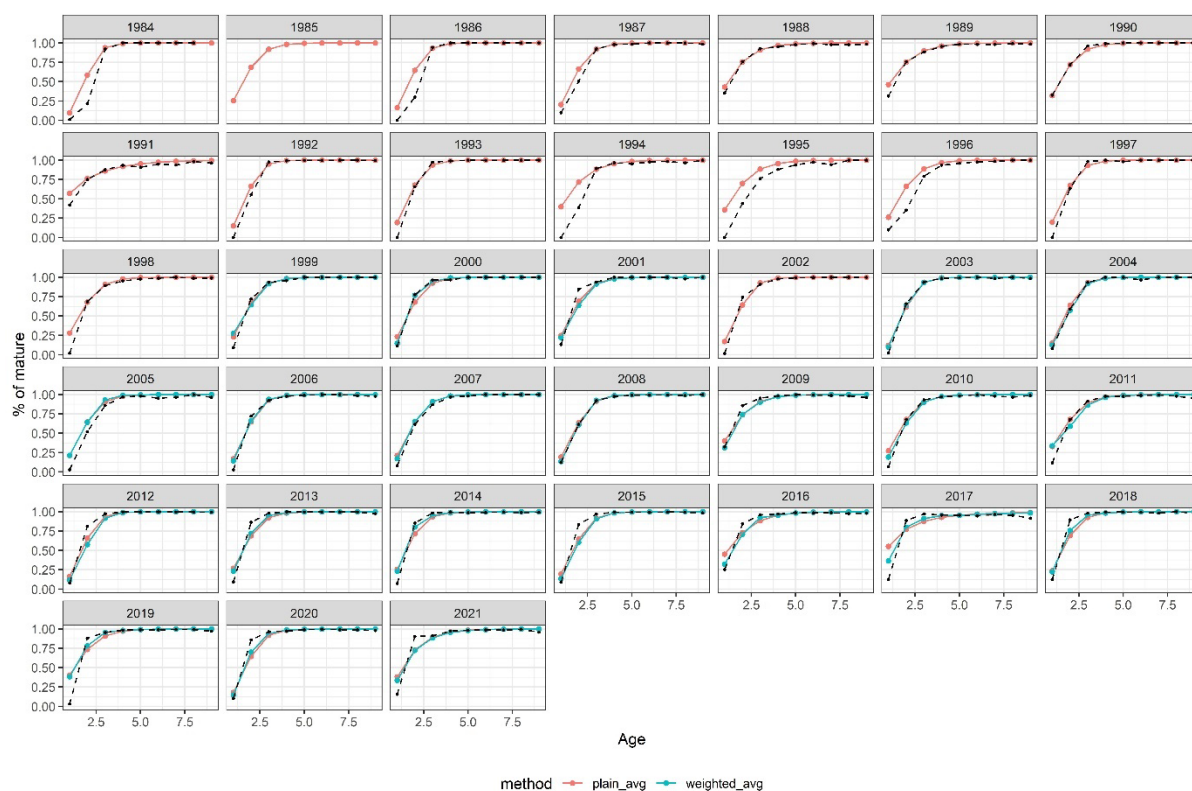
#### 4.4. Maturity ogive matrix

Maturity ogives to be used in the stock assessment were produced as predictions by area and year from the best model. However, since the current stock assessment configuration does not allow results to be used by area, predictions must be averaged over the total area to obtain final matrix of percentage of mature by age and by year. Two procedures were proposed and used here:

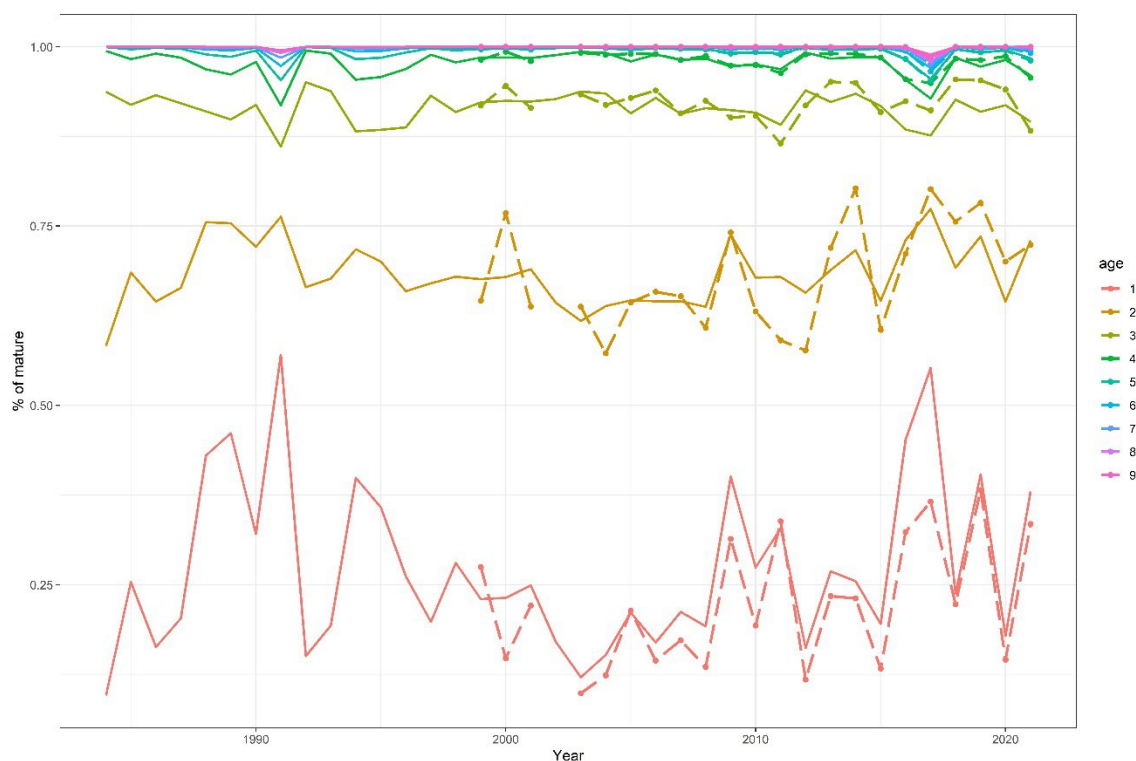
- 1) Plain average over the total area: the predictions of the model were average (by year) without taking into account the different distribution of the species in the study area between different SDs.
- 2) Weighted average: the predictions of the model were average (by year) against the «true» abundance of the species in each SD using a matrix of spatial distribution by BIAS survey (1999-2021)

Since the second option was considered optimal, weighted average was used to build the final ogive matrix by year to be used in stock assessment. Before 1999, where no spatial distribution by BIAS survey was available, option 1 has been used (Figure 10 and 11). Despite being considered less accurate, plain average option still proved to mimic weighted average option and raw data quite well. Final values used in stock assessment model are provided in table 3.

There are no data derived maturity estimates from 1974-1984, hence for this period it is suggested to use the value of the first year available (1984).



**Figure 10.** Proportion of mature fish estimated by the GLMM model and averaged by year following the two different methods. Red line represents plain average (method 1) while blue line weighted average by survey abundance (method 2). Additionally, dotted line represents raw estimates (not modelled).



**Figure 11.** Time series of proportion of mature fish by ages 1-9+ (9+ represents a plus group) estimated by the GLMM model and averaged by year following the two different methods. Continuous lines represent plain averages (method 1) while dotted lines weighted averages by survey abundance (method 2).

**Table 3.** Final proportion of matures at age estimated for central Baltic herring between 1984-2021.

Year	Method	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9+
1984	Plain Avg	0.10	0.58	0.94	0.99	1.00	1.00	1.00	1.00	1.00
1985	Plain Avg	0.25	0.69	0.92	0.98	1.00	1.00	1.00	1.00	1.00
1986	Plain Avg	0.16	0.64	0.93	0.99	1.00	1.00	1.00	1.00	1.00
1987	Plain Avg	0.20	0.66	0.92	0.98	1.00	1.00	1.00	1.00	1.00
1988	Plain Avg	0.43	0.76	0.91	0.97	0.99	1.00	1.00	1.00	1.00
1989	Plain Avg	0.46	0.75	0.90	0.96	0.99	0.99	1.00	1.00	1.00
1990	Plain Avg	0.32	0.72	0.92	0.98	0.99	1.00	1.00	1.00	1.00
1991	Plain Avg	0.57	0.76	0.86	0.92	0.95	0.97	0.98	0.99	0.99
1992	Plain Avg	0.15	0.66	0.95	0.99	1.00	1.00	1.00	1.00	1.00
1993	Plain Avg	0.19	0.68	0.94	0.99	1.00	1.00	1.00	1.00	1.00
1994	Plain Avg	0.40	0.72	0.88	0.95	0.98	0.99	1.00	1.00	1.00
1995	Plain Avg	0.36	0.70	0.88	0.96	0.98	0.99	1.00	1.00	1.00
1996	Plain Avg	0.26	0.66	0.89	0.97	0.99	1.00	1.00	1.00	1.00
1997	Plain Avg	0.20	0.67	0.93	0.99	1.00	1.00	1.00	1.00	1.00
1998	Plain Avg	0.28	0.68	0.91	0.98	1.00	1.00	1.00	1.00	1.00
1999	Weighted Avg	0.27	0.65	0.92	0.98	1.00	1.00	1.00	1.00	1.00
2000	Weighted Avg	0.15	0.77	0.95	0.99	1.00	1.00	1.00	1.00	1.00
2001	Weighted Avg	0.22	0.64	0.91	0.98	1.00	1.00	1.00	1.00	1.00
2002	Plain Avg	0.17	0.64	0.93	0.99	1.00	1.00	1.00	1.00	1.00
2003	Weighted Avg	0.10	0.64	0.93	0.99	1.00	1.00	1.00	1.00	1.00

2004	Weighted Avg	0.12	0.57	0.92	0.99	1.00	1.00	1.00	1.00	1.00
2005	Weighted Avg	0.21	0.64	0.93	0.99	1.00	1.00	1.00	1.00	1.00
2006	Weighted Avg	0.14	0.66	0.94	0.99	1.00	1.00	1.00	1.00	1.00
2007	Weighted Avg	0.17	0.65	0.91	0.98	1.00	1.00	1.00	1.00	1.00
2008	Weighted Avg	0.14	0.61	0.92	0.99	1.00	1.00	1.00	1.00	1.00
2009	Weighted Avg	0.31	0.74	0.90	0.97	0.99	1.00	1.00	1.00	1.00
2010	Weighted Avg	0.19	0.63	0.90	0.97	0.99	1.00	1.00	1.00	1.00
2011	Weighted Avg	0.34	0.59	0.87	0.96	0.99	1.00	1.00	1.00	1.00
2012	Weighted Avg	0.12	0.58	0.92	0.99	1.00	1.00	1.00	1.00	1.00
2013	Weighted Avg	0.23	0.72	0.95	0.99	1.00	1.00	1.00	1.00	1.00
2014	Weighted Avg	0.23	0.80	0.95	0.99	1.00	1.00	1.00	1.00	1.00
2015	Weighted Avg	0.13	0.61	0.91	0.99	1.00	1.00	1.00	1.00	1.00
2016	Weighted Avg	0.32	0.71	0.92	0.95	0.98	0.99	1.00	1.00	1.00
2017	Weighted Avg	0.37	0.80	0.91	0.95	0.95	0.97	0.98	0.98	0.99
2018	Weighted Avg	0.22	0.76	0.95	0.98	1.00	1.00	1.00	1.00	1.00
2019	Weighted Avg	0.38	0.78	0.95	0.98	0.99	1.00	1.00	1.00	1.00
2020	Weighted Avg	0.15	0.70	0.94	0.99	1.00	1.00	1.00	1.00	1.00
2021	Weighted Avg	0.33	0.72	0.88	0.96	0.98	0.99	1.00	1.00	1.00

## 5. Discussion and recommendations

This document highlights that, despite no clear overall increasing or decreasing trend is seen, higher variability in maturity (A50 used as proxy) can be detected looking at the data by Subdivisions (spatial-temporal variability). Although this variability could also be influenced by the different sample size by area and year, these results are in concordance with general evidence of spatial-temporal trends in the life histories of the herring stock (ICES, 2022).

These analyses support the effort to move towards an area-based stock assessment model in future benchmark but further work needs to be done in this regard. For instance, considering the amount of work done for reviewing and (partially) correcting the raw data coming from the official data call, it is evident that there are discrepancies in the interpretation of internationally agreed maturity scale SMSF by the different nations. Hence, further exploration of the previously identified issues (section 2 and 4.1) by WGBIOP (Working Group on Biological parameters) is strongly recommended.

Nevertheless, time-varying component addressed in this document are considered a good improvement compared to the constant maturity ogive previously adopted for the stock (ICES, 2013). Moreover, moving to time-varying maturity ogive gives the ability to detect changes in maturity ogive in real time.

## 6. References

ICES, 2022. Baltic Fisheries Assessment Working Group (WGBFAS). ICES Scientific Reports. 4:44. 659 pp. <http://doi.org/10.17895/ices.pub.19793014>.

ICES, 2018. WKASMSF - Report of the Workshop for Advancing Sexual Maturity Staging in Fish (WKASMSF). ICES Expert Group reports (until 2018). Report. <https://doi.org/10.17895/ices.pub.19212915.v1>.

ICES. 2013. Report of the Benchmark Workshop on Baltic Multispecies Assessments (WKBALT), 4–8 February 2013, Copenhagen, Denmark. ICES CM 2013/ACOM:43.399 pp.

ICES, 2008. Report of the Workshop on Maturity Ogive Estimation for Stock Assessment (WKMOG), 3–6 June 2008, Lisbon, Portugal. ICESCM2008/ACOM:33. 72 pp.

## WD7: Natural Mortality estimated using Life History based methods

Stock - Herring (*Clupea harengus*) in subdivisions 25-29 and 32, excluding the Gulf of Riga (central Baltic Sea)

BWKBALPEL 2023

Authors: Francesco Masnadi, Massimiliano Cardinale, Szymon Smoliński, Mikaela Bergenius Nord

The natural mortality rate ( $M$ ) of fish populations is one of the most important parameters for population dynamics and stock assessment models (Mannini et al., 2021). Unfortunately, it is also one of the most difficult parameters to estimate. Moreover, in this particular case, the natural mortality has to be divided in two fraction cause of the cod predation: residual natural mortality ( $M1$ ) and predation mortality ( $M2$ ). In the current assessment,  $M1$  is assumed as known and constant over time and ages while SMS (Lewy and Vinther, 2004) provides predation mortality for cod ( $M2$ ). As part of the 2019 SMS keyrun it was decided to use 0.1 for herring (annual values), with the justification that this are the values used for the same species for the North Sea keyrun. However, there are no data to actually support the exact values chosen for  $M1$ .

In this context, the Barefoot Ecologist's Toolbox ([http://barefootecologist.com.au/shiny\\_m](http://barefootecologist.com.au/shiny_m)) can be used to derive different values of single  $M$  or to derive composite  $M$  value weighting different method. This toolbox, developed by Jason Cope, provides a straightforward method for obtaining the estimated value of natural mortality from a range of life-history based methods (different life-history input requirement) that can be used to derive proxy of  $M1$  to be used in SMS.

In this WD length-at-age data from 1984 to 2021 (Fig. 1) has been used to derive VB parameters ( $L_{inf}$ ,  $k$  and  $t_0$ ) to be used as input parameters in the toolbox to provide empirical estimation of natural mortality. Moreover, according to ICES (ICES, 2022), spatial-temporal trends in growth has been also analyzed and take into account in the estimation of VB parameters supporting the use of a time-varying  $M$ .

### 1) Data exploration

The data used for the following analyses has been requested by the ICES benchmark workshop on Baltic Pelagic stocks (BWKBALPEL) to be used in assessments model and as basis for advice for the Central Baltic herring stock. The data contains biological information for herring, *Clupea harengus*, in the Central Baltic Sea, ICES Division 3d (Subdivisions 25 to 27, 29, and 32). During the Official Data preparation meeting (17/11/2022), the data has been cleaned to avoid under-sampling of certain area/quarters and to remove outlier as follows:

- use the threshold of at least 250 observations for the combination of Year\*Quarter\*SD;
- remove quarter 1&2 from age-0 data (biologically illogical);
- remove remaining outlier values one by one.

Figure 2 shows length-at-age data by Subdivision and quarter after the cleaning process.

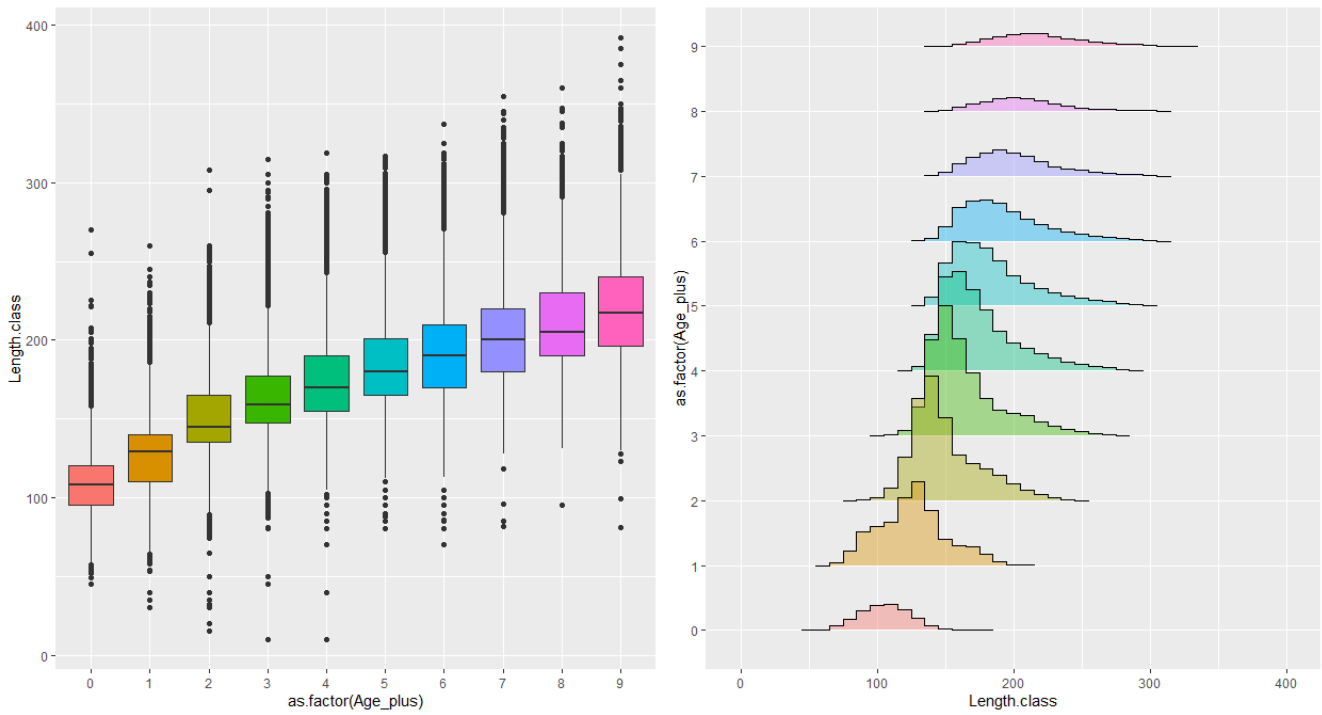


Figure 1. Length-at-age data for central Baltic Sea herring (raw data).

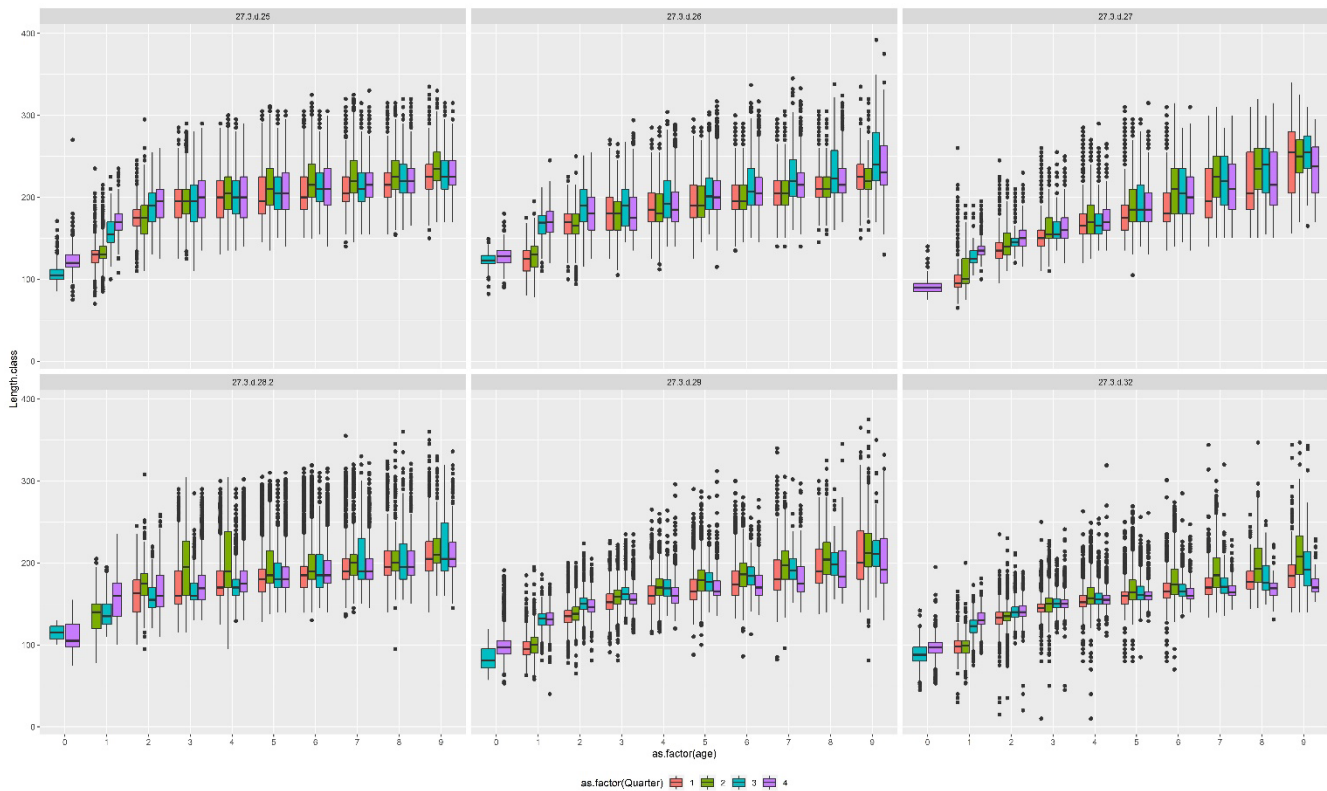


Figure 2. Length-at-age data for central Baltic Sea herring by Subdivision and quarter (cleaned data).

## 2) Spatial-temporal trends in mean length-at-age

According to the last report from the Baltic Fisheries Assessment Working Group (ICES, 2022), a marked decrease in mean weights-at-age (kg) that started in the early 1980s ceased around the mid-1990s and remains at this low level. Moreover, marked geographical differences were also detected in the mean weight with higher values in Subdivisions 25 & 26 than in the more northern ones. The decrease in weight-at-age does not by itself represent evidence of differences in growth because a population can decrease in average weight over a period (e.g. due to resource scarcity) without necessarily decreasing in length. It is therefore aim of these analyses to investigate if decreasing trends are also present in length-at-age data and if they could be considered as symptoms of a real change in growth rate. Length-at-age data reveals a decreasing trend all along the timeseries, especially for the older ages (Fig. 3). Differences by Subdivision were also detected with higher mean values in southern SDs (25 & 26) and lower in the northern ones (29 & 32). A mixture behavior was detected for SDs 27 & 28.2. As a final confirmation, an overall decreasing trend over time is present in all SDs (Fig. 4).

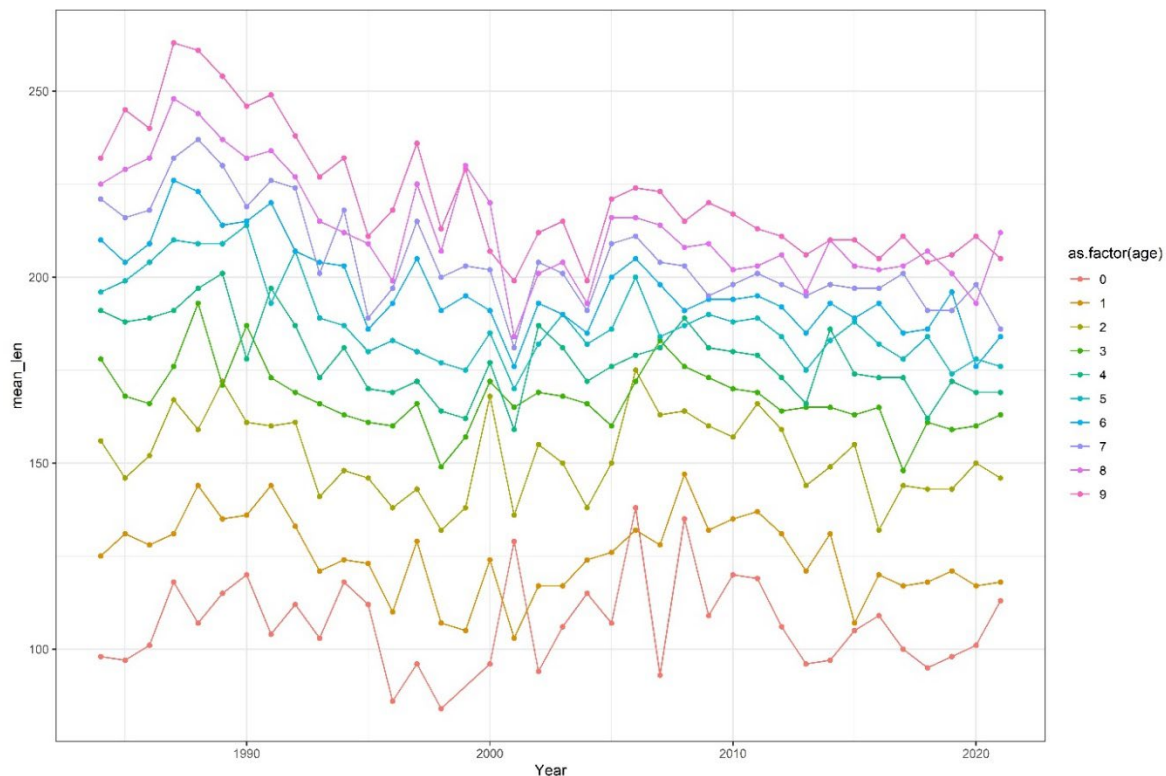


Figure 3. Temporal trends in the mean Length-at-age (mm) in central Baltic Sea herring commercial data.

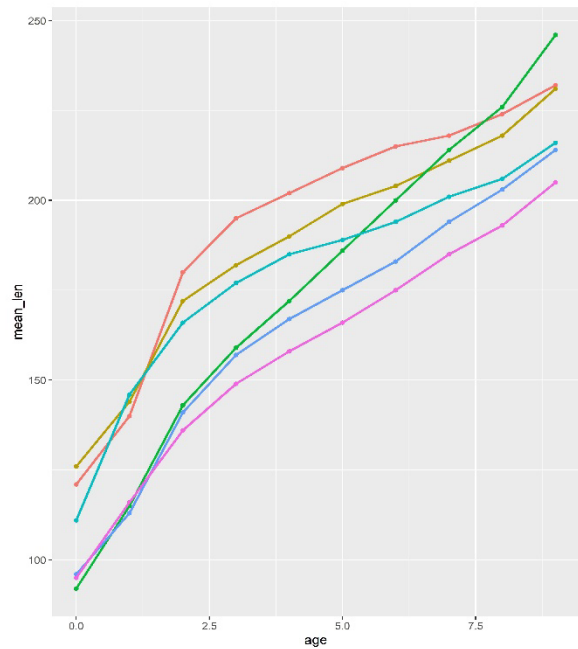
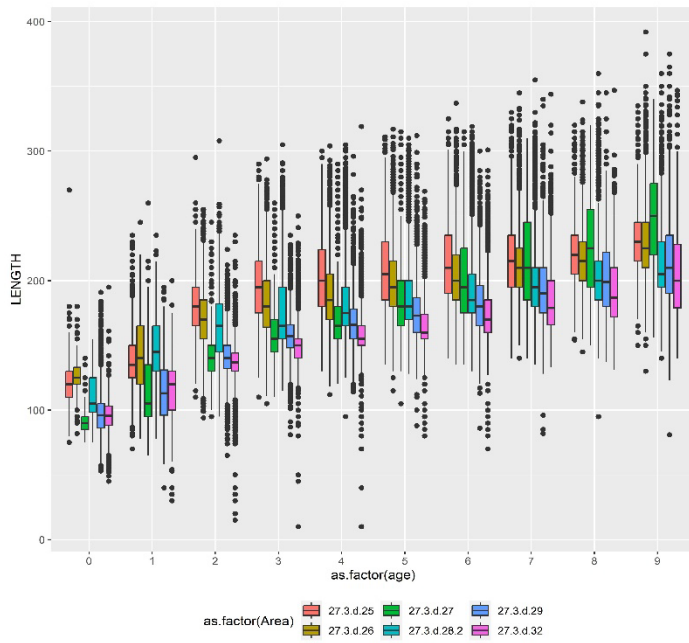
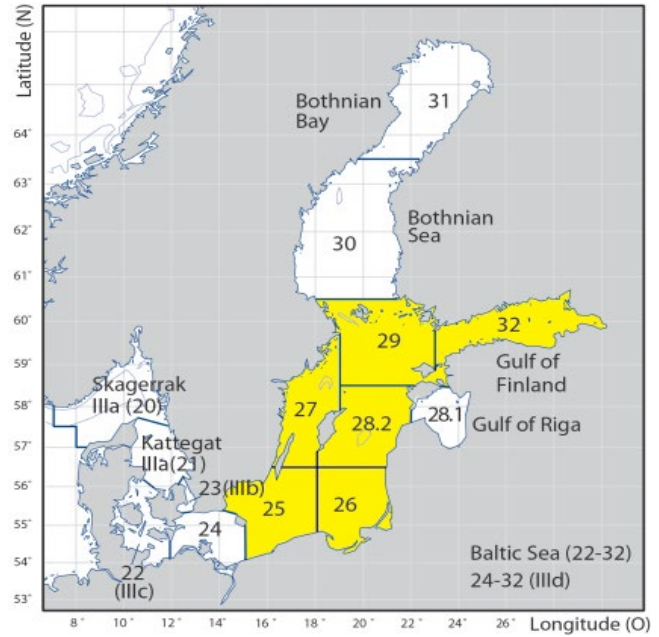
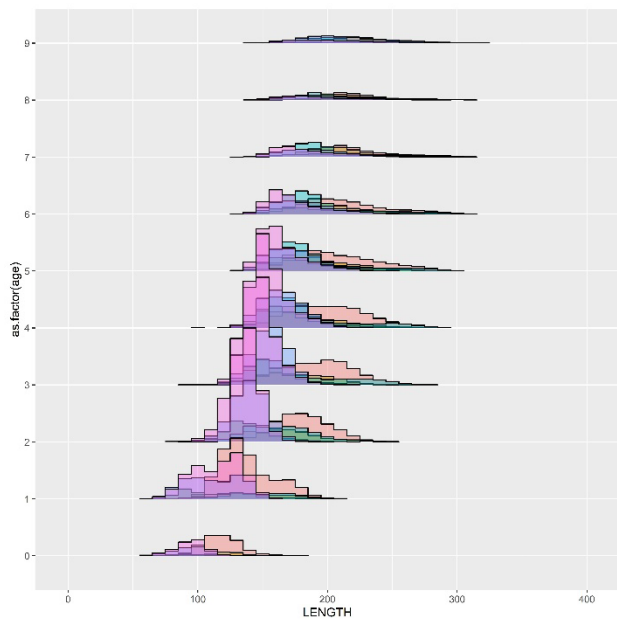


Figure 3. Spatial trends in the mean Length-at-age (mm) by Subdivisions in central Baltic Sea herring commercial data.



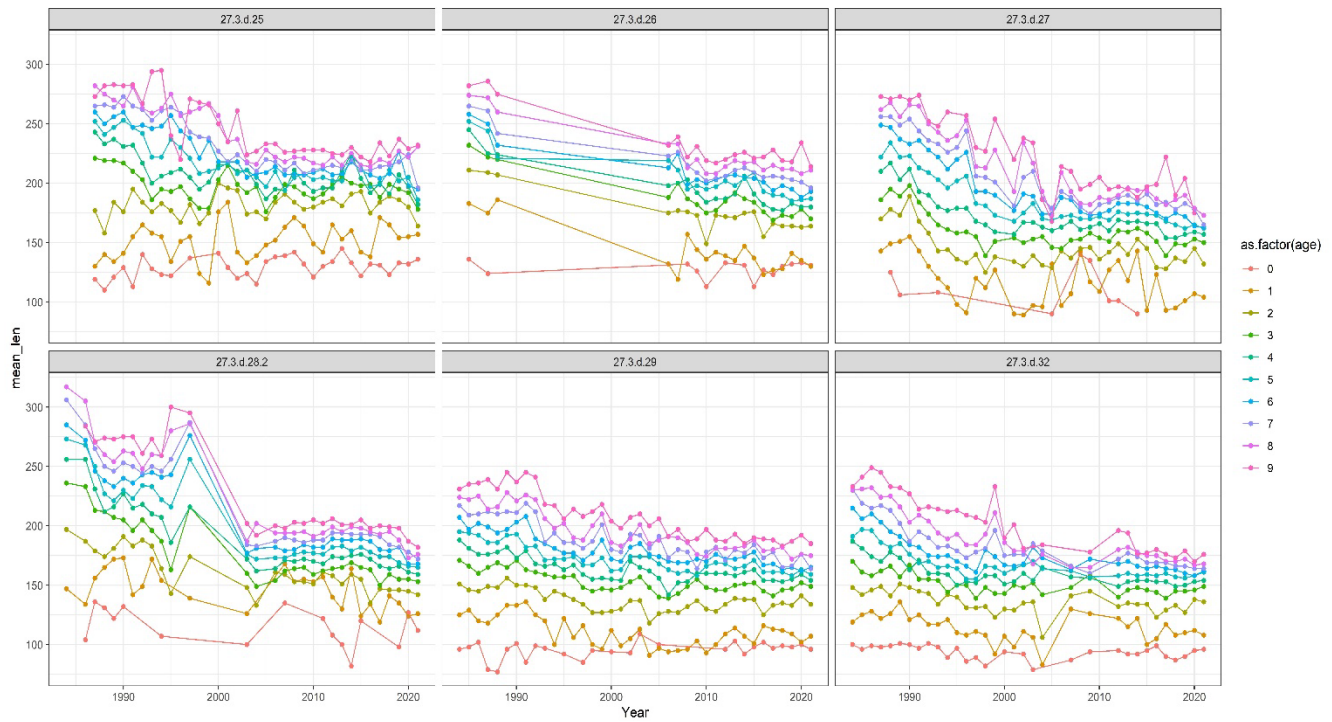


Figure 4. Spatial-temporal trends in the mean length-at-age (mm) by Subdivisions in central Baltic Sea herring commercial data.

### 3) von Bertalanffy Growth Parameters

Based on this consideration, VB growth curve equation has been fitted to the data from 1984 to 2021 using nonlinear mixed effect model (*nlme* package in R) using Year and SD as random effects to be able to taking into account variability in growth between year and areas.

Four different models have been fitted and compared using Akaike Information Criterion (AIC). First, we tested a simple model without effects at all (no variability in time and space). Then, we tested both the single term effect (spatial and temporal variability separately) and the interaction between effect (spatial-temporal variability).

Here the result in terms of AIC value:

Model	AIC
1) No random effect	4183324
2) Year	4121555
3) SD	4033885
4) SD in Year	3801474

The inclusion of a spatial-temporal structure is justified as the AIC of the model with random effect in both Year and SD is substantially lower than the AIC of the other models. Model (4) was retained as the best growth function. Good fitting is confirmed by residual plots below (by Year and SD). High variability in residuals mirrors the general high variability in length-at-age data.

Additionally, a good match of mean length to age derived using the model prediction and raw data is found in Figure 6.

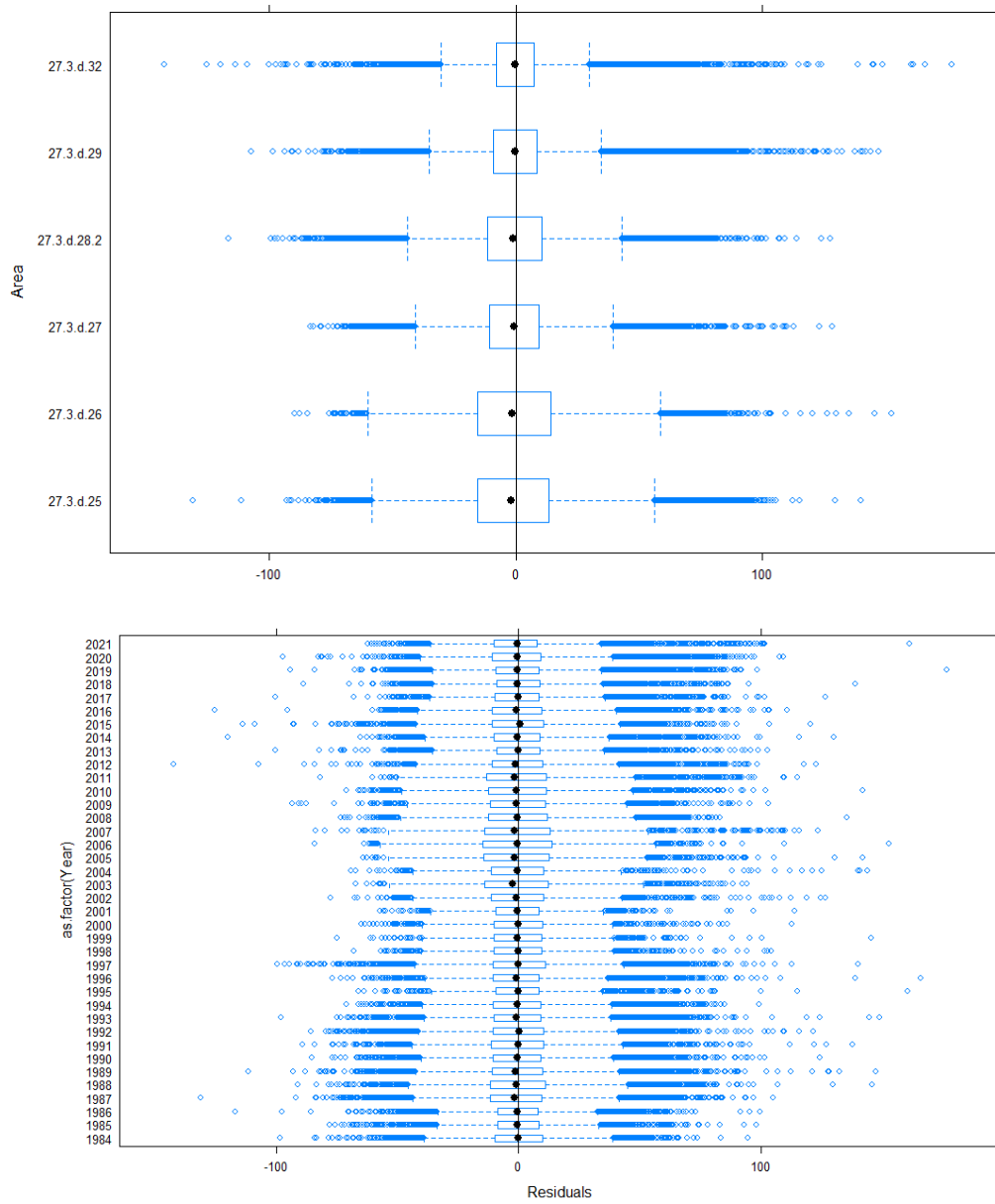


Figure 5. Residual boxplot of the best model (4) by Subdivision (top panel) and year (bottom panel)

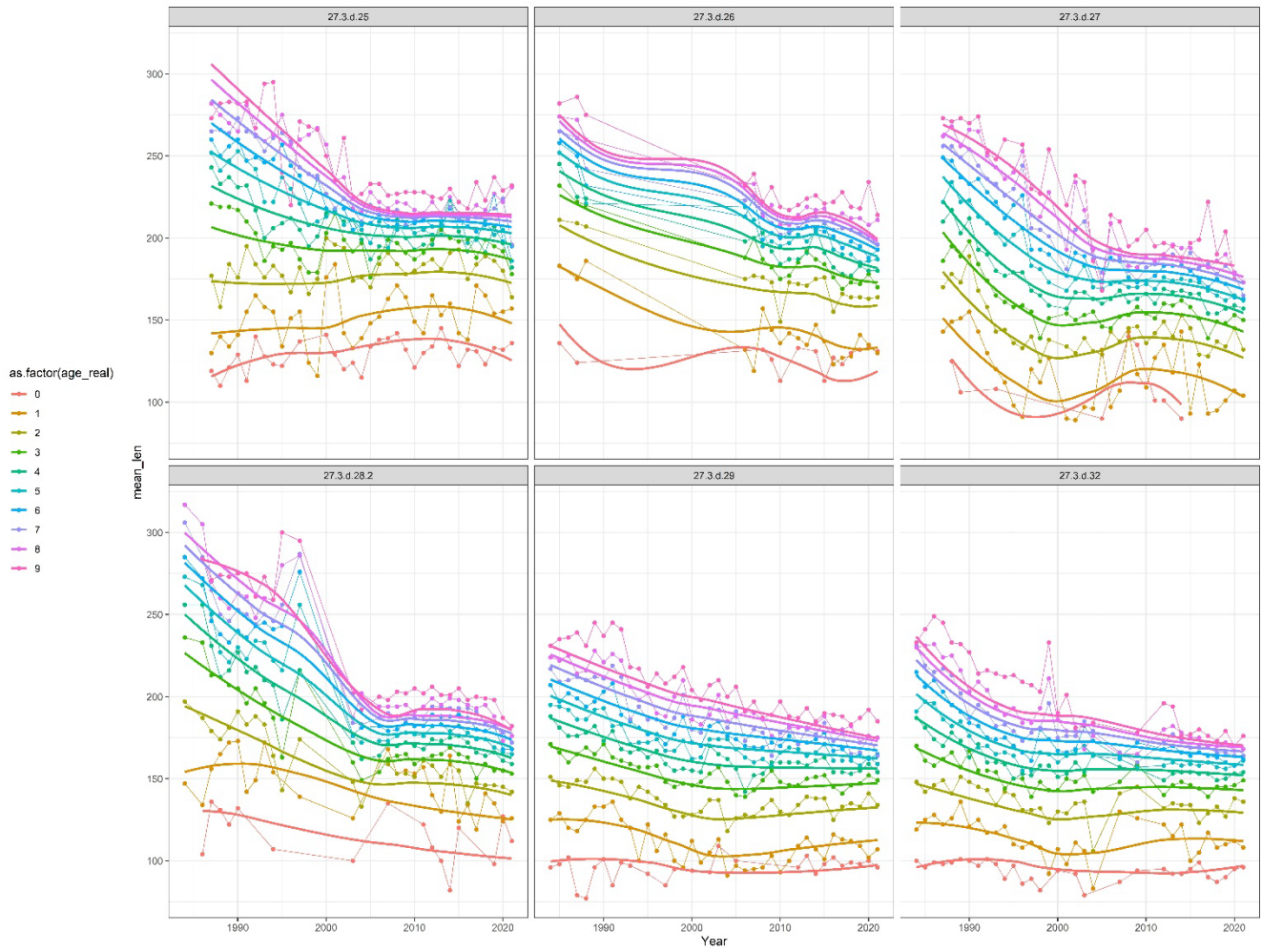


Figure 6. Mean length-at-age by Year and SD derived using model prediction (smoothers) against raw data (points and lines).

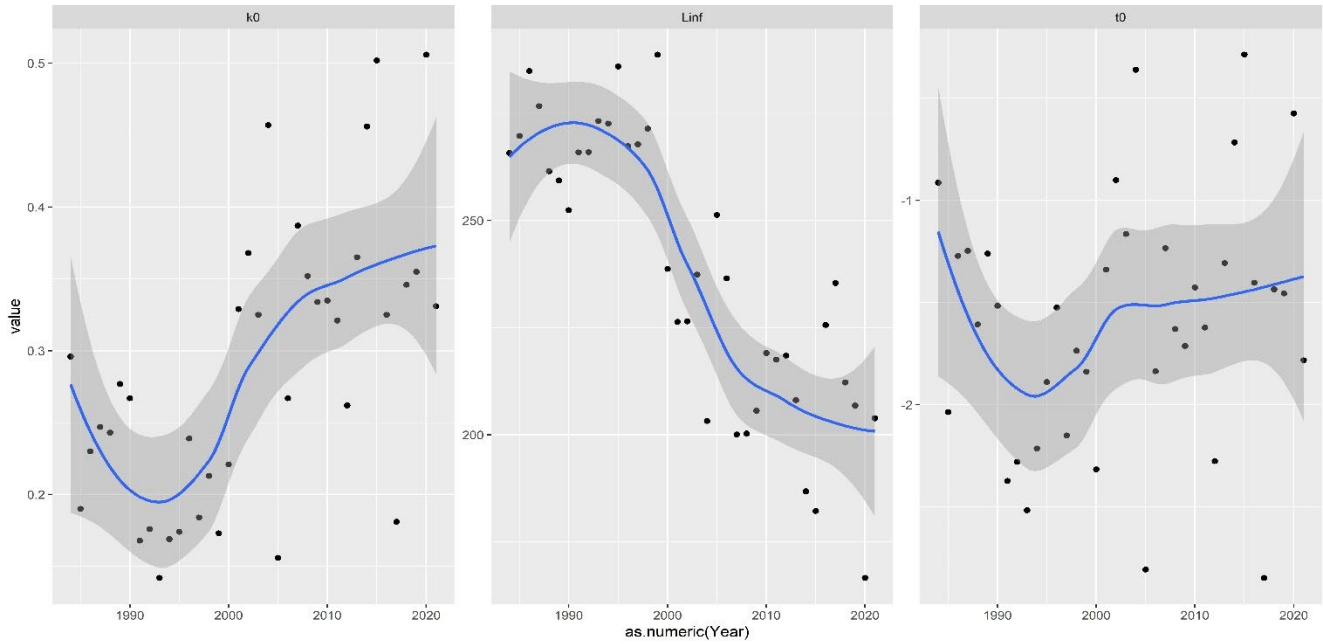


Figure 7. VB equation parameter ( $k$ ,  $Linf$  &  $t0$ ) per year, smoothers have been added to shown trends in time.

Table 1. Value of VB equation parameter ( $k$ ,  $Linf$  &  $t0$ ) per year.

Effect	Year	$k$	$Linf$	$t0$
Year	1984	0.30	265.8	-0.91
Year	1985	0.19	269.7	-2.04
Year	1986	0.23	284.9	-1.27
Year	1987	0.25	276.7	-1.25
Year	1988	0.24	261.5	-1.61
Year	1989	0.28	259.4	-1.26
Year	1990	0.27	252.4	-1.52
Year	1991	0.17	265.9	-2.37
Year	1992	0.18	265.9	-2.28
Year	1993	0.14	273.2	-2.52
Year	1994	0.17	272.6	-2.22
Year	1995	0.17	285.9	-1.89
Year	1996	0.24	267.4	-1.52
Year	1997	0.18	267.8	-2.15
Year	1998	0.21	271.4	-1.74
Year	1999	0.17	288.7	-1.84
Year	2000	0.22	238.7	-2.32
Year	2001	0.33	226.4	-1.34
Year	2002	0.37	226.5	-0.90
Year	2003	0.33	237.3	-1.16
Year	2004	0.46	203.2	-0.36
Year	2005	0.16	251.3	-2.81

Year	2006	0.27	236.5	-1.84
Year	2007	0.39	200.1	-1.23
Year	2008	0.35	200.3	-1.63
Year	2009	0.33	205.6	-1.71
Year	2010	0.34	219.1	-1.43
Year	2011	0.32	217.5	-1.62
Year	2012	0.26	218.5	-2.28
Year	2013	0.37	208.1	-1.31
Year	2014	0.46	186.8	-0.72
Year	2015	0.50	182.2	-0.29
Year	2016	0.33	225.6	-1.40
Year	2017	0.18	235.4	-2.85
Year	2018	0.35	212.2	-1.44
Year	2019	0.36	206.8	-1.46
Year	2020	0.51	166.6	-0.57
Year	2021	0.33	203.8	-1.78

Given the overall decreasing trend of growth over the years, the VB equation parameters per year (Figure 7 & Table 1), i.e. fixed effects plus Year random effects component, have been used as input value in the Barefoot Ecologist's Toolbox to produce a time-varying natural mortality vector. The VB parameters  $Linf\ k$  and  $t0$  obtained from the analyses together with a proxy of longevity (yrs) set at age 20, allowed the use of the following 7 methods of the toolbox:

- *Then\_nls, Then\_lm, Then\_VBGF*; based on Then et al., 2015;
- *Hamel\_Amax, Hamel\_k*; based on Hamel, 2015;
- *Jensen\_k1, Jensen\_k2*; based on Jensen, 1996 & 1997.

A CV of 0.2 has been provided to add additional uncertainty to the point estimates and the results from the 7 different methods has been combined in a composite M weighting the contribution of each method in the final distribution based on redundancies of methods using similar information. For instance, the three longevity-based methods (*Then\_nls, Then\_lm, Hamel\_Amax*) are given a weight of 0.33, so all weighted together equal 1. Median value of the composite M distribution has been calculated for all the year of the time series producing a time-varying natural mortality vector. Once obtain the final vector by year, OLS-based CUSUM test (Ploberger & Krämer, 1992) has been used to detect possible breakpoints in the M value time series. CUSUM (or cumulative sum control chart) is a sequential analysis technique that can be used for testing, monitoring and dating structural changes in regression models. The test detected a significant breakpoint corresponding to year 2000 of the time series (Figure 8;  $p\text{-value} < 0.001$ ) with M showing a higher values after the breakpoint.

Given the change in mortality after 2000, the mean M value calculated before and after the breakpoint (M before 2000: 0.28; M after 2000: 0.38) have been used to re-scale the assumed annual M1 (0.05, 0.1, 0.2) in SMS (scenarios for "likely" M1 presented in WD **XX**).

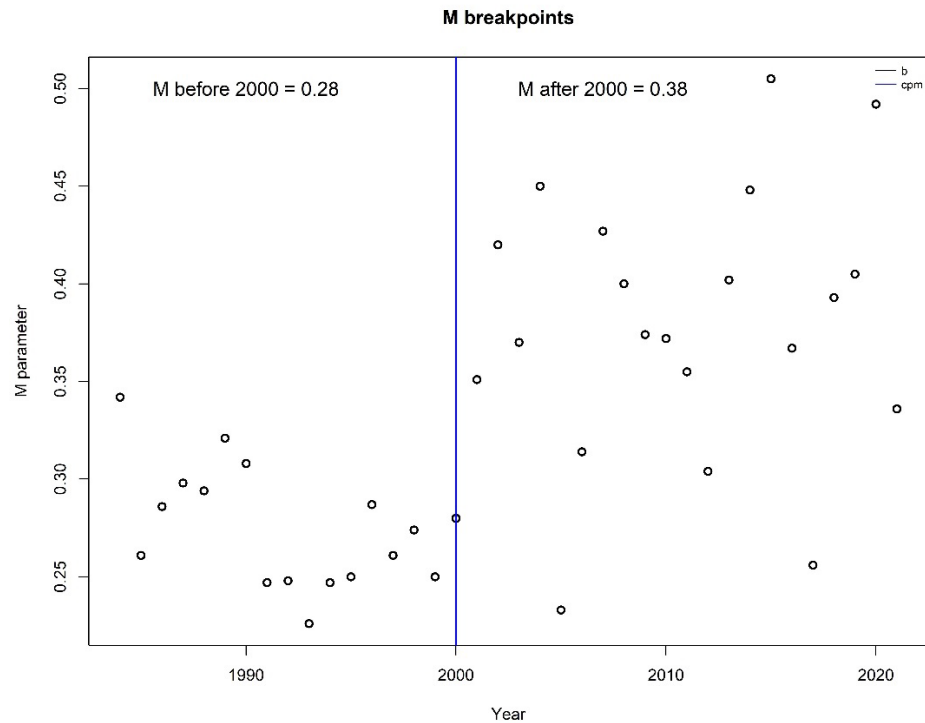


Figure 8. Composite natural mortality time series obtain combining different methods of Barefoot Ecologist's Toolbox. The blue line represents the breakpoint detected from the OLS-based CUSUM test. Mean M value before and after the breakpoint are given in the plot.

## References

- Hamel, O.S., 2015. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. *ICES Journal of Marine Science* 72, 62-69. <https://doi.org/10.1093/icesjms/fsu131>.
- ICES. 2022. Baltic Fisheries Assessment Working Group (WGBFAS). *ICES Scientific Reports*. 4:44. 659 pp. <http://doi.org/10.17895/ices.pub.19793014>.
- Jensen, A.L. 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. *Can. J. Fish. Aquat. Sci.* 53: 820-822.
- Jensen, A.L. 1997. Origin of the relation between K and Linf and synthesis of relations among life history parameters. *Can. J. Fish. Aquat. Sci.* 54: 987-989.
- Lewy, P., & Vinther, M. 2004. A stochastic age-length-structured multispecies model applied to North Sea stocks. *ICES CM*, 33.
- Mannini, A., Pinto, C., Konrad, C., Vasilakopoulos, P., & Winker, H. 2020. "The Elephant in the Room": Exploring Natural Mortality Uncertainty in Statistical Catch at Age Models. *Frontiers in Marine Science*, 7, 585654.
- Ploberger W., Krämer W. 1992. The CUSUM test with OLS residuals, *Econometrica*, 60, 271-285.

- Then, A.Y., J.M. Honeig, N.G. Hall, D.A. Hewitt. 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. *ICES Journal of Marine Science* 72(1): 82-92.

## WD8\_Working Document on potential corrections of national catch data of Baltic sprat and central Baltic herring

To be submitted to the datacompilation meeting

Authors: All participants in the ISSG

### 1. Background

Issue list from AWG, long standing discussion on species misreporting of herring and sprat, other types of errors that the national catch data might be associated with.

### 2. Approach taken by the ISSG Small Pelagic Fisheries Baltic

In the 2021 RCG decision meeting (D07), 8 member states (MS) (Germany, Denmark, Finland, Poland, Lithuania, Estonia, Latvia, Sweden) agreed to :

Each MS with trawlers fishing small pelagics in the Baltic need to decide if they can commit to an analysis of potential “historical” misreporting of the proportion of herring and sprat in their national data. The commitment includes to perform an analysis, to present it at the ISSG small pelagics in the Baltic and to decide if historical catch data should be corrected on the basis of the analysis. Deadline for the analysis is October 2022. The aim is to feed in the overall outcome to the benchmark process of central Baltic herring and sprat 2023.”

Two meetings has been conducted in 2022 (18-19 January and 10 May)

In the first meeting the stock assessors for the sprat and herring stock were invited to the meeting to get the end-users perspective. It was decided during the meeting to:

- Document present WGBFAS time series in respect to corrections.
  - Fill in a template about corrections done (or not done) in connection to historic misreporting.
- Analyze if it is possible for MS’s to use some quality indicators to check if there has been inconsistency between official numbers in catch composition and data from alternative sources (national control data, Danish control data, observer trips, scientific surveys)
- Collate quota shares by year and country

### 3. Country specific chapters on potential corrections

**(each country have the same headers)**

#### 3.1. Denmark

The Danish fishery for sprat and herring in the Baltic is mainly used for industrial purpose, see figure 1. Denmark has a relatively small part (<3%) of the CBH EU quota and for Baltic sprat Denmark has 10% of the quota and landings (figure 2).

There are two different types of Danish fleets conducting the fishing. Some fewer relatively large vessels >25 meters fishing in many ICES squares and some smaller local fishermen fishing very close to the Island of Bornholm fishing mainly in SD 25.



The relative Danish quota share has changed over time and in 2021 79% of the total sprat and herring landings were sprat.

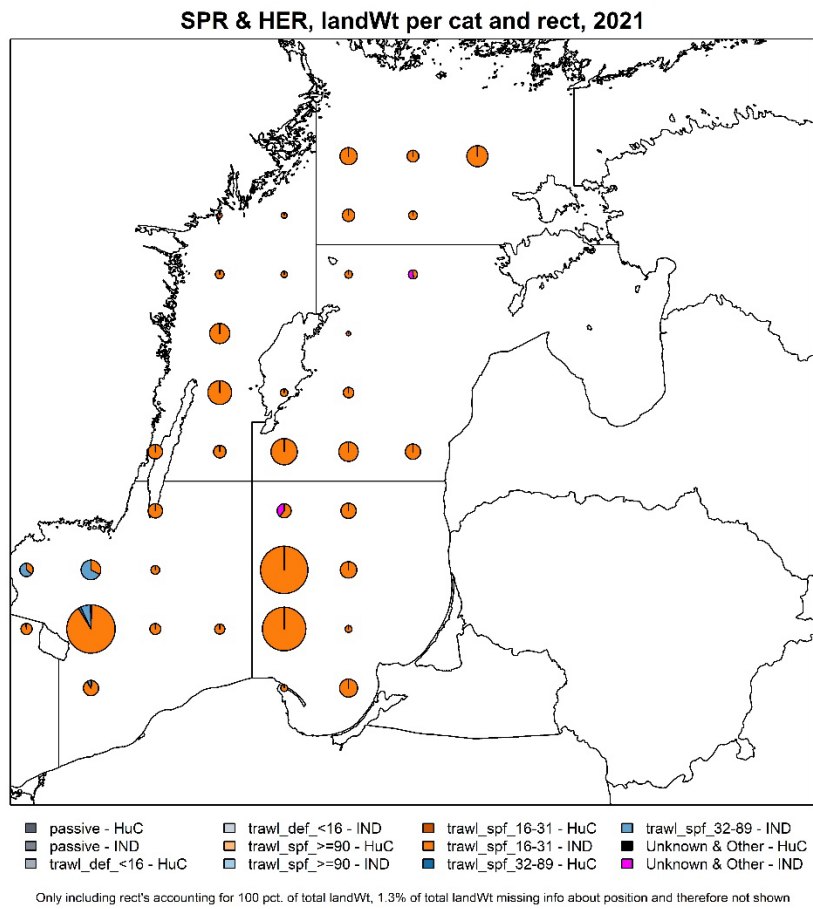


Figure 1 Landings in 2021 from Danish vessels by métier and purpose (Human consumption/ Industry)

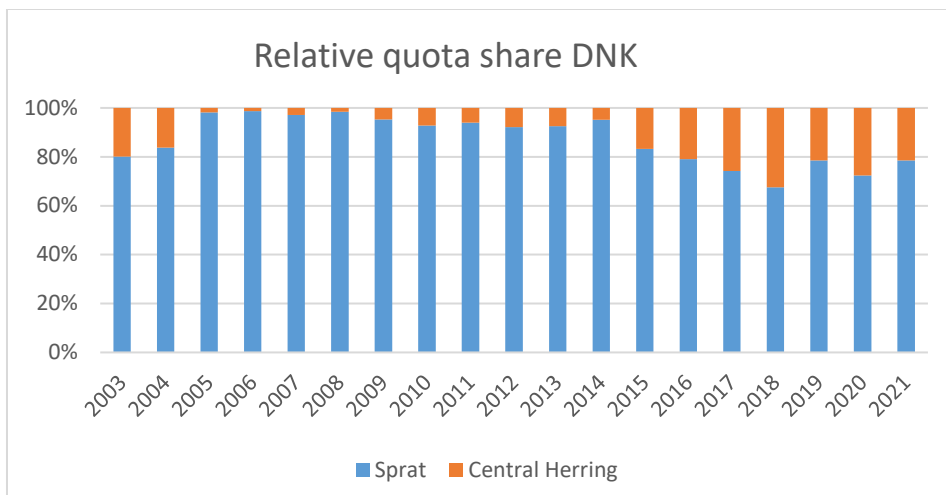


Figure 2 The relative share of sprat (blue) and central herring quota in the Danish Baltic fishery.

### *Approach taken to analyze if there are errors in the time series of catch data due to inadequate reporting of species and/or other reasons*

#### Reporting of species

In Denmark sale notes are used to subtract quota and delivering data for ICES for all fisheries, but fisheries for reduction.

In the fisheries for reduction a species composition is estimated, either with the 9-square method, sub-sampling of all landings or by the fisherman (license 1205) and the results are reported to ICES, see table 1 for an overview of where and when the different methods have been used. In respect to quotas subtraction the estimated species composition is only used for license 1205 and herring with a by-catch quota, herring in the North Sea, Skagerrak and Kattegat, all other species are subtracted at the level of the target species (at least this is true for the time after 2000). Therefore, the total amount caught are often only reported for the target species in the sales notes, see figure 3.

Since 1991, the Danish control has sub-sampled a certain percentage of the landings for reduction, number of landings sampled depends on year and fishery. These samples have been used to estimate the official species composition with the 9-square method. Looking at the species composition in these samples it is evident that the sprat fishery in the Baltic is very rarely a clean sprat fishery, see figure 4. Therefore, it makes sense to correct the species composition reported in the sale slips.

Since the methods for estimating the species composition before 2012 was not official, DTU Aqua is uncertain on how this was handled in the past and a comparison was made between the uploaded historic time series to ICES WGBFAS and the information we presently could find in the sale slips. From this exercise it was evident that especially the historic data on herring were much higher in the WGBFAS report than can be documented from the sale slips available, see figure 5, which indicate that some kind of species composition has been estimated in most year in the past.

#### Handling of spatial information

Presently DTU Aqua use the area and rectangle declared in the logbooks for assigning spatial information, in more recent years VMS and AIS is use when no information exists or a mismatch in declared area and square is found. For vessels without logbook, the area from the sale notes is used.

In the past landings from the Eastern Baltic was not always declared by subdivision, but only as 27.3.d, so it is needed to assign a subdivision to these based on other information available. In most cases it is possible to find a subdivision based on rectangle, see figure 6.

It is unknown how missing and too coarse spatial information was handled in the past, but the methodology for handling this has been developed and refined over the years.

### *Main outcomes of the analysis done*

#### By-catch of herring in the fisheries for reduction

Since it is unknown how the species compositing was handled in the past, it was decided to implement the 9-square method for the sprat fishery for all years in the Baltic with a standard method for imputations. The only samples we have available for this is the samples from the Danish control.

The results are compared with the present ICES WGBFAS timeseries in figure 7.

#### Handling of spatial information

Since it is unknown how missing and too coarse spatial information was handled in the past, it was decided to use the present methodology to assign subdivision to landings back to 1987.

As can be seen in figure 8, then distribution between subdivisions is quite different in the two time series, especially in the years before 1997, where all landings were assign to 27.3.d.25 in the old ICES WGBFAS time series.

*Advice to the benchmark*

Use the new time series, where the methodology is known and documented.

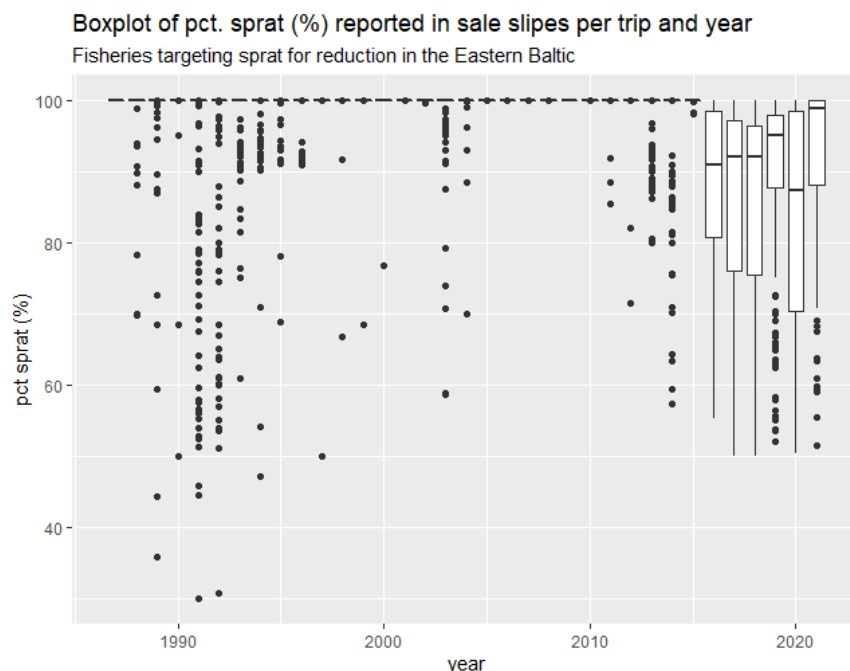


Figure 3 Boxplot of the percentage of sprat per trip in the sale slipes from the industrial sprat fishery in the Eastern Baltic.

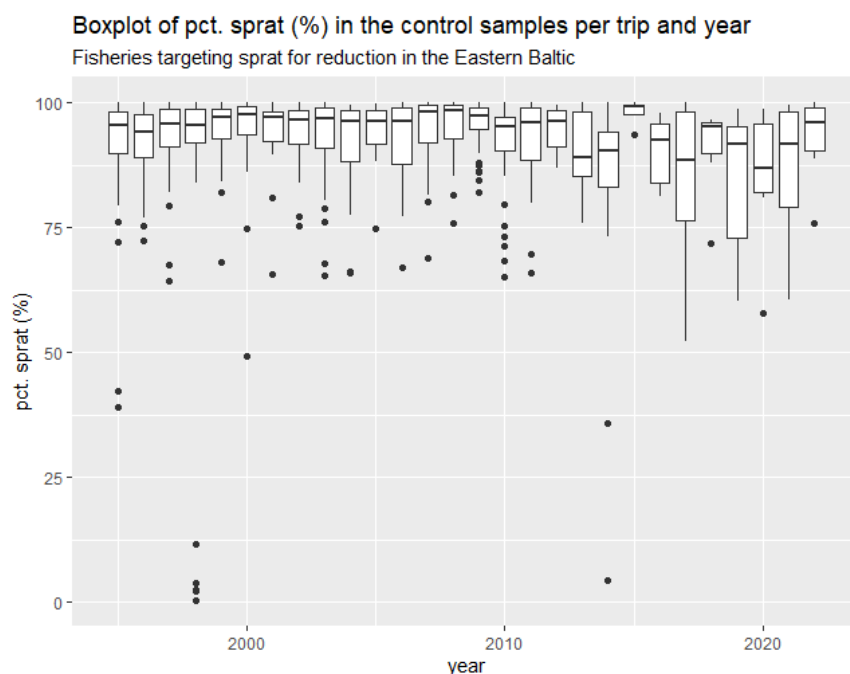


Figure 4 Boxplot of the percentage of sprat per trip in the Danish control samples from the industrial sprat fishery in the Eastern Baltic.

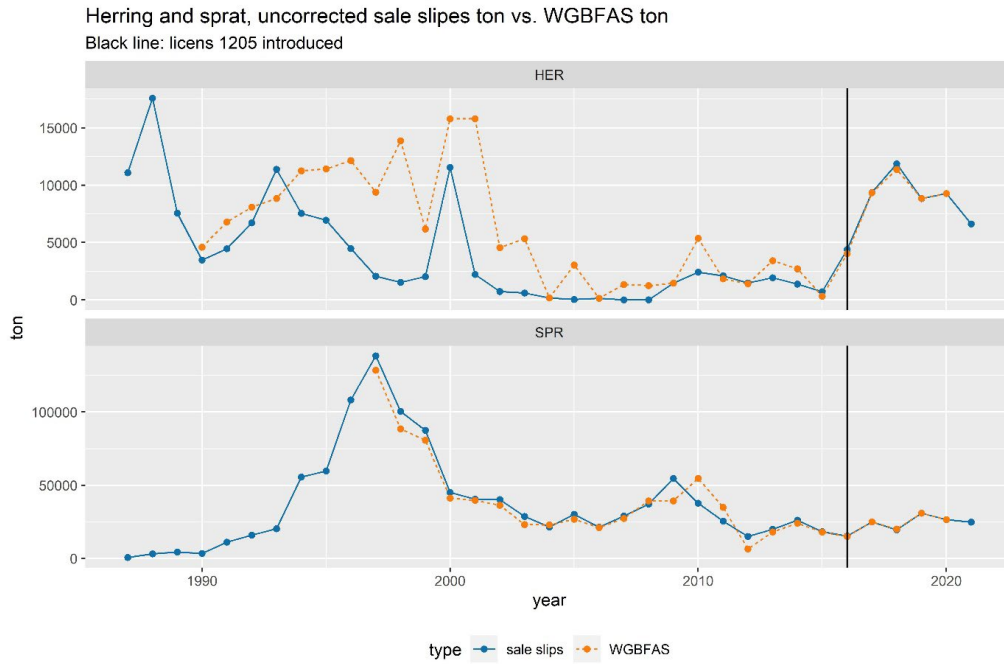


Figure 5 Comparison between the uploaded historic time series to ICES WGBFAS and the information we presently could find in the sale slipes. The figure only has data from the Eastern Baltic

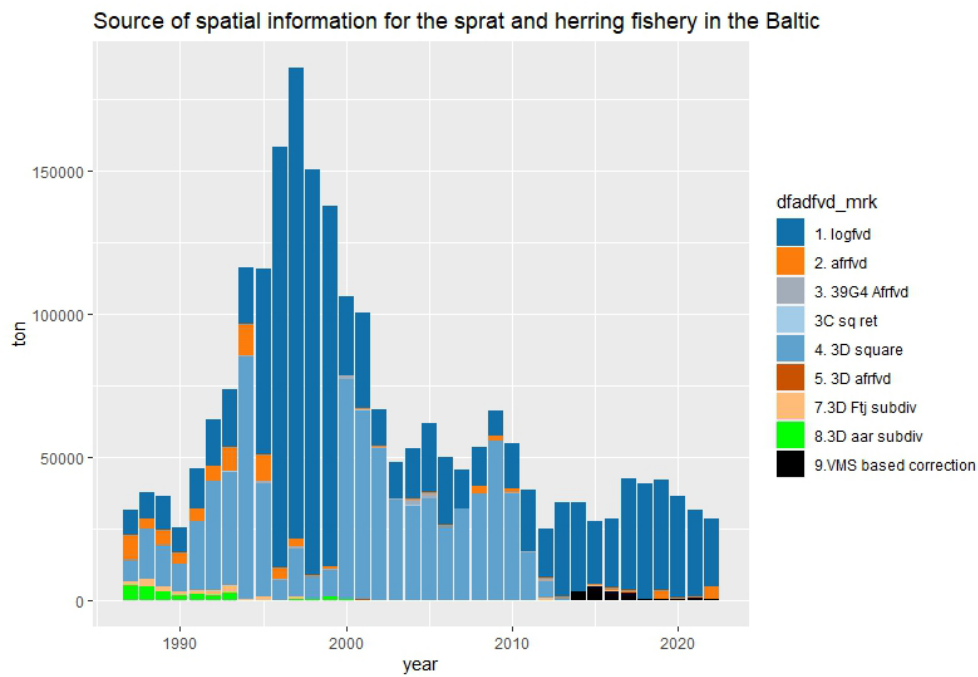


Figure 6 Source used for spatial information at the subdivision level with the present methodology

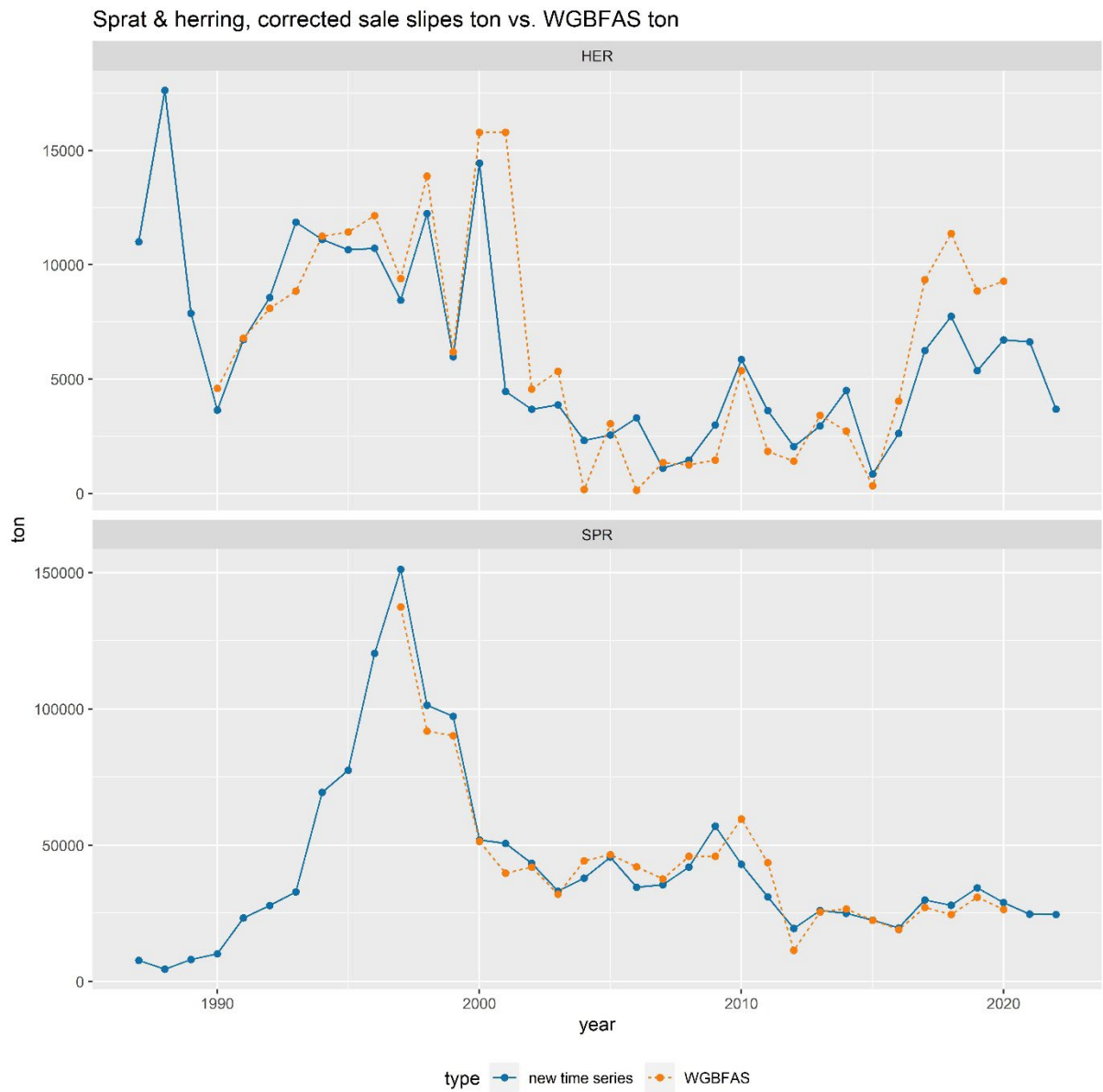


Figure 7 Comparison between the total amount in the historic ICES WGBFAS time series and the new time series. The figure shows sprat from the Baltic and herring from the Eastern Baltic

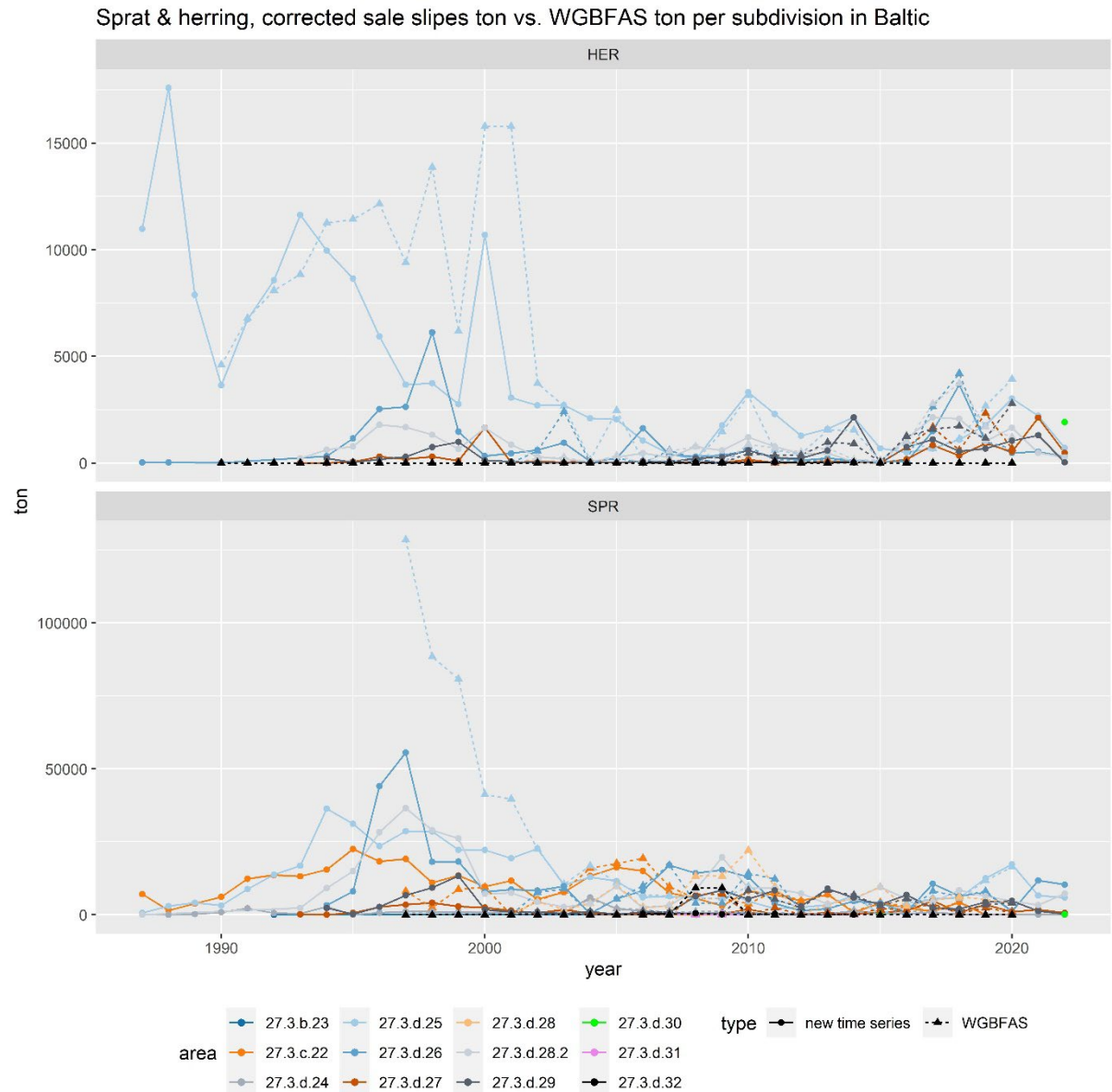


Figure 8 Comparison between the total amount in the historic ICES WGBFAS time series and the new time series per subdivision. The figure shows sprat from the Baltic and herring from the Eastern Baltic

Table 1 Overview of data sources and method used when submitting data from the Baltic to WGBFAS

Overview of data sources used when submitting data from the Baltic to WGBFAS			
MS	Landing category	Time period	Data source
Denmark	IND	2020-present	Sale slipes  (In 2020, Denmark introduced a new system for estimating the species composition in the landings for reduction. The Danish 1 <sup>st</sup> buyers of these landings now oblige to sub-sample every landing and use these to estimate the species composition in that landing. The estimated figures are reported in the sale slipes. The number of sub-samples depends on species, area and total amount landed e.g. in

		<p>2020 landings of sprat from the Baltic was sub-sampled in the following way;</p> <table border="1"> <thead> <tr> <th>Tons</th> <th>Number of sub-samples</th> </tr> </thead> <tbody> <tr> <td>0-25</td> <td>5</td> </tr> <tr> <td>25-200</td> <td>10</td> </tr> <tr> <td>&gt; 200</td> <td>15</td> </tr> </tbody> </table> <p>(The two biggest 1<sup>st</sup> buyers of landings for reduction use 3<sup>rd</sup> party companies to sample the landing))</p>	Tons	Number of sub-samples	0-25	5	25-200	10	> 200	15
Tons	Number of sub-samples									
0-25	5									
25-200	10									
> 200	15									
	2017- 2019	<p>Sale slip figures. No correction with control samples</p> <p>(All vessels had the 1205 license in the period)</p>								
	2016	<p>55% sale slips (1205 license) 45% Sale slips figures corrected with the 9-square method.</p> <p>(A new license, 1205, was introduced in the Eastern Baltic fishery for reduction. Vessel fishing with that license is oblige to report the species composition caught and the sale slip figures was not corrected with the 9-square method for these vessels. Sale slip figures from vessel fishing without was still corrected with the 9-square method)</p>								
	2012-2015	<p>Sale slip figures corrected with the 9-square method.</p> <p>(In 2012, The Danish Fisheries Agency took over calculation of by-catch with the 9-square method and it became a routine to use the method in the Baltic)</p>								
	1991-2012	<p>Sale slips figures has been corrected with the 9-square method some years, others not.</p> <p>(The so-called <b>9-square method</b> was introduced in 1991. The method use the Danish control samples to estimate the species composition in the fisheries for reduction. A species composition is calculated per square and month based on samples from the square and the 8 surrounding squares within month. The estimate is then applied to the figures from the sale slips per square and month (Logbooks are used to get information about ICES square).</p> <p>The method was routine for the North Sea, Kattegat and Skagerrak and used as official figures. For the Baltic it has not been an official routine, but some years it has been used, when submitting data to WGBFAS</p> <p>DTU Aqua was responsible for the calculations, but the results for North Sea, Kattegat and Skagerrak was used by the Danish Fisheries Agency)</p>								
	Before 1991	No clue								
HUC	All years	Sale slips								

## 4.1. Estonia

### 4.1.1. Fishery

The Estonian fishery of herring and sprat is mainly for human consumption, however this trend has been decreasing with recent years due to the development of fishmeal and oil factory located in Paldiski, Estonia. Estonia TAC share from Central Baltic herring stock is around 8-9%, 46% from Gulf of Riga herring TAC and 9-10% from sprat TAC.

### 4.1.2. Approach taken to analyze if there are errors in the time series of catch data due to inadequate reporting of species and/or other reasons

Since 2014 Estonian control agency has conducted regular controls to determine the accuracy of the species composition and weight of landed fish. Legally,  $\pm 10\%$  difference in landed weight per species is allowed. The control agency has the leniency to determine based on visual inspection if biological samples need to be taken to determine the species composition, and total landed weight per species. This means that biological samples might not be taken during every inspection event.

When difference is detected between logbook and inspected data then the control agency suggests for the skipper to change/update the logbook data to correspond to what has been determined by the inspection. If this data is updated/changed then corrected data will be used when catches/landings are reported to ICES. However, it is not possible to track in which cases the data was corrected or not.

### 4.1.3. Main outcomes of the analysis done

The control agency has been collecting samples since 2014. From 2014-2021 total of 1466 fishing trips were inspected (Table 1). From all inspected trips (N=1466) total of 819 trips were sampled for species composition (Table 1). This makes 55.83% of all inspected trips. The number of trips sampled has increased from year to year.

Reported and inspected species composition does not seem to differ a lot between the reported species composition and inspected species compositions (Figure 1, 2). For years 2014-2021 the overall impression is that herring might be slightly overreported compared to sprat. The direction of false reports in general indicates an over reporting of herring in the catch composition as over 45% of inspected trips over reported herring (Table 2).

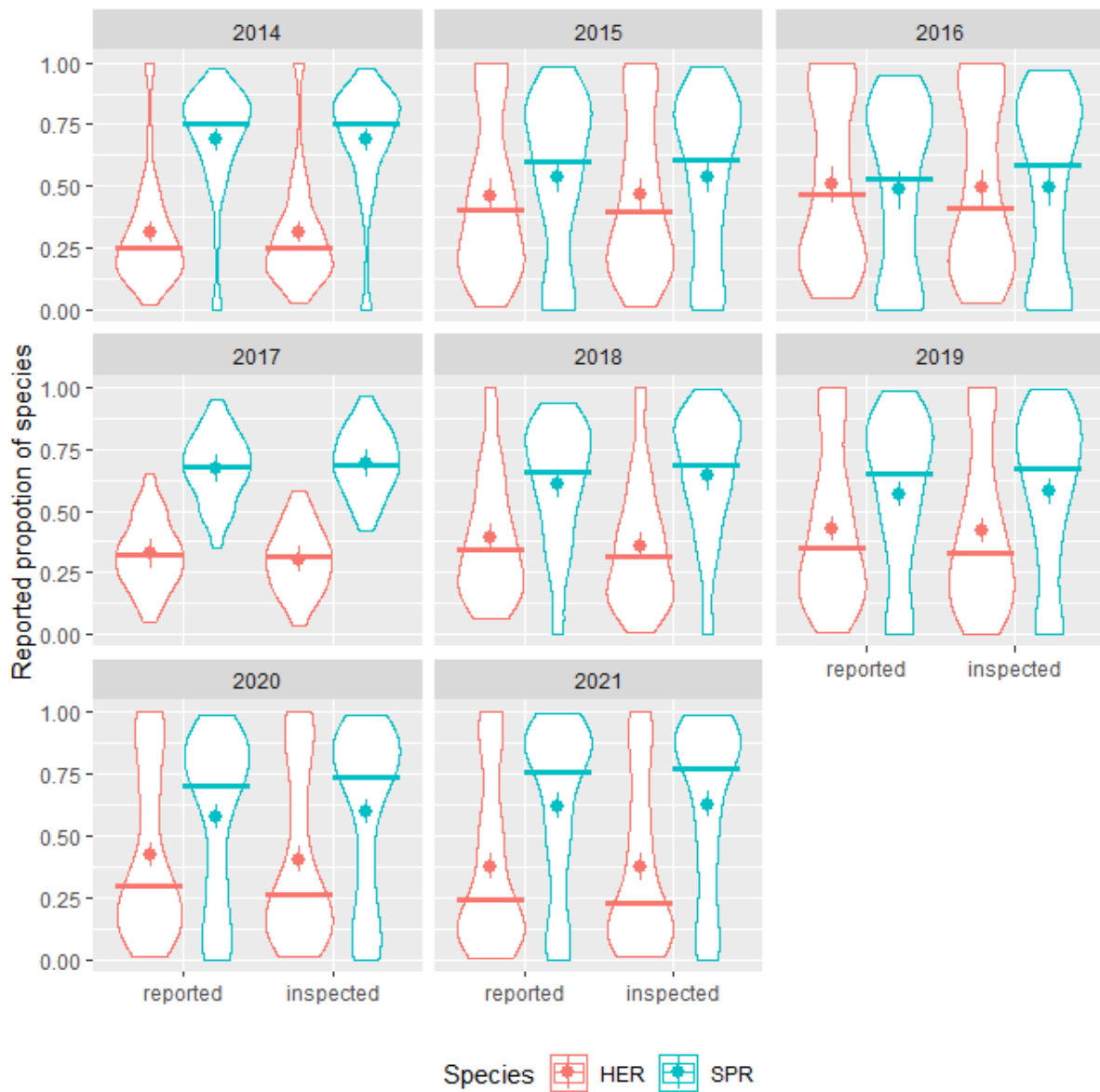
**Table 1.** Total number of Estonian trawl fleet fishing trips, and number of trips that were inspected in years 2014-2021.

Year	Inspected	Landed	Sampled	% Inspected	% Sampled
2014	234	2732	95	8.57	40.60
2015	240	3229	93	7.43	38.75
2016	119	2656	78	4.48	65.55
2017	128	3107	23	4.12	17.97
2018	169	2966	84	5.70	49.70
2019	199	2893	143	6.88	71.86
2020	209	2823	164	7.40	78.47
2021	168	2066	139	8.13	82.74

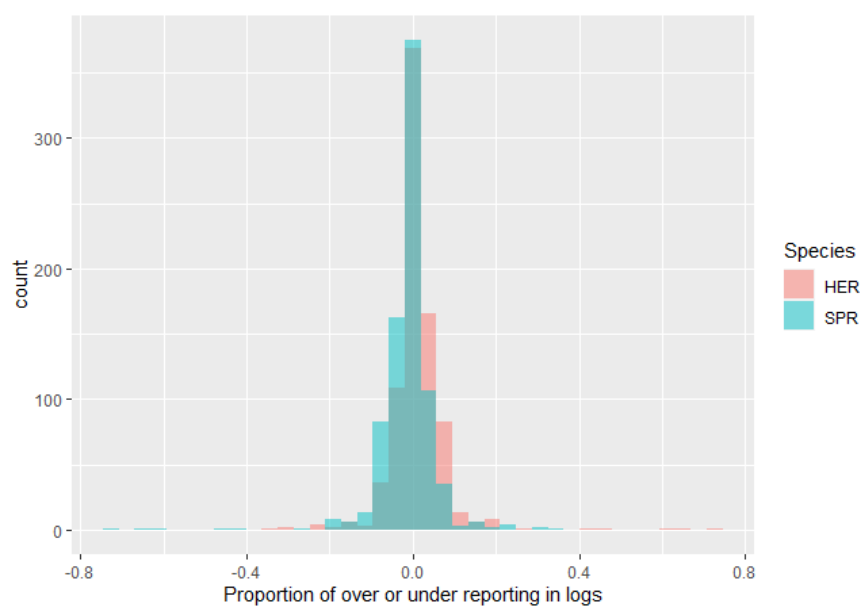


**Table 2.** Proportion of over or under reporting of herring and sprat in catch composition based on sampled inspected trips for years 2014-2021. N=number of sampled trips; median% - median proportion of over or under reporting; mean% - mean proportion of over or under reporting; sd% - standard deviation of the mean proportion of over or under reporting; overreportedN – number of trips where over reporting was detected; overreported% - % of overreported trips.

Year	Species	N	median%	mean%	sd%	overreportedN	overreported%
2014	HER	94	0.0	0.2	5.7	45	47.9
2014	SPR	94	0.0	-0.2	5.7	36	38.3
2015	HER	90	0.0	-0.5	5.0	42	46.7
2015	SPR	90	0.0	0.5	5.0	39	43.3
2016	HER	78	0.1	1.2	4.6	44	56.4
2016	SPR	78	-0.1	-1.3	4.6	23	29.5
2017	HER	23	2.2	2.3	3.4	18	78.3
2017	SPR	23	-2.2	-2.3	3.4	5	21.7
2018	HER	84	1.2	3.6	9.8	55	65.5
2018	SPR	84	-1.2	-3.6	9.8	24	28.6
2019	HER	143	0.7	1.1	8.2	84	58.7
2019	SPR	143	-0.6	-1.2	8.2	49	34.3
2020	HER	163	1.1	2.0	8.1	105	64.4
2020	SPR	163	-1.1	-2.0	8.1	57	35.0
2021	HER	139	0.5	0.3	5.0	76	54.7
2021	SPR	139	-0.5	-0.3	5.0	59	42.4



**Figure 1.** Species proportion distribution comparison between logbooks data and and results from the sampled inspection trips.



**Figure 2.** Proportion of over or under reporting of herring and sprat in catch composition. Figure is symmetrical as over reporting of one species comes with an under reporting of the other species. Dark blue indicates overlap. Bars are set at 5% intervals.

#### 4.1.4. Advice to the benchmark

- *The national data to not be updated* as there are no indications after the analysis that data can be improved. Country will not provide new time series.
- It is important to note that based on the current available data and analysis conducted we are not able to improve the current data. This however does not mean that the data should not be improved. Currently available data is not enough to conduct such improvements.

Choose from three options

*The national data to be updated* as there are clear indications that data can be improved. Country will provide new time series

*The national data to not be updated* as there are no indications after the analysis that data can be improved. Country will not provide new time series

*The national data might be updated* as there are indications after the analysis that data can be improved. Country will provide one or more new time series that the stock assessors can explore.

## 5.1. Finland

### 5.1.1. Fishery

Finnish fishery targeting herring and sprat is conducted mostly with pelagic trawls, but also to a minor extent by coastal trapnets (FPN, FYK) during spawning time on the emphasis to springtime. The stocks

concerned are Central Baltic Herring stock (SD 25-29, 32), Gulf of Bothnia Herring stock, i.e. Bothnian Sea Herring (SD 30) and Bothnian Bay Herring (SD 31) – the latter two have always belonged to the same management unit and to the same assessment unit since 2017, and the Baltic Sprat stock (SD’s 22-32). Biological data are collected mostly from sampling of commercial trawl fisheries (OTM\_SPF and PTM\_SPF), but also from trapnets.

The Finnish quotas for the stocks are: GoB, 81,99%, CBH 21,93 % and Baltic Sprat 5,16%. The catches are mostly used for industrial purposes (Figures x.1. And x.2.).

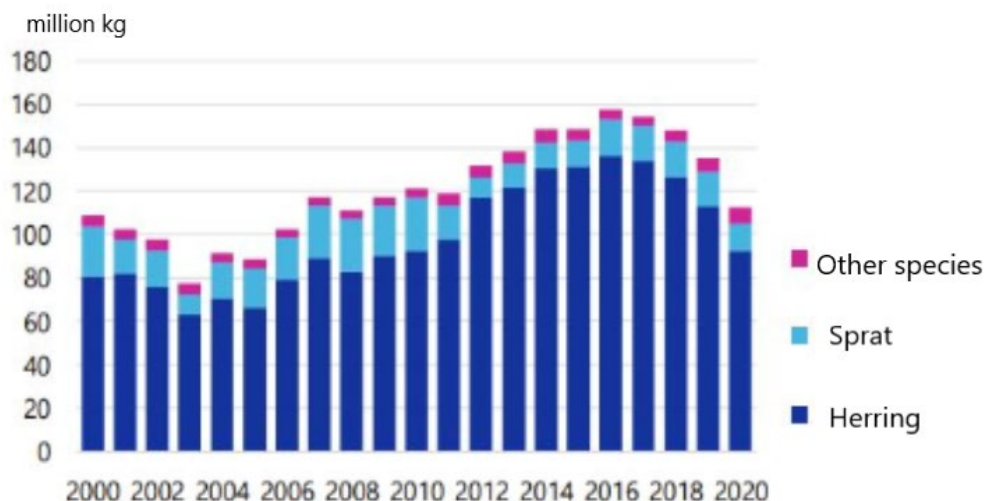


Figure x.1. Catches in Finnish Fisheries in 2000-2020

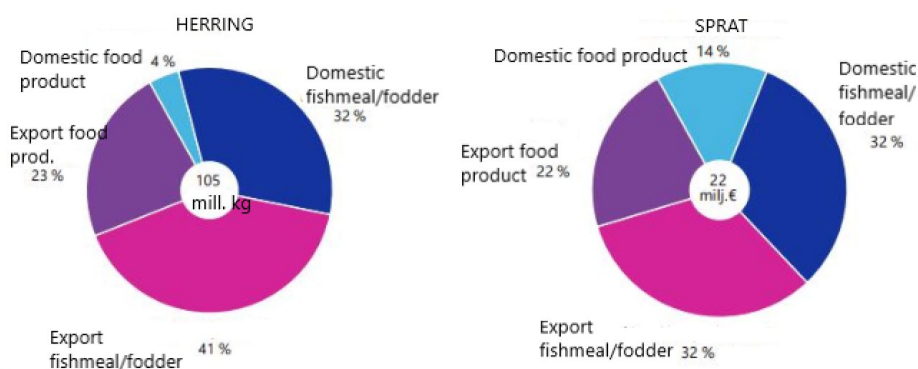


Figure x.2. Shares for human consumption, industrial purposes, domestic use and export in Finnish official herring and sprat catches, in millions of kilograms (left) and millions of Euros (right). (Source: Natural Resources Institute Finland)

5.1.2. Approach taken to analyze if there are errors in the time series of catch data due to inadequate reporting of species and/or other reasons

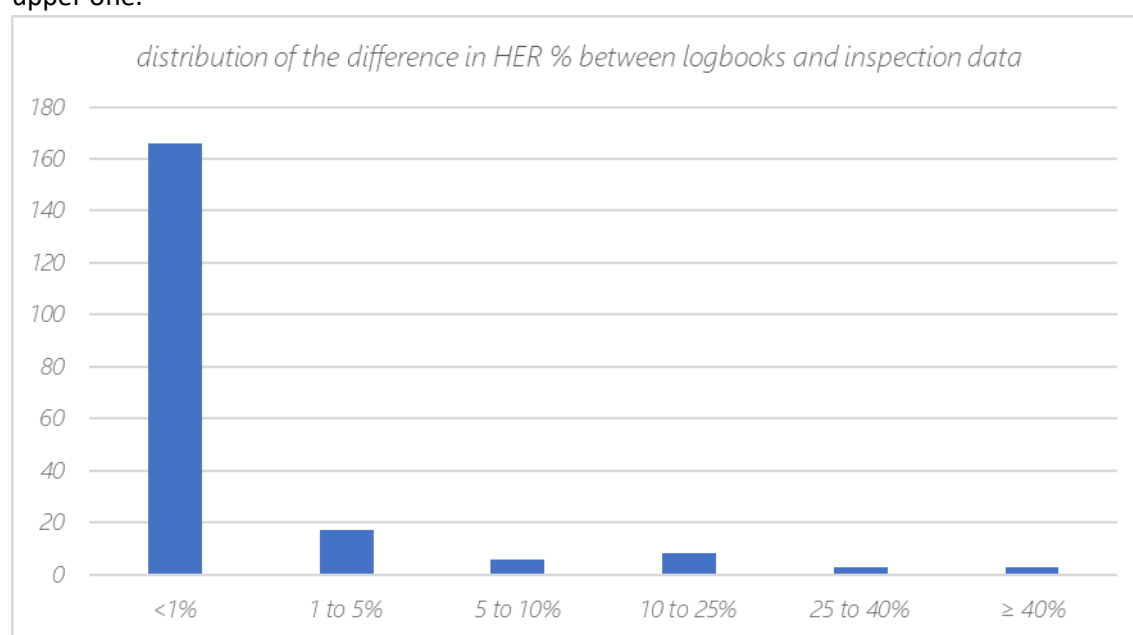
To assess the existence and potential magnitude of misreporting of sprat for herring and vice-versa in the catches of the Finnish trawlers, we were provided by the fisheries inspection services of ELY-Keskus with data corresponding to a set of 203 catch events distributed in the years 2007 to 2022, in ICES SD 28 (1 catch event), 29 (178 catch events) and 32 (24 catch events). The data provided by ELY-Keskus are anonymous (i.e. no vessel ID) and contain the following variables: ICES rectangle, catch date, inspected herring weight and sprat weight. For most cases the combination of catch date and statistical rectangle in these inspection data was pointing to a single event in the logbooks. There were also occurrences of several catch events on the same day and rectangle both in the inspection data and in

the logbooks, most of which were also easily relatable under the assumption that exact or close (i.e. <2% difference) matches in terms of species proportions indicated a same catch event.

No other species than herring and sprat was present in these catches. The value used for the analysis was the difference between the percentage of herring in the catch weight reported in the logbooks and the corresponding percentage in the inspection data.

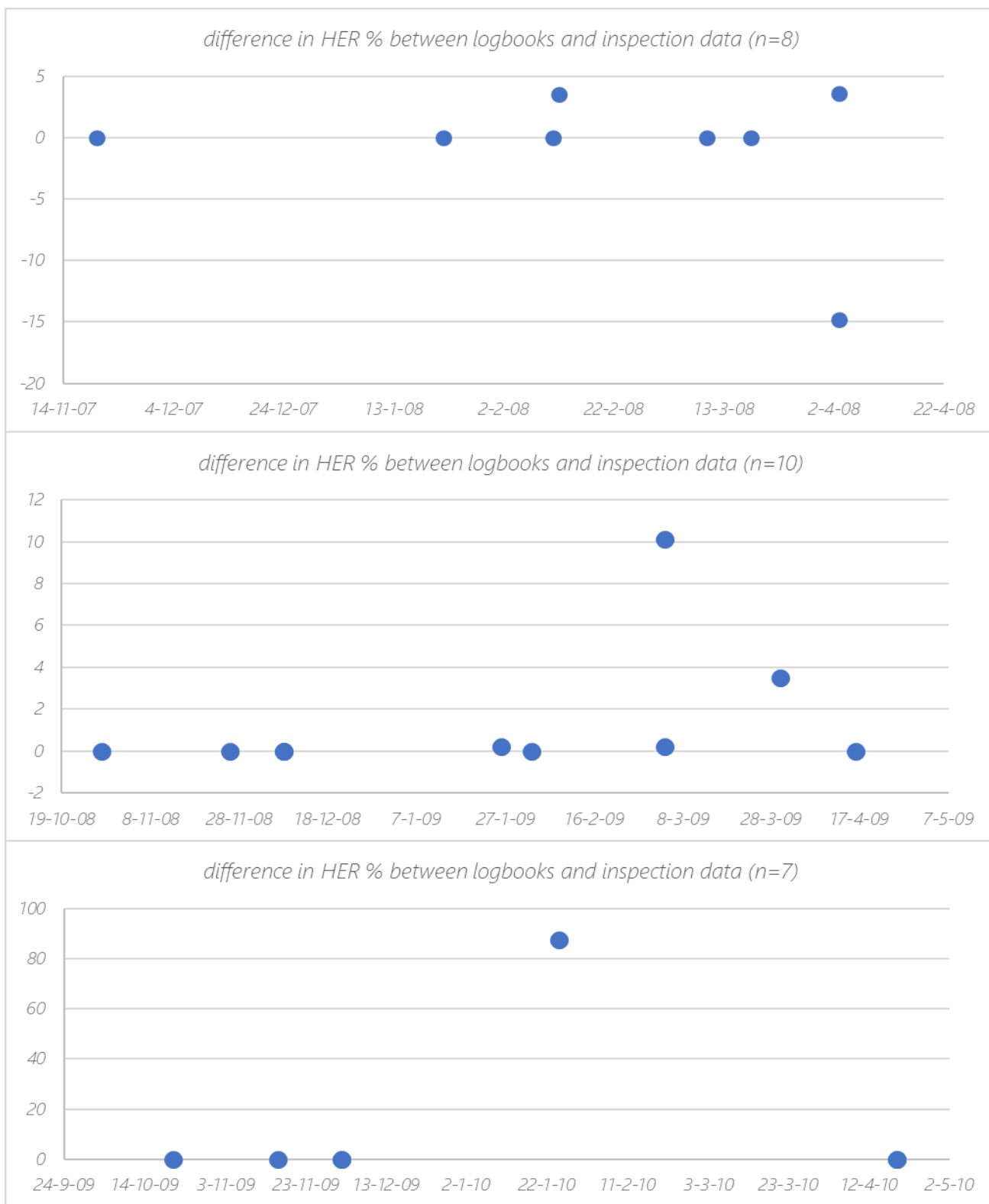
### 5.1.3. Main outcomes of the analysis done

Among the 203 catch events inspected, 166 displayed an exact match (<1% difference in herring percentage between logbook and inspection data), 17 displayed a close match (difference ranging from 1 to 5%) and 20 displayed a difference superior to 5% (Figure x.3). Within this last case, 6 events displayed a difference between 5 and 10%, 8 a difference from 10 to 25%, 3 a difference from 25 to 40% and 3 a difference over 40%. All ranges expressed include the lower boundary and exclude the upper one.



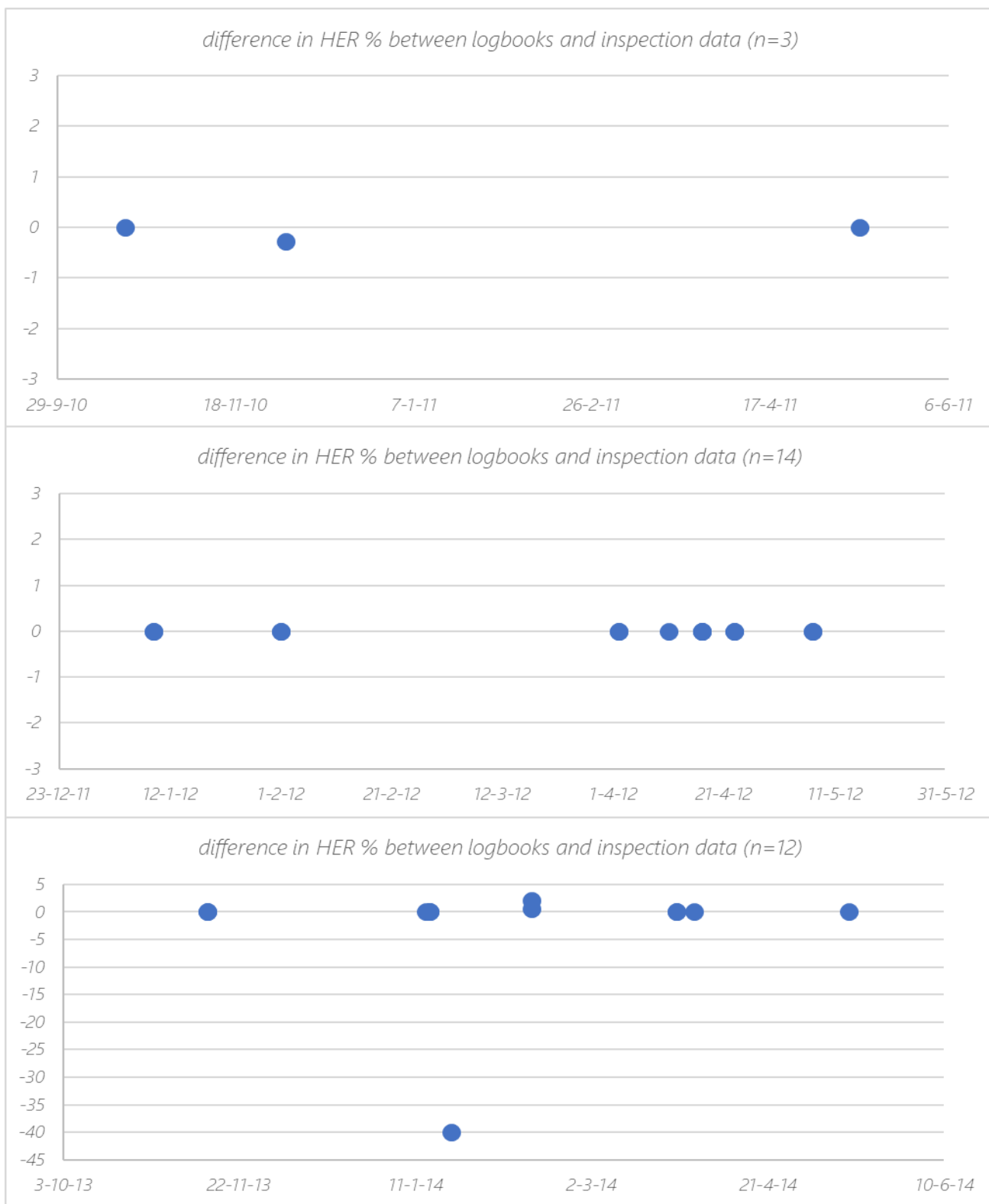
**Figure x.3:** difference in herring percentage between logbook and inspection data, grouped by percentage range

For the 2007-2008 winter, most differences are 0, two cases show a positive difference by less than 5% and one case shows a negative difference by 15%. For the 2008-2009 winter we observe one positive difference by 10%, and in the following winter an 87.5% negative difference, for which the logbook contains no sprat data. (Figure x.4)



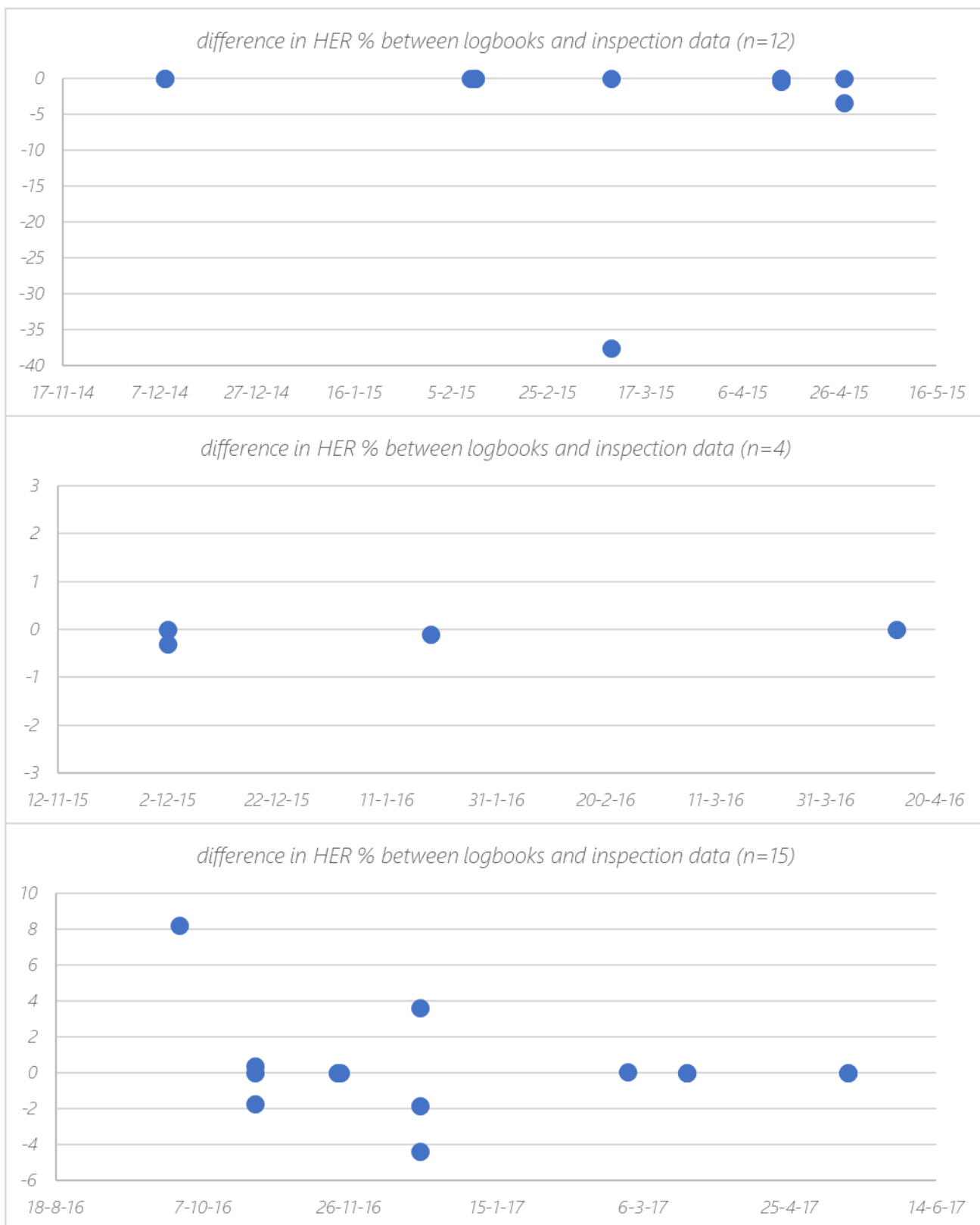
**Figure x.4:** difference in herring percentage between logbook and inspection data, split by date for the seasons 2007-2008, 2008-2009 and 2009-2010

For the winters 2010-2011, 2011-2012 and 2013-2014, the most meaningful discrepancy is a negative difference of 40% in January 2014, for which the inspection data shows no sprat (Figure x.5).



**Figure x.5:** difference in herring percentage between logbook and inspection data, split by date for the seasons 2010-2011, 2011-2012 and 2013-2014

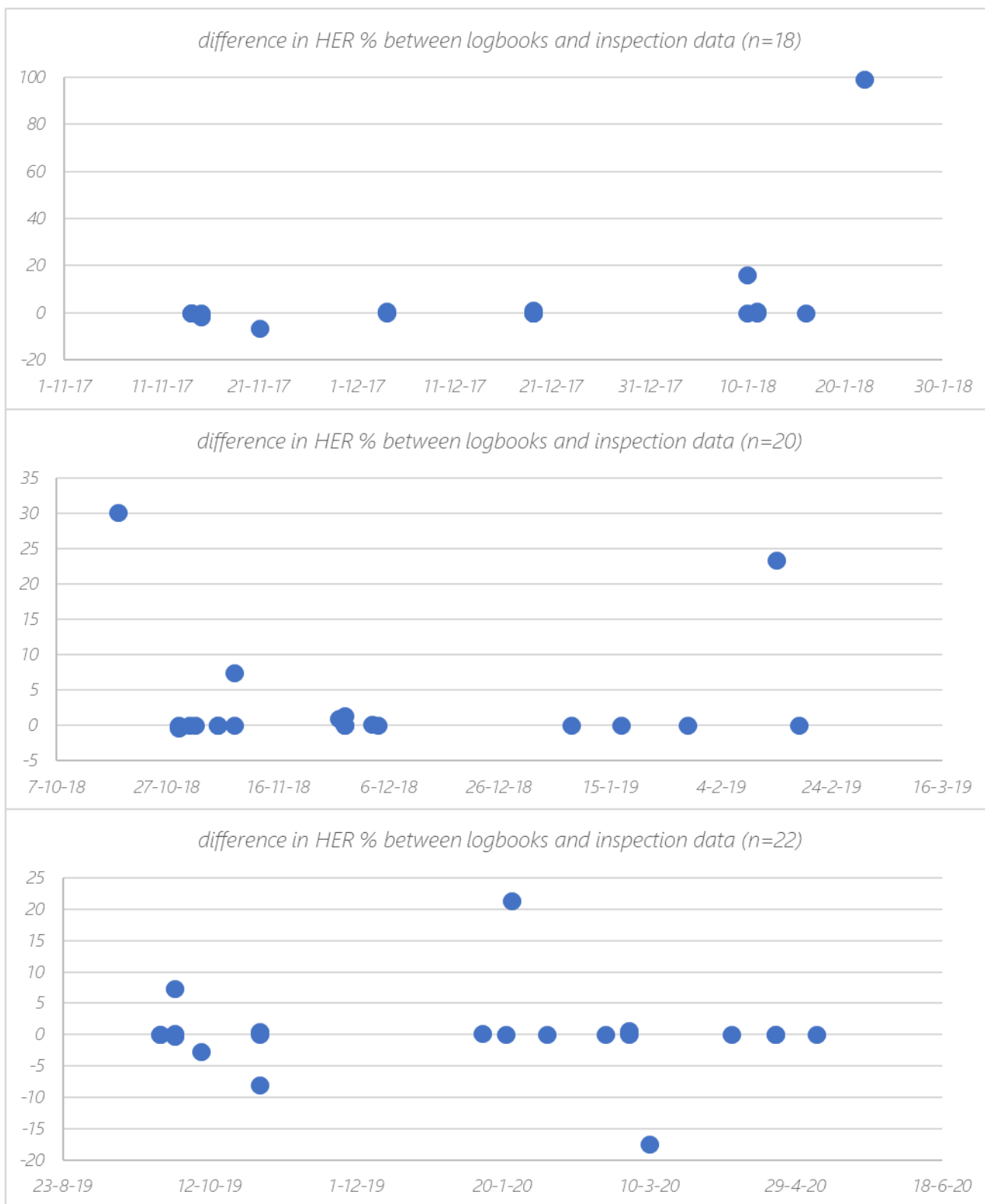
For the seasons 2014-2015, 2015-2016 and 2016-2017, the most meaningful discrepancy is also a negative difference of 37.6%, for which the inspection data shows much more herring than the logbook (Figure x.6).



**Figure x.6:** difference in herring percentage between logbook and inspection data, split by date for the seasons 2014-2015, 2015-2016 and 2016-2017

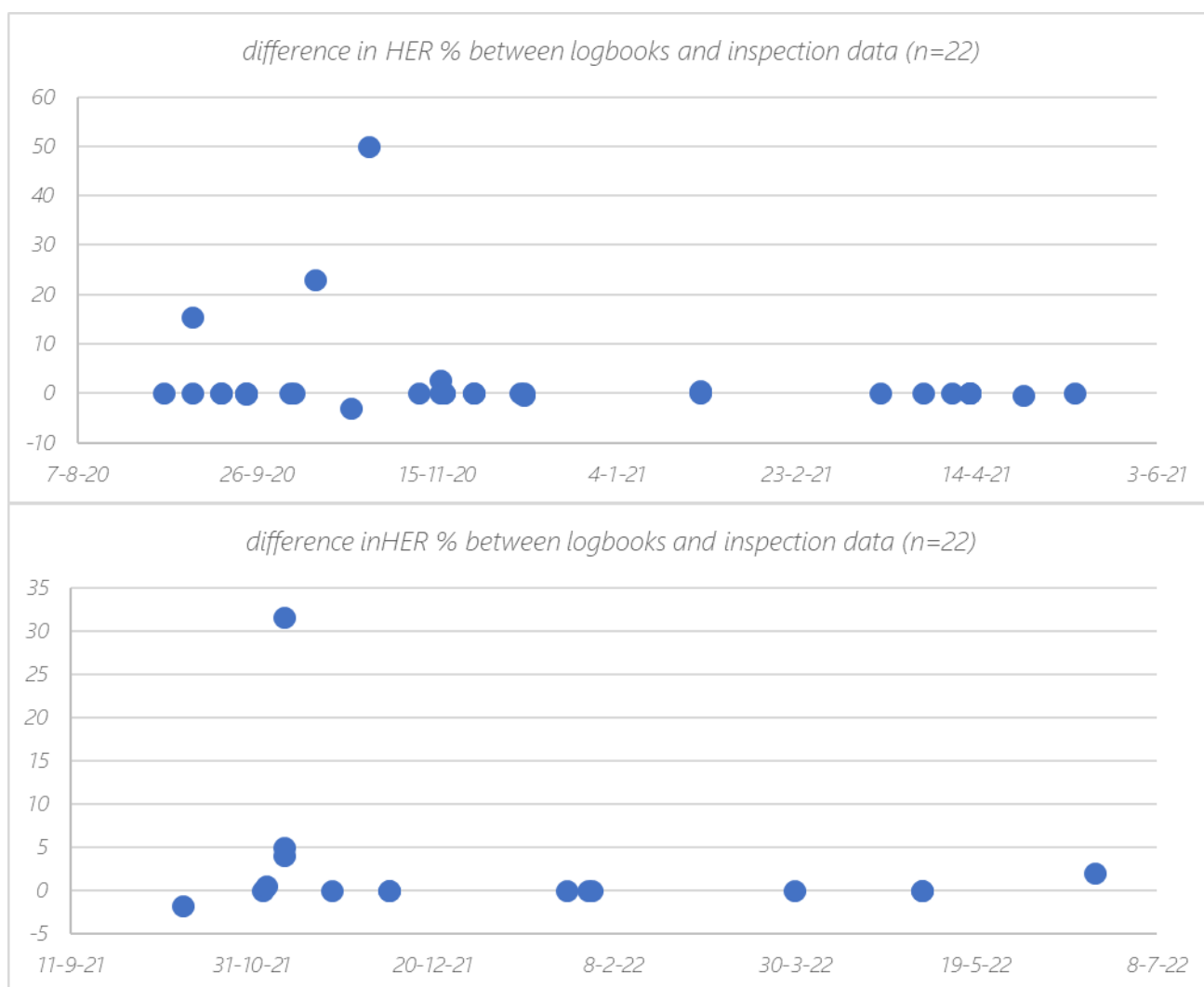
For the seasons 2017-2018, 2018-2019 and 2019-2020 we observe in early 2018 the most extreme positive difference, for which no herring appear in the inspection data. A 30% positive of herring in October 2018 corresponds to a higher proportion of sprat in inspection data. Reversely the negative difference in March 2020 corresponds to a higher proportion of sprat in the logbook (Figure x.7).





**Figure x.7:** difference in herring percentage between logbook and inspection data, split by date for the seasons 2017-2018, 2018-2019 and 2019-2020

The 50% positive difference in October 2020 is due to a near absence of sprat in logbook data, and the 31.5% positive difference in November 2021 is due to a higher proportion of herring in the logbook data (Figure x.8).



**Figure x.8:** difference in herring percentage between logbook and inspection data, split by date for the seasons 2017-2018, 2018-2019 and 2019-2020

#### 5.1.4. Discussion, pending issues and advice to the benchmark

The differences observed do not show any temporal pattern, neither in terms of seasonality nor in terms of year. We do not identify a pattern of over-declaration of one species versus the other to adjust *a posteriori* the catch to the available quota, i.e. these data do not suggest any intended misreporting. However, the most extreme discrepancies correspond to rare cases where one species was absent or near absent in either the logbook data or the inspection data, which might also be due to human error i.e. unintended skipping a number when reporting. Additionally, we noticed several cases in which the respective *proportions* of herring and sprat match between the logbook data and inspection data whereas the *amounts* do not, and other cases in which the amounts (and therefore proportions) match. To clarify these issues, we will need further exchanges with the personnel of ELY-Keskus in charge of the inspections. This was not possible for this deadline due to the leave *sine die* of the contact person in ELY who provided the inspection data, and the unavailability of other personnel potentially contributing to this task (no reply to the messages sent so far). Although we hope to get feedback before the WGBFAS meeting, we cannot guarantee that we will.

- *Should data be updated or not*

Considering the information we have at this stage, the data should be kept as such. Pending further clarifications to be obtained from ELY-Keskus, data may require some minor update.

- *Are there particular years / periods in the time series that are more or less trustworthy than others*

The differences observed do not show any temporal pattern

## 6.1. Germany

### 6.1.1. Fishery

The German fishery for Baltic sprat and Central Baltic herring is mainly used for industrial purposes. The quota share of Germany of the CBH and Baltic sprat EU quota is <1% and <5%, respectively. Two trawlers (>40 m LOA) take most of these fishing opportunities. They fish from SD 25 to SD 29 and land their catches usually in Denmark. In addition, a small number of mid-sized trawlers fish in SD 25 on both stocks. Until 2021 sprat was bycaught in the pelagic trawl fishery targeting Western Baltic spring spawning herring off the island of Rügen (SD 24) (the trawl fishery is closed since then). And a few trawlers caught minor amounts of sprat in SD 22.

### 6.1.2. Approach taken to analyze if there are errors in the time series of catch data due to inadequate reporting of species and/or other reasons

Two approaches were taken to check the official landings data of the German fishing fleet for species misreporting:

1. Check official landings declarations and logbook entries and compare them with Danish control data, covering the last five years.
2. Compare species compositions from co-sampling of the commercial trawlers analysed by Thünen-OF and compare them with the species composition reported by the vessels (landings declaration and logbooks).
- 3.

### 6.1.3. Main outcomes of the analysis done

No indication of misreporting was found in neither approach and for neither species or stocks. In most cases, control data, co-samples from the fishery and the official records did match with >95% similarity.

In some cases, the compared values differed, but could be explained after consulting the logbook entries and feedback from the fisheries control authority in Denmark:

1. **Differences between Danish control data and official landings records:** A total of 48 trips, covering the years 2019 to 2021, was compared (Fig.1-3). Of the 48 trips, 4 trips showed a <95% similarity in species composition (Fig. 1 to 3, blue squares). Two trips (one in 2019 and one in 2021) had a larger herring ratio in the sales notes than registered by the control data. In both cases, the trip was done with a partner vessel, conducting a pair trawl (PTM) and landings were assigned between vessels based on quota availability and agreements between the vessels. Two other trips in 2020 (also PTM) swapped landings in the respective landings' declarations, possibly due to similar reasons (quota restrictions or internal agreements between the vessels).

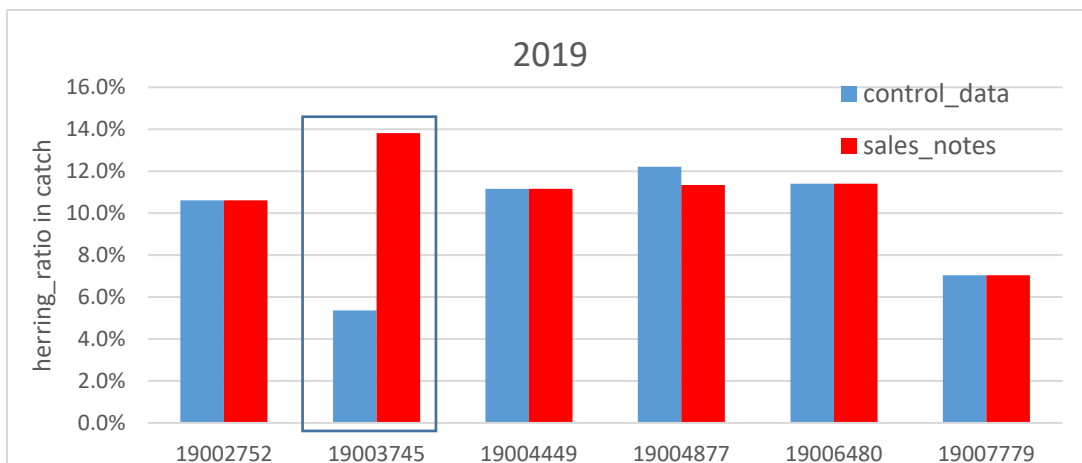


Fig. 1: Ratio of herring in Danish control data (blue) vs. official sales notes (red) in 2019. Trip with a low similarity is marked.

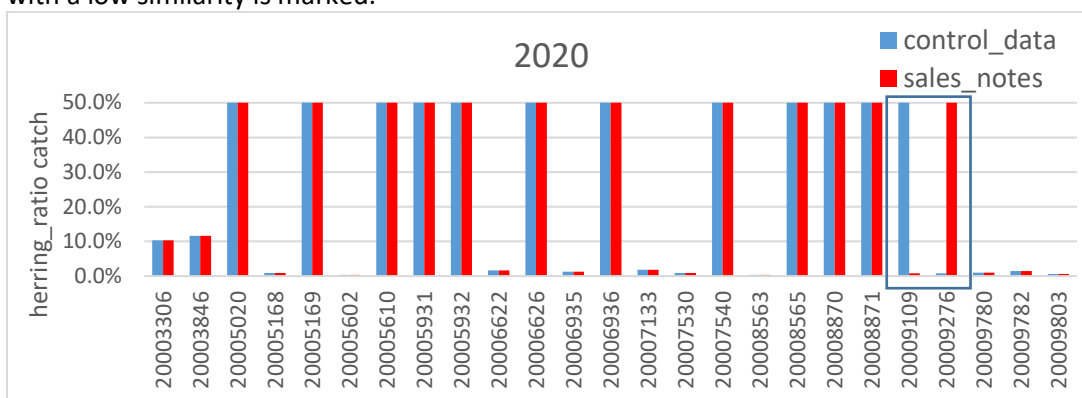


Fig. 2: Ratio of herring in Danish control data (blue) vs. official sales notes (red) in 2020. The Y-axis was cut at 50% to better display the trips with a low herring ratio. Trips with low similarity are marked.

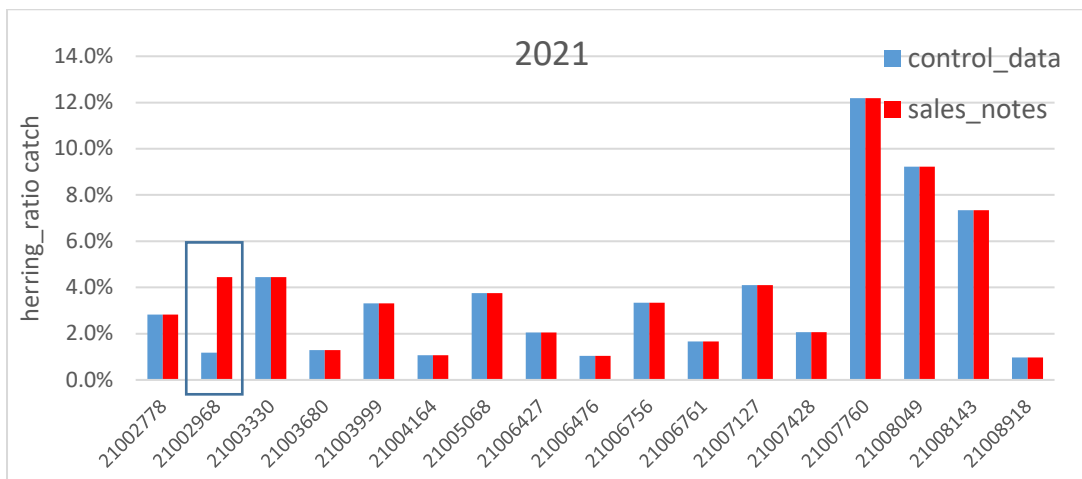


Fig. 3: Ratio of herring in Danish control data (blue) vs. official sales notes (red) in 2021. Trip with a low similarity is marked.

- Differences between co-samples from the fishery and official landings records:** A total of 32 trips, covering the years 2019 to 2021, was compared (Fig.4-6). Of the 32 trips, 11 trips showed a <95% similarity in species composition (Fig. 4 to 6, blue squares). Most differences

could be assigned to the design of the co-sampling where the fishery collects an unsorted catch sample in a 5 kg bucket from each trip. Each co-sample from the fishery is analysed in detail by Thünen-OF using fisheries biology standards. This amount is sufficient to provide useful biological data for sprat (and CBH when present) but cannot provide unbiased results on species mixing from different hauls taken during a trip.

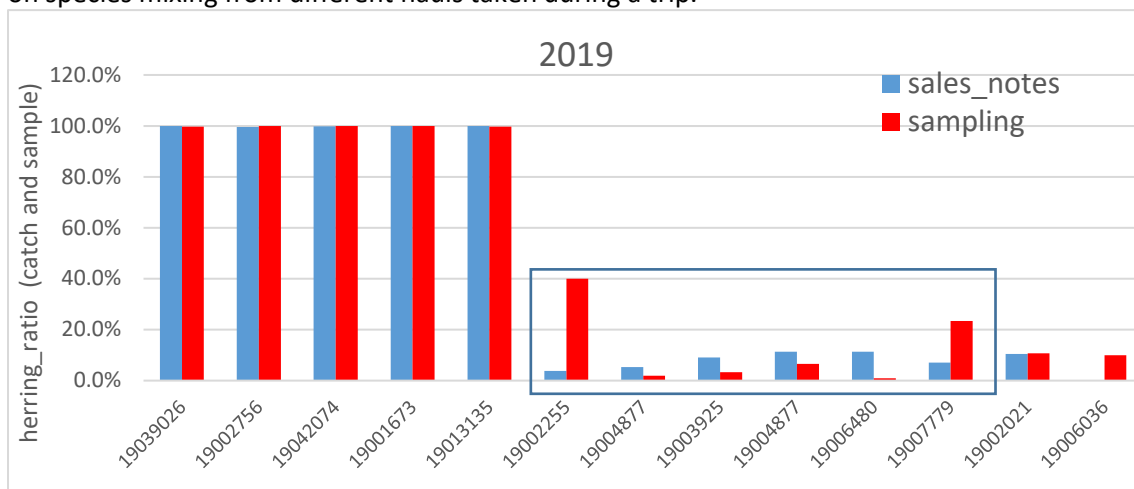


Fig. 4: Ratio of herring in official landings data (blue) vs. co-samples from the fishery (red) in 2019. Trips with a low similarity are marked.

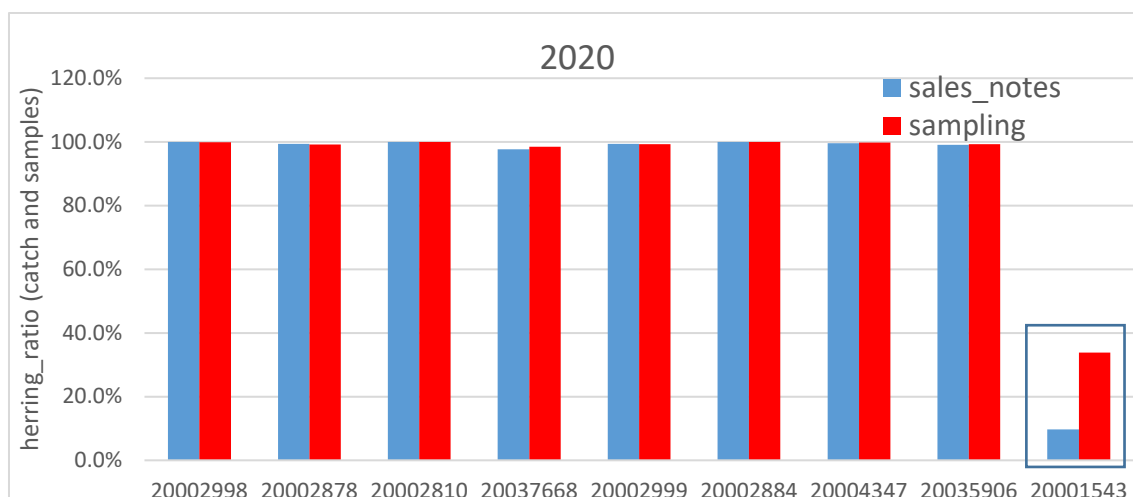


Fig. 5: Ratio of herring in official landings data (blue) vs. co-samples from the fishery (red) in 2020. Trip with a low similarity is marked.

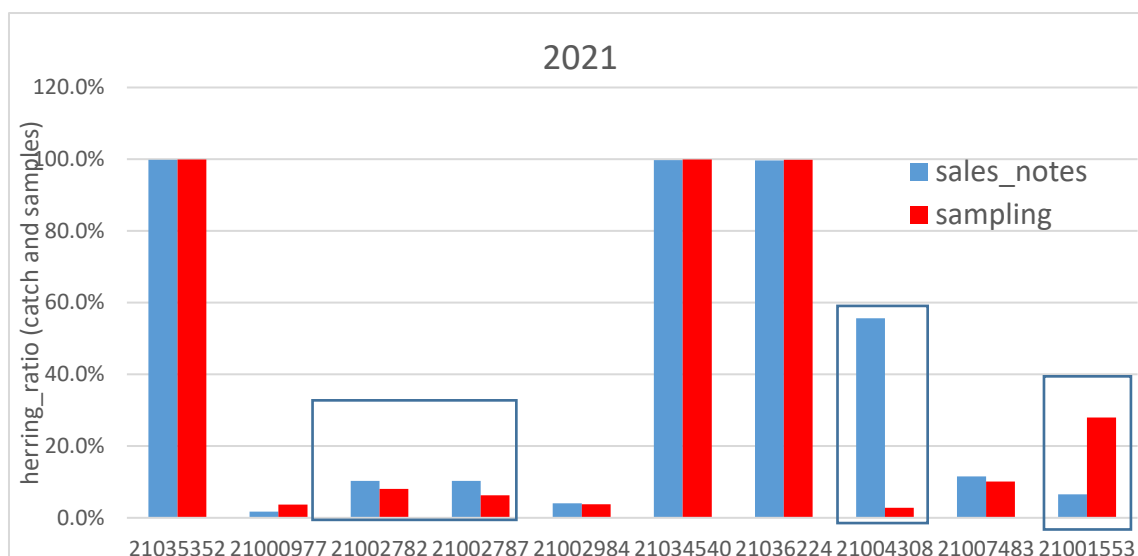


Fig. 6: Ratio of herring in official landings data (blue) vs. co-samples from the fishery (red) in 2021. Trips with a low similarity are marked.

6.1.4. Advice to the benchmark

**No evidence of misreporting has been found for the two major German vessels targeting Central Baltic herring and Baltic sprat. The benchmark group can therefore use the submitted data without adjustments or changes.**

7.1. Latvia

7.1.1. Fishery

The Latvian fishery for sprat and herring in the Baltic is mainly used for human consumption. Latvia has a relatively small part (<4 %) of the Central Baltic herring quota and for Baltic sprat, Latvia has approx. 12 % of the quota. The relative Latvian quota share has changed over time and in 2021 around 87 % of the total sprat and herring quota in the Central Baltic was related to sprat (Figure 7.1.1).

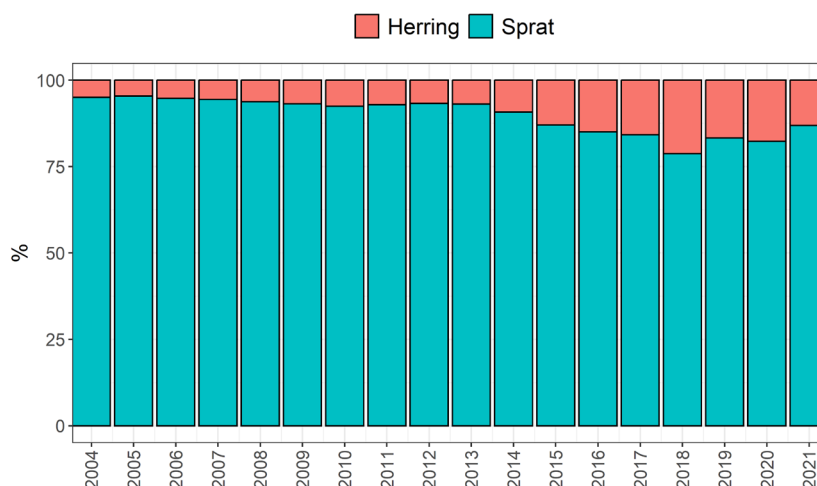


Figure 7.1.1. The relative share of sprat and herring quota in the Latvian Central Baltic fishery (excluding SD 28.1 (Gulf of Riga)).

Latvian pelagic fishery is mainly conducted with pelagic trawls targeting sprat or herring (métiers OTM\_SPF\_16\_31\_0\_0 and OTM\_32\_104\_0\_0). Most landings are taken in ICES SD 28.1 and 28.2 showing differences between both regions – in SD 28.1 (Gulf of Riga) main target stock is the Gulf of Riga herring, whereas in SD 28.2 (Central Baltic) main stock is sprat (Figure 7.1.2). Herring in the Gulf of Riga has a separate management unit. Herring fishery in the Gulf of Riga is performed, using both trawls and trapnets. Herring catches in the Gulf of Riga include the local Gulf of Riga herring and the Central Baltic herring, entering the Gulf of Riga for spawning. Discrimination between the two stocks is based on the different otolith structure due to different feeding conditions and growth of herring in the Gulf of Riga and the Baltic Proper. The Latvian fleet also takes Gulf of Riga herring outside the Gulf of Riga in Subdivision 28.2. In 2021 these catches were 775 t.

In 2020 Latvian fleet consisted of 49 registered offshore vessels (12–40 m) and 603 coastal vessels (< 12 m).

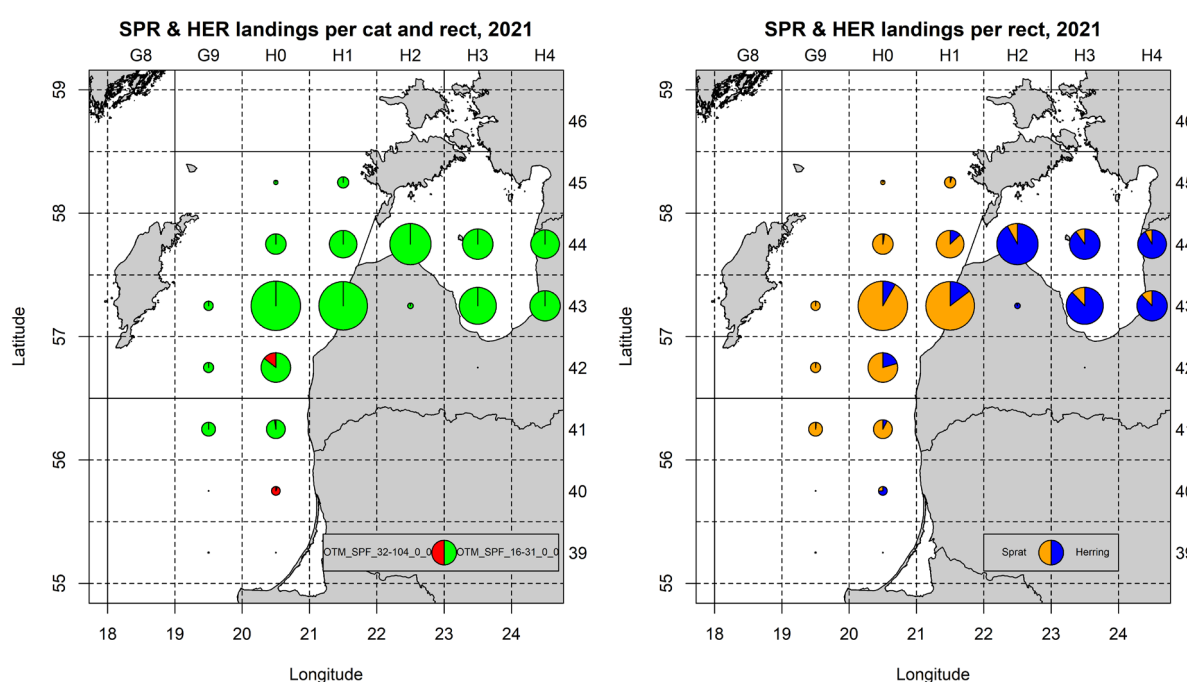


Figure 7.1.2. Sprat and herring landings in 2021 from Latvian vessels by métier and species.

#### *7.1.2. Approach taken to analyze if there are errors in the time series of catch data due to inadequate reporting of species and/or other reasons*

Latvian logbook data were compared with Danish control samples provided by the ISSG Small Pelagic group. In total 69 matching trips were identified for the analysis and covered the period from 1998-2019. The majority of samples were from SD 25 near Bornholm (Figure 7.1.3).

Another source of information was Latvian fishery sales notes from Denmark ports which were compared to Latvian logbook data.

Additional information was also asked from the Latvian control agency. Latvian control agency is conducting regular controls to determine the accuracy of the species composition and weight of landed fish.

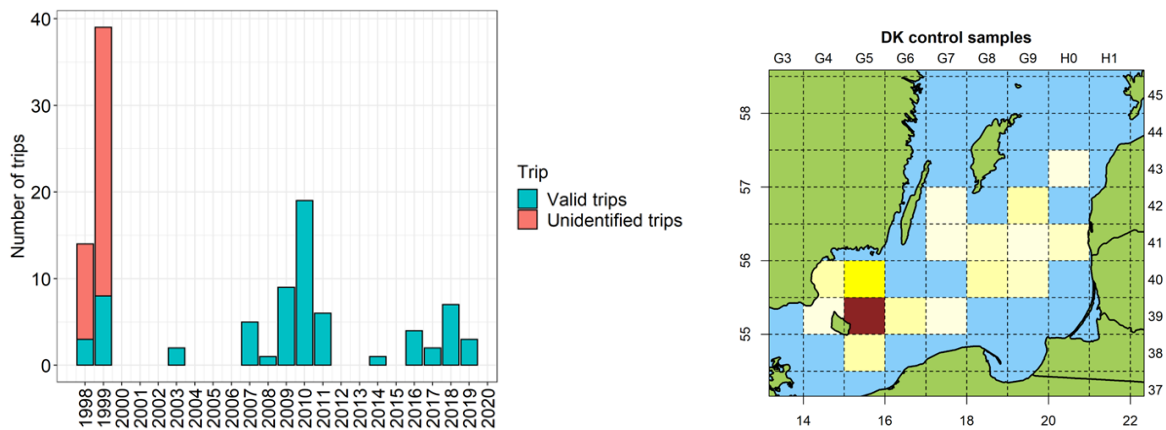


Figure 7.1.3. Coverage of Latvian pelagic fishing trips in Danish control samples.

7.1.3. Main outcomes of the analysis done

Total landings seem quite consistent between Danish control samples and Latvian logbook data. Only 6 trips have a difference larger than 10 %. In the analysed period 91.3 % of trips have a difference of less than 10 %. Fluctuations show no clear trend (Figure 7.1.4).

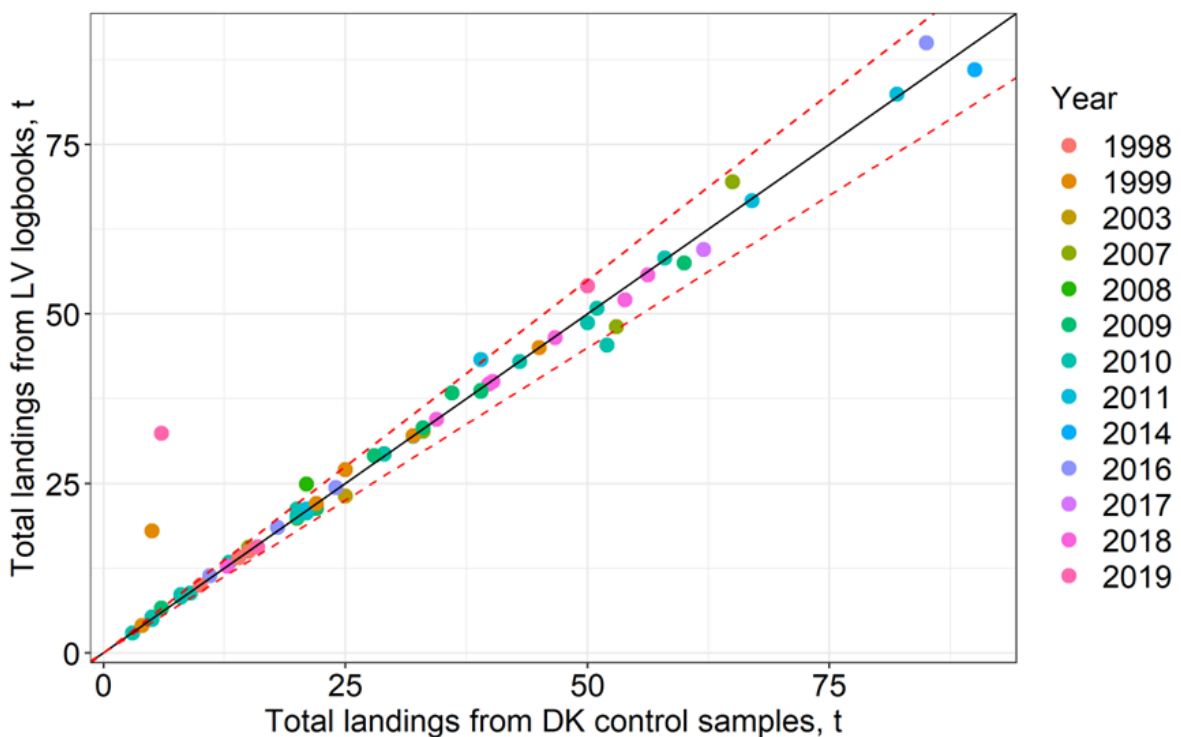


Figure 7.1.4. Relationship between total landings per trip estimated by Danish control samples and Latvian logbook data (1 dot = 1 trip). The middle black line represents a 1:1 ratio. Red dashed lines correspond to a 10 % range.

Sprat landings seem consistent between Danish control samples and Latvian logbook data. In the analysed period 84.1 % of trips have a difference of less than 10 %. Fluctuations have no clear trend. For herring differences are larger, however, herring overall landings are significantly smaller than sprat, thus differences by landing weight are considered negligible compared to the total weight of pelagic landings (Figure 7.1.5).



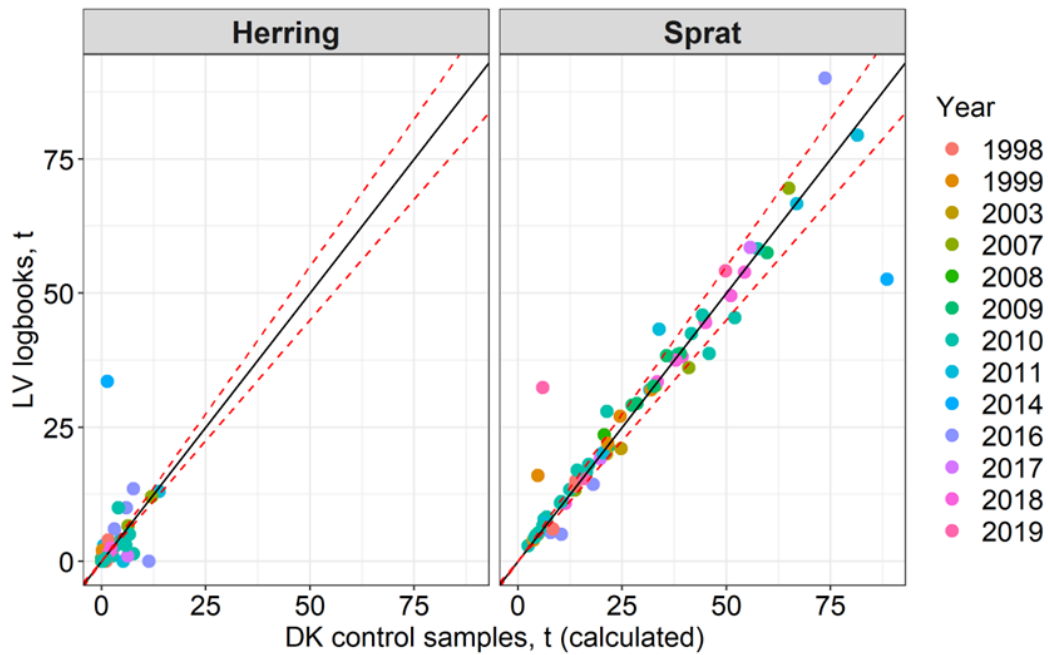


Figure 7.1.5. Relationship between herring and sprat landings per trip estimated by Danish control samples and Latvian logbook data (1 dot = 1 trip). The middle black line represents a 1:1 ratio. Red dashed lines correspond to a 10 % range.

Sales notes show good consistency and in most cases are in line with Latvian logbook data. Few observed differences are likely due to typing errors (Figure 7.1.6).

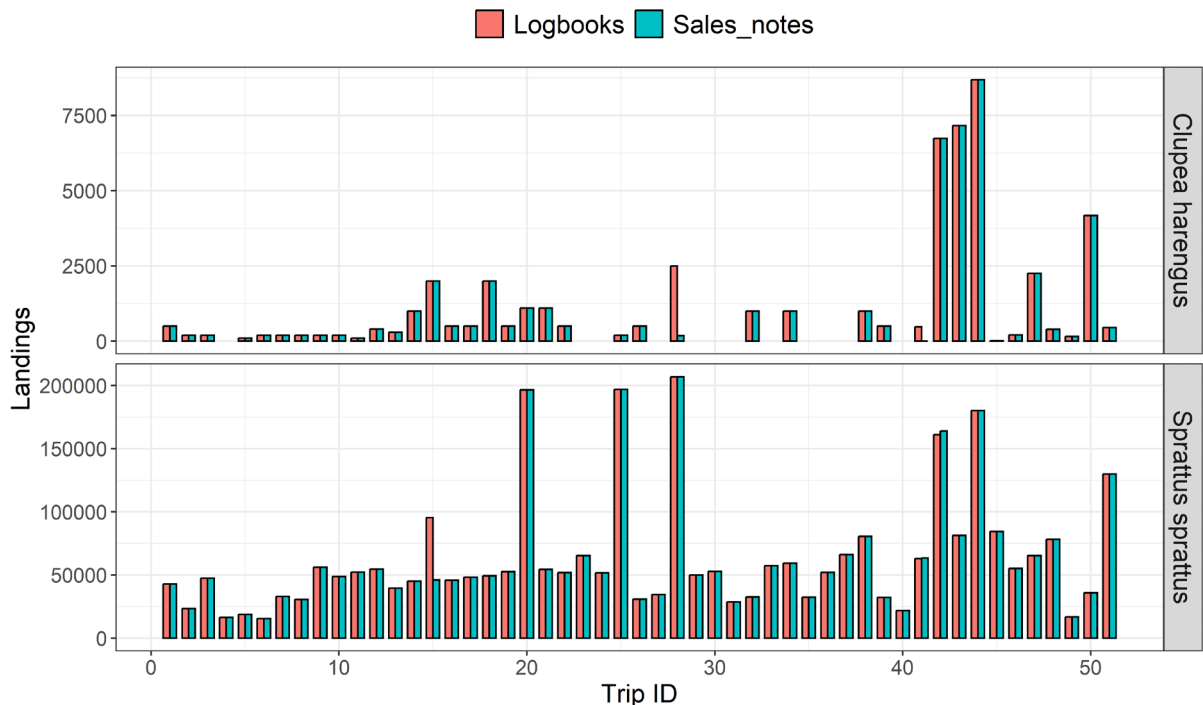


Figure 7.1.6. Comparison between Latvian landing data in sales notes (Denmark) and Latvian logbooks (51 trips in 2019-2022).

According to received information from the Latvian control agency, the agency controls the accuracy of species composition determination and landing weight estimation. Thus, no separate biological samples are taken by the agency and there are no additional data for the analysis. Although fishing

trips with agency participation can be identified, analysis of overall control intensity and potential differences when comparing with trips without agency oversight were not analysed at this point.

7.1.4. Advice to the benchmark

**The national data to not be updated as there are no indications after the analysis that data can be improved. The country will not provide new time series.**

8.1. Lithuania

Fishery

Lithuania has 5% of Baltic sprat and 2.6% of Central Baltic Herring (CBH) quotas. Relative share of SPR vary from 64 to 89 percent depending on TAC allocated by EU Regulations. (Fig1.)

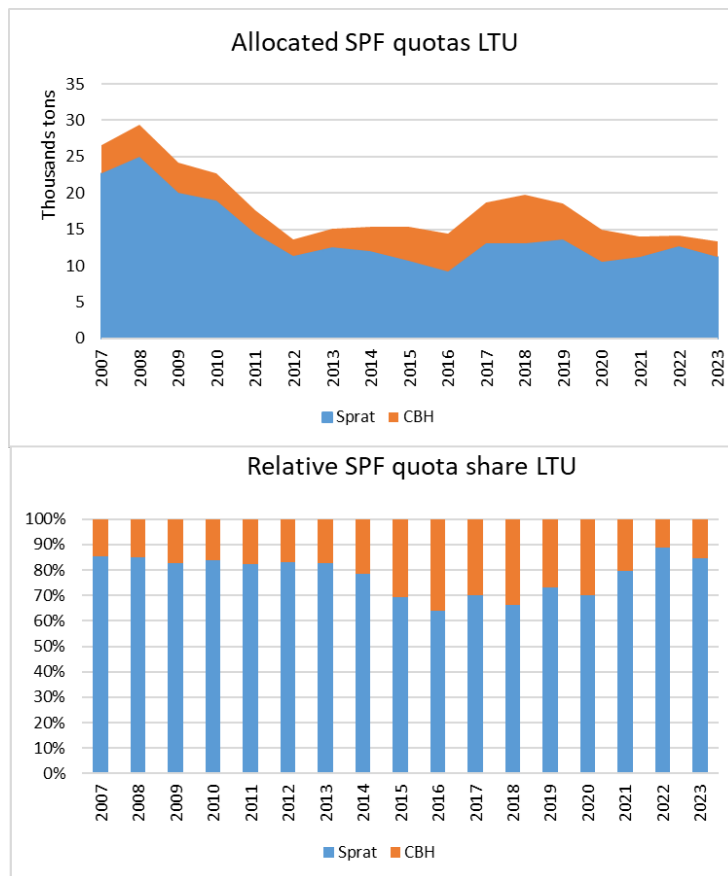


Figure 1. Allocated SPF quotas and relative sprat rate (blue) in the Lithuanian Baltic fishery.

Trawlers with LOA 24 meters and more are taking about 95% catches of SPF. They are fishing in subdivisions 25, 26, 28.2 and 29 (Fig.2). Sprat takes the biggest share of catches in almost all ICES statistical rectangles, except 40HO (close to Lithuanian coast) herring takes the biggest share of catch.

Up to 5% of herring and very tiny quantity of sprat is fished by small scale fishing vessels mainly with fyke-nets. All these catches are landed for HUC in Lithuanian ports.

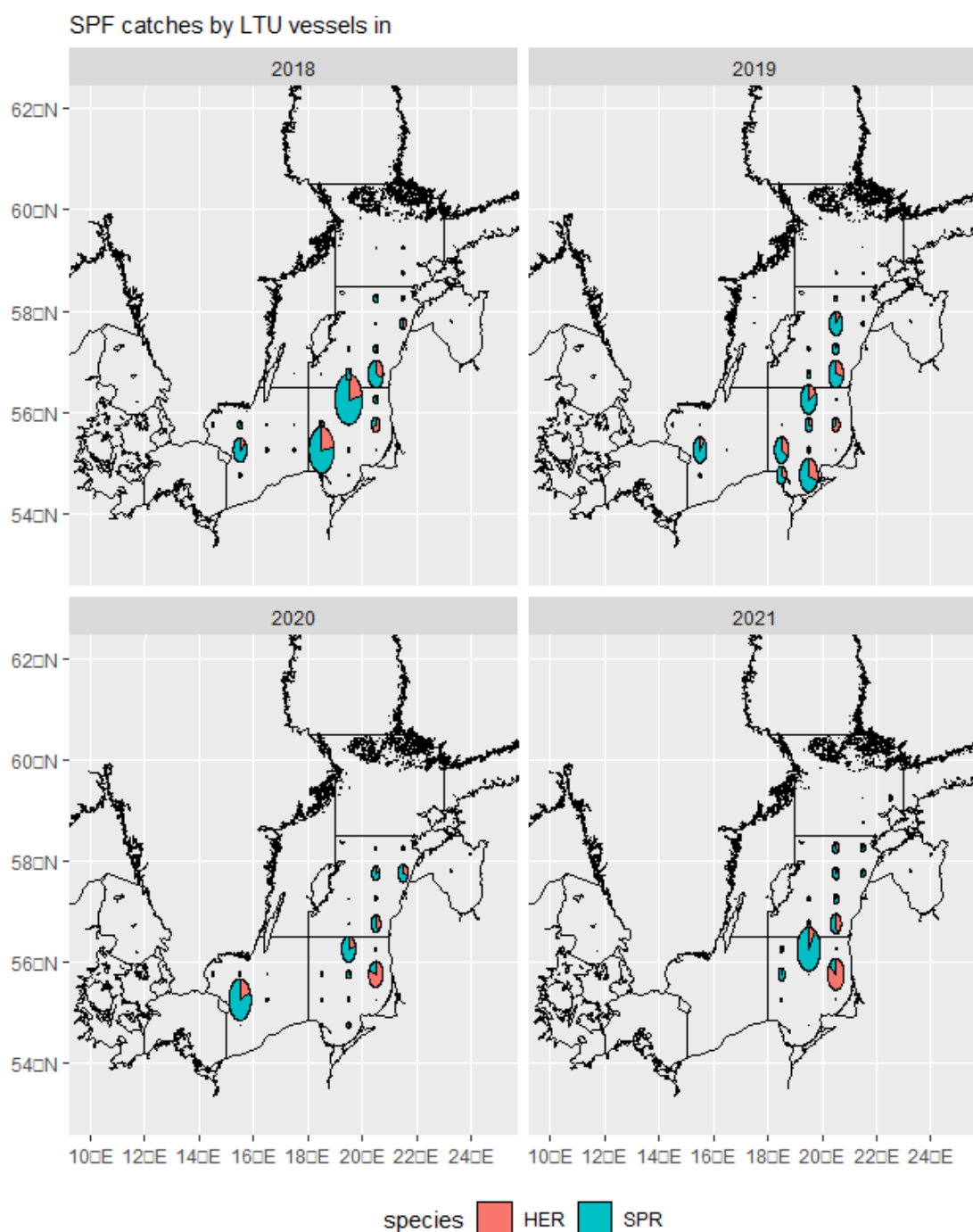


Figure 2. SPF catches by Lithuanian vessels in 2018-2021 by CES statistical rectangles

Despite that most of SPF catches are made in eastern part of Baltic Sea most of SPF catches are landed in Denmark (Fig.3). Less than 1% of total SPR catches are landed in Lithuania (Klaipeda port). Until introduction ban for direct fishing for Eastern Baltic cod, about 10% from total HER catches were

landed in Lithuania. Then in 2020 – 2021 share of HER landings in Lithuania increased up to 50% from total landings, however from 2022 Lithuanian trawlers shifted to do landings in foreign ports again.

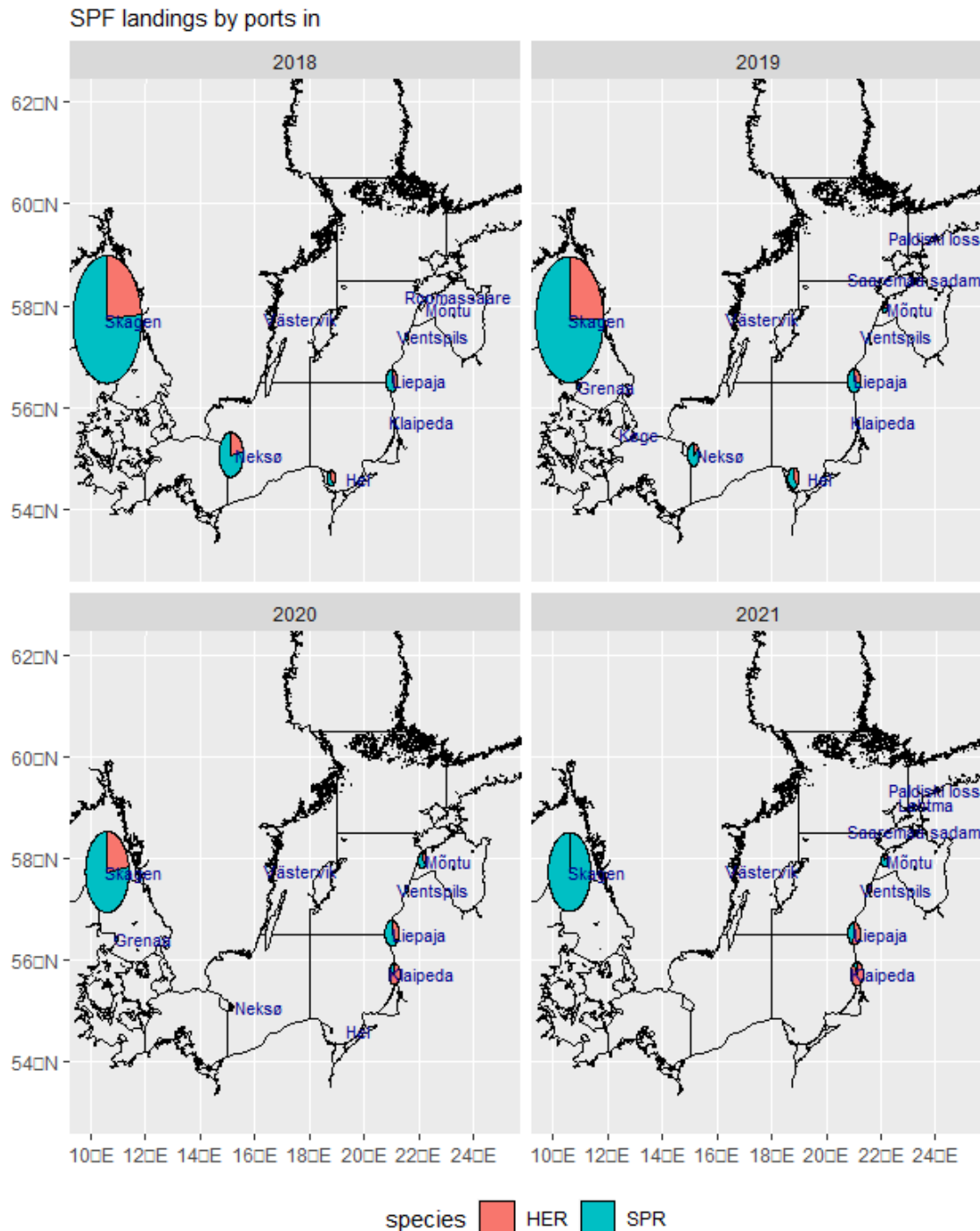


Figure 3. SPF landings from Lithuanian vessels in 2018-2021

Before 2004 the Lithuanian fishery for sprat and herring (SPF) in the Baltic was mainly for human consumption and most of landings were made in Lithuania. From about 2007 sprat started to fish mainly for industrial purposes and landed mostly in Denmark. Direct fishing for sprat for industrial landings (IND) is conducted with vessel with LOA 24 and more meters using trawls with mesh size 16 mm. Bycatch of herring caught by these gears is landed for IND. Most of IND landings are made in Skagen. Trawlers with mesh size 36 and 40 mm are fishing for herring for human consumption (HUC). Landings for HUC are made in Lithuania, mostly, and in Latvia (Fig.3).

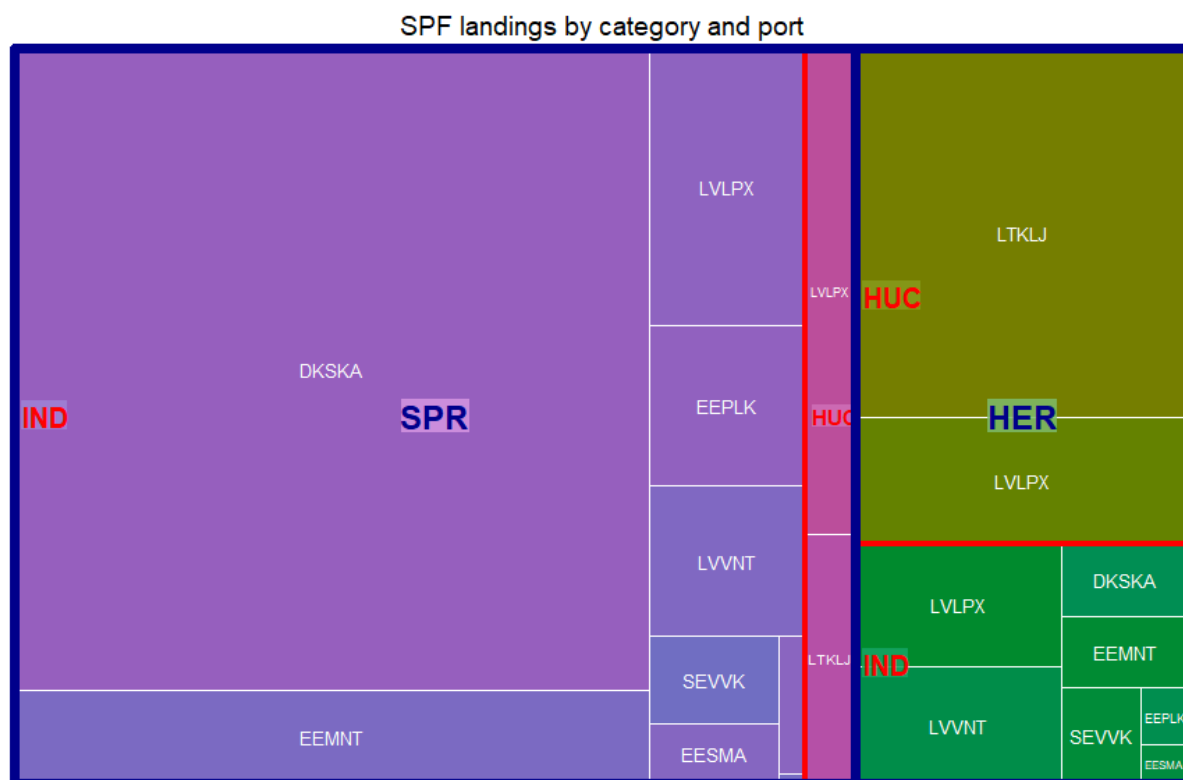


Figure 3. Distribution of SPF landings by species, landing category and landing port in 2021.

*Approach taken to analyze if there are errors in the time series of catch data due to inadequate reporting of species and/or other reasons*

Fisheries Service under Ministry of Agriculture (FS) is responsible for collecting logbook, landing declaration and sales notes data. FS is responsible for control of quota uptake.

Landing declaration figures are used for quota uptake control from 2004. If some inconsistencies detected during import of landing declaration data corrections can be made only after consulting of master or owner of the vessel concerned. It is the possibility to update/correct figures in the data system if master of the vessels provides reasonable proofs and in the reasonable period. Any other corrections are illegal. Scientific analyses and estimations of catches or landings may be used for discussions, but not as a basis for correction of official landing figures, except if it was court decision.

Earlier (1992-2004) figures from monthly reports (paper format) were used for quota uptake. These reports were based on logbook figures, and it was vessel owner’s responsibility to ensure reliability of these reports. Cross checks between logbook and monthly report figures were made regularly. Unfortunately, most of these primary data were lost during the transition period and are not imported in the present data system.

Overview of data sources used when submitting data from the Baltic to WGBFAS			
MS	Landing category	Time period	Data source
Lithuania	All categories	2004 – present	Landing declarations – available in data system.
		1992-2003	Paper monthly reports, not imported in the data system.
		Before 1992	No clue

As it was stated earlier most of SPF landings are made abroad. Only landings for HUC are made in Lithuanian ports therefore, sampling for catch composition of these landings does not adequately cover the whole SPF landings which are mostly designated for IND.

To achieve better sampling coverage cooperation with data collection institutions in Denmark is ongoing. Thanks to this cooperation, analysis of catch composition of industrial landings made by Lithuanian vessels in Danish ports in the period from 2009 to 2020 was done. Results of this analysis are provided to ISSWG “CS small pelagic in the Baltic” in 2021. However, according to Lithuanian law these estimations could not be used for correction of official landing figures.

## 9.1. Poland

### 9.1.1. Fishery

According to the current fishing opportunities in the Baltic Sea for 2023, Poland has about 25% of the EU quota for central Baltic herring and 29% of the sprat (Regulation (EU) 2022/2090). Most of the Polish herring catches come from ICES subdivisions 25 and 26 and midwater trawlers. These herring catches are mainly directed toward human consumption. Herring is also fished in the coastal areas and lagoons (Vistula Lagoon, Szczecin Lagoon) by small-scale fishery using trapnets and gillnets, but the contribution of these catches typically constitute less than 10% of the total Polish herring catches. Most of the Polish sprat catches are taken by midwater trawlers for industrial purposes.

### *9.1.2. Approach taken to analyze if there are errors in the time series of catch data due to inadequate reporting of species and/or other reasons*

A misreporting of central Baltic herring and sprat can exist, and where possible, it is partly accounted for by Poland. Historically, when the data on the bycatch of small herring in the sprat catches were considered representative, the correction of the Polish catches reported to the WGBFAS has been made. Based on the case-by-case expert assessment, when representative data collected by the onboard observers from a given ICES subdivision and quarter were available, corrections have been made. The estimated proportion of herrings in the sprat landings has been used to correct the input figures on the national level and provide as accurate data as possible for assessment.

In line with ICES CM 2012/ACOM:10: WD 5 Walther et al., it is hard to make an accurate estimate on the proportion of herring and sprat in the landings from industrial trawl fisheries with small meshed trawls. These types of trawlers account for the majority of Polish catches. According to the current legal regulations, the permitted margin of tolerance in estimates recorded in the fishing logbook of the quantities in kilograms of fish retained on board shall be 10 % for all species (Article 14(3) of Regulation (EC) No 1224/2009). However, by way of derogation from Article 14(3) of Regulation (EC) No 1224/2009, for catches that are landed unsorted the permitted margin of tolerance in estimates recorded in the fishing logbook of the quantities in kilograms of fish retained on board shall be 10 % of the total quantity retained on board (Article 13 of Regulation (EU) 2016/1139). This mainly affects estimates of the catches obtained by trawlers, especially using Refrigerated Sea Water (RSW) systems.

The data used to report official catches are based on the amounts registered in logbooks and not on landing declarations or sales slips. This approach results from the data analysis which has shown that catches registered in logbooks are more accurate. In addition, this information is more detailed in

terms of fishing area, gear, time, etc., which is important for the level of data aggregation required by the ICES assessment WGs.

Not all fishing vessels are controlled for the adequacy of the reported catches. The controlling agency has the discretion to determine whether or not to collect biological samples for species composition and weight determination through visual inspection. Not every inspection event will result in the collection of samples. If discrepancies are found between the logbook and the inspection results, the agency may recommend the skipper update the logbook to match the inspection findings. When the data is updated, the corrected information will be reported to ICES for reporting purposes. However, it is not possible to track whether the data was corrected or not.

As part of the DCF, Poland constantly conducts at-sea observed trips from all types of fisheries. Samples collected at sea are considered to be more reliable than those collected on shore. Therefore at-sea sampling data were used in the main analysis of misreporting. However, a comparison of the results with the data obtained by the foreign controlling agency confirmed the trends observed in the Polish data.

The analysis of misreporting of herring and sprat consisted of the following steps. First, data from at-sea observed trips targeting pelagic species in the period 2013-2020 were extracted from the database. The catch composition at a trip level was then calculated. The dataset was combined with official catch statistics from the same trips, which allowed us to compare the shares of different species in total catches. The results were visualized on a set of plots presented below.

### *9.1.3. Main outcomes of the analysis done*

- Typically, one species (herring or sprat) is dominating in the catches of fishing vessels from which biological samples have been obtained (Fig. 1).
- Overall, there is a relatively good agreement between % of species observed by the onboard observers and reported by fishers. Most of the points in the density plots are located close to the extremes (0 or 100% contribution) with relatively low deviation (Fig. 1). If the misreporting is present, it is skewed towards overreporting the herring catches, and rarely the opposite situation is observed (overreporting of sprat).
- When misreporting at higher levels (>10%) occurs, it is mainly observed in the fishing vessels that report the lowest total catches in the given trip (Fig. 2).
- Overall, through the years, the median differences in the percentage of species observed and reported in both species are within the range of ~2.5%, except in 2019, when a higher level of misreporting was observed (Fig. 3). This pattern can be partially caused by the overrepresentation of single fishing vessel selected for biological sampling, which may result in biased results.

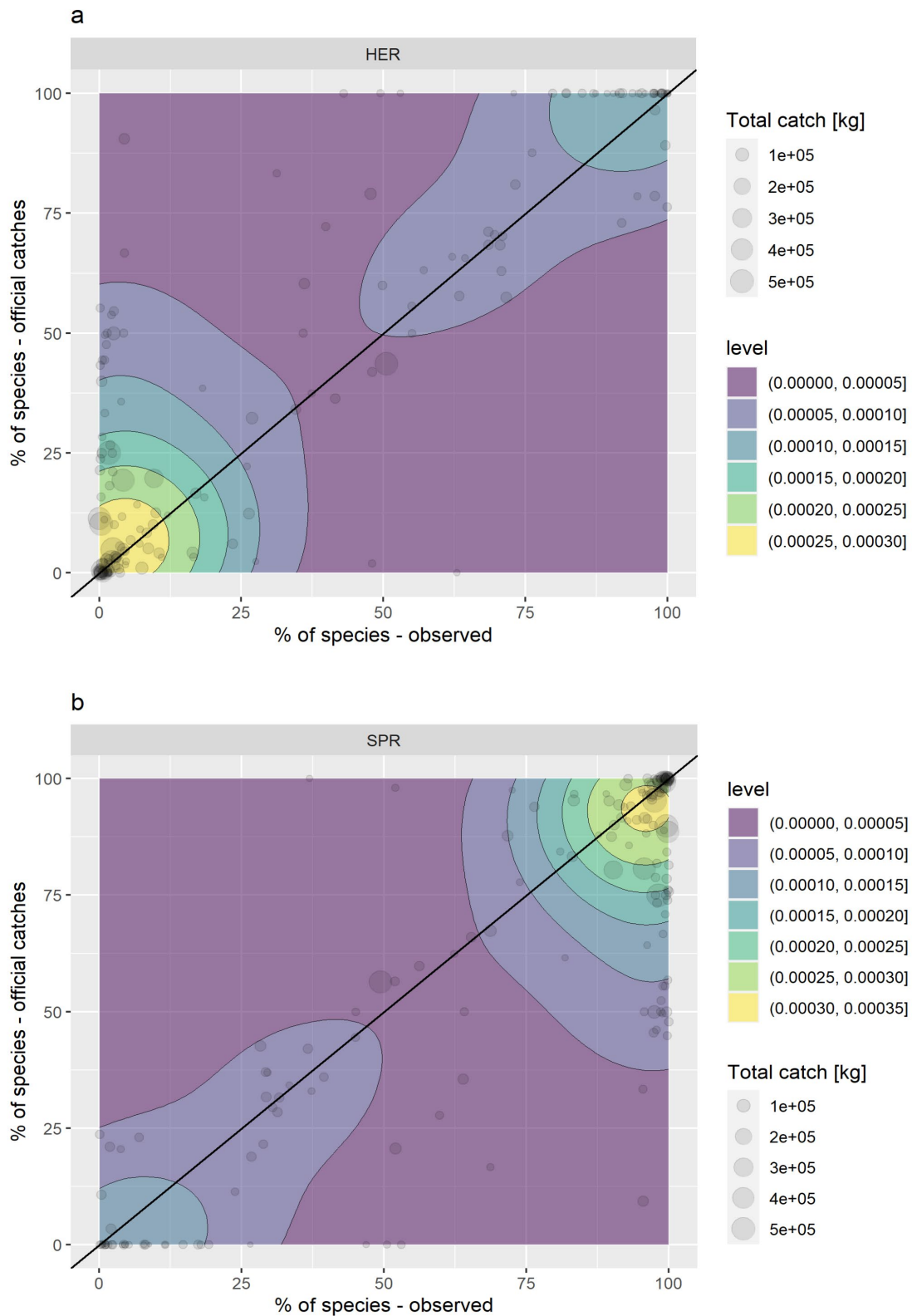


Fig. 1. 2D kernel density of the percent of herring (a) and sprat (b) observed and in official catches. The distribution of points was visualized using ggplot2::geom\_density\_2d function. The color gradient indicates the density of points, while the size of points indicates the total catch of the fishing vessel in the given trip.



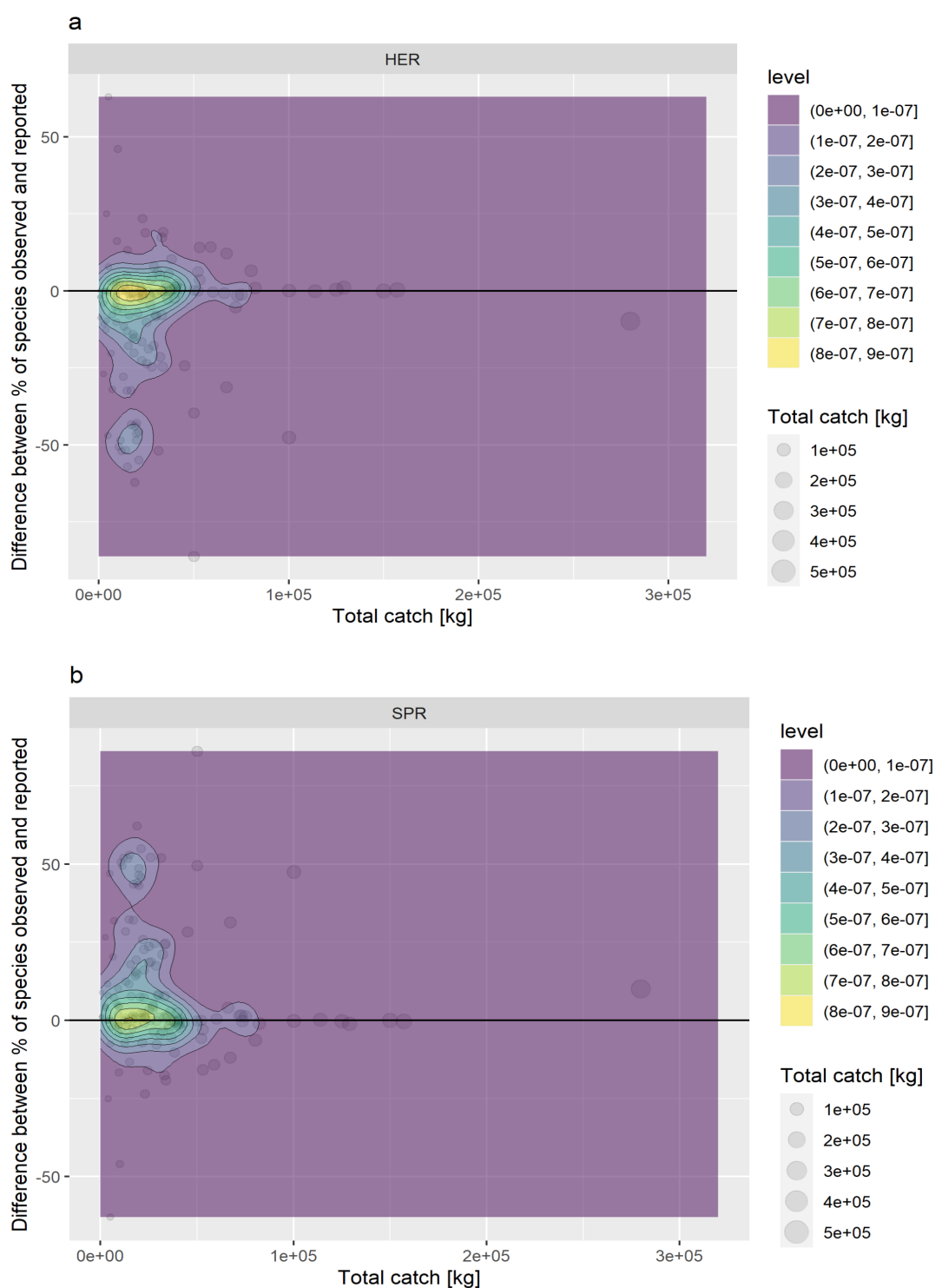


Fig. 2. 2D kernel density of the difference between the percentage of herring (a) and sprat (b) observed and reported as a function of the total catch of the fishing vessel in the given trip. The distribution of points was visualized using `ggplot2::geom_density_2d` function. The color gradient indicates the density of points, while the size indicates the fishing vessel's total catch in the given trip.

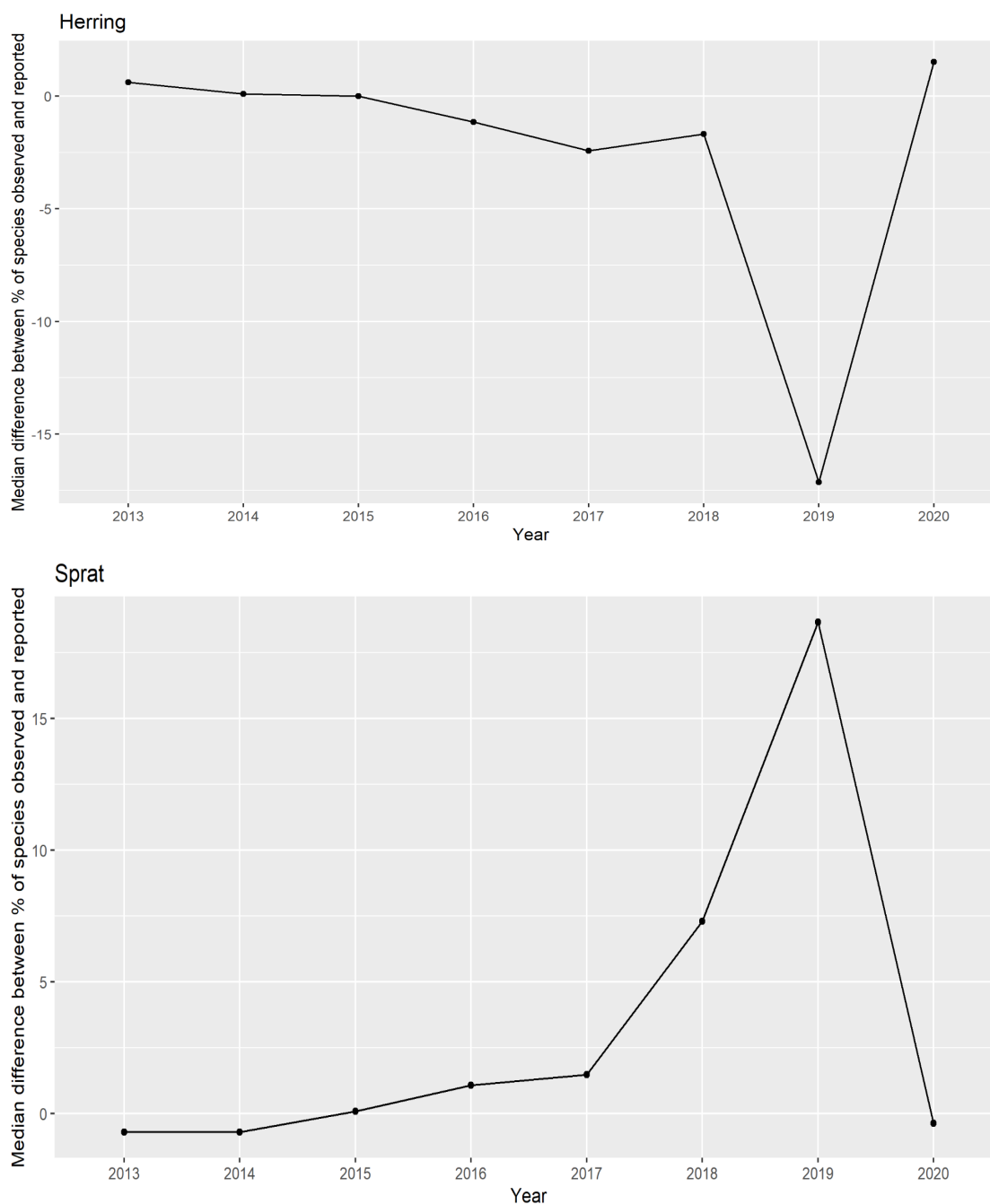


Fig. 3. Time series of the median difference between the percentage of species observed and reported in herring (upper panel) and sprat (lower panel).

#### 9.1.4. Advice to the benchmark

- *The national data to not be updated* as there are no indications after the analysis that data can be improved. The country will not provide new time series.
- A higher level of misreporting was observed in 2019, after which the misreporting in 2020 moved back to the acceptable level of ~0%. The precise causes of this pattern are not known. It might be related to the ratio of the quota in herring and sprat or be a product of chance (poor representativeness of the samples). However, that year was indicated as potentially less trustworthy than others.

## 10.1. Sweden

### 10.1.1. Fishery

Sweden has about 33% of the EU quota for central Baltic herring and 19% of the sprat. Presently, most of the catches are taken by trawlers that fish herring and sprat for industrial purposes and land in Denmark. Smaller-scale coastal fisheries targeting herring for human consumption also take place and have strong cultural significance even if not large landings.

### 10.1.2. Approach taken to analyze if there are errors in the time series of catch data due to inadequate reporting of species and/or other reasons

A set of unsupervised anomaly detection techniques was used to try detecting the possible presence of misreporting in the Baltic small pelagic fishery for Herring and Sprat and quantify it.

The datasets used in analysis contained information relative to the study area (Subdivisions 27.3.d.25,...,27.3.d.32, excluding 27.3.d.28.2, 27.3.d.30, 27.3.d.31) in a 22 year time-span (1999-2021). Commercial data included logbook data as well as landing declaration data. Logbooks contain information in time and space on the effort (e.g. vessel and gear features) and the species caught (quantities, taxonomy, contribute of the species to the catch, among the others), being the primary source of commercial information submitted to stock assessment. Landing declarations are generally considered a more accurate estimates of the amount landed by the fishermen but, because they integrate the output of sometimes long trips, taking place over multiple subdivisions, they frequently do not have the spatio-temporal resolution needed by end-users and required for more sophisticated anomaly detection. Environmental information was also considered, namely main temperature at a specified depth interval (-15 to -45 meters) for the Baltic Sea extracted from a pre-existing model (CMEMS Baltic Sea Physical Reanalysis BALTICSEA\_REANALYSIS\_PHY\_003\_011, Liu, 2019) and bathymetry information extracted from NOAA databases (ETOPO Global Relief Model, NOAA 2022). Finally, information on the TAC was compiled for the years 1997 to 2020 and merged to the remaining data.

The analytical approaches used in this study include the application of the Newcomb-Benford Law (NBL, for a complete review see Nigrini, 2012) and the application of both regression based (RB here-after) and Isolation Forest algorithms (IF here-after Liu, 2008).

NBL was applied both to logbook data and to landing declarations in order to highlight the possible presence of anomalies in the data overall. In the NBL analysis approach First (F1T here-after) and First Two digits (F12T here-after) tests were used to determine whether the data were consistent or not with the NBL model (Nigrini, 2012). In particular, mean absolute deviation statistic (MAD here-after, Nigrini, 2012) statistic was used (as in Silva Azevedo et al., 2021) with critical cut-off scores reported in Nigrini and Drake (2000).

First Digits	First-Two Digits	Conformity
0.000 to 0.006	0.000 to 0.0012	Close conformity
0.006 to 0.012	0.0012 to 0.0018	Acceptable conformity
0.012 to 0.015	0.0018 to 0.0022	Marginally acceptable conformity
above 0.015	above 0.0022	Nonconformity

Tab.1: Cut-off critical values of mean absolute deviation to assess the conformity of data to NBL (Drake and Nigrini, 2000).

Logbook data (filtered,  $n = 145680$  records, all species included), did not conform to the NBL both at the F1T ( $MAD = 0.016$ ) and at F12T ( $F12T, MAD = 0.012$ ). Landing declarations of both Sprat and Herring (108747 observations) exhibited acceptable conformity at F1T ( $F1T, MAD=0.0077$ ) but not at the F12T, ( $F12T, MAD=0.0035$ ), (Fig. 1).

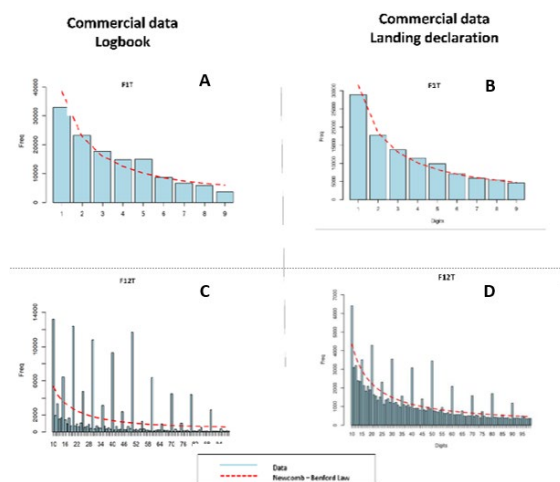


Fig.1: Digit analysis on commercial landing declaration catches of Herring and Sprat. Database unfiltered (145680 records) tested at first digit (F1T, case 'A', 'B') and first-two digits (F12T, case 'C', 'D') with second order test included in both case (SO1T and SO12T, case 'B' and 'C')

Unsupervised approaches were used in an attempt to estimate the amount eventually misreported. Information on space (e.g. subdivision), time (e.g. year, month), features of the boat (e.g. gear type), abiotic (e.g. bathymetry interval) and legislative environment (i.e. TAC) was included in the analysis. Data used encompassed most of the catches namely those of mid-water and bottom trawlers (PTM, OTM, PTB, OTB) landing in Denmark and Sweden. Observations were re-assigned in two main groups: Pelagic (aggregating PTM + OTM) and Bottom Trawlers (aggregating PTB + OTB) and split into two bathymetry classes: coastal (> -70 m) and offshore hauls (< -70 m).

The dataset used in IF analysis was stratified on the categorical variables (gear, ICES Subdivision, bathymetry class). Two different parameterization were tested: IFSB (from “Isolation Forest Basic Variables”) and IFALL (from “Isolation Forest Basic Variables”). In IFSB the analysis was built using as features: Proportion of Herring, Quarter (ordinal, encoded), Year. Temperature, TAC for Herring and TAC for Sprat were included when IFALL was performed. In both settings, on each unit of the stratification presented, the algorithm was run by fitting 500 isolation trees and using a sample size equal to one fourth of the total number of observations in the level. A threshold of half of the possible anomaly score (anomaly score < 0.50) defined the inliers, while the outliers were furthermore divided in possible outliers (0.50 < anomaly score < 0.55) and likely outliers (anomaly score > 0.55). Strata with less than 100 observations were not classified and the relative observed proportions were considered not anomalous.

RB approach was based on two modelling framework: Generalized Additive Models, GAM here-after (Hastie and Tibshirani 1986, see Wood 2006 for a complete review) and Generalized Additive Models for Location Scale and Shape (GAMLSS here-after, Rigby and Stasinopoulous 2006, see Rigby and Stasinopoulous 2017 for a complete review). A series of models was parametrized using the proportion of Herring in each haul as a response and a set of covariates including information on i) gear, ii) vessel, iii) time, iv) space, v) auxiliary effects. The models had the general formula:

$$Y = \beta_0 + f_i(\text{year}) + f_i(\text{month}) + f_i(\text{lat, lon}) + f(\text{length}) + f_j(\text{vessel\_ID})$$

Where *i* defines intercepts (one for PT and one for BT), a random effect is assigned to the vessel call-sign. RB models were compared, when possible, by using the diagnostic Akaike Information Criterion (AIC), R-squared (R<sub>2</sub>) and the Cox & Snell Generalised (Pseudo) R-squared (R<sup>2</sup><sub>cs</sub>), visual inspection of the residuals and a parsimonious approach (choose the simplest model i.e. with less *knots*). Two models performed best: Quasibinomial Generalized Additive Model (qbGAM, here-after) and Beta Zero and One Inflated GAMLSS (beinfGAMLSS here-after). Both are reported as these cannot be directly compared using the same diagnostics. The inspection of residuals from these models found them to be normal only in certain cases (i.e. when z-scores for beinfGAMLSS as shown in Appendix Fig. 5 and on a lesser extent when scaled Pearson type residuals for qbGAM are considered as shown in Appendix Fig. 4).

Both models approaches were tested also using the de-trended qq-plot (worm plot). The results were not satisfactory: several points are falling outside the confidence band of the plot. The performance of the `beinfGAMLSS` against the worm plot had improved when fit complex splines (~ 200 knots) for the term relative to the interaction between latitude and longitude (Appendix Fig. 6). On the other hand, since the model may be influenced by biased data points, if any, incrementing further the number of knots in order to fit the data was avoided in order to avoid the influence of the eventually biased information. Discrepancies highlighted indicate that the model should be improved. Models coming from both frameworks (`qbGAM` and `beinfGAMLSS`) and relatively complex spatial interaction (no more than an amount of nodes “k” = 200) are presented but should be considered with extra-caution and as preliminary.

The classification in both the models constituting the RB approach was based on the definition of residual as the discrepancy between the observed and the predicted value of the response was used to quantify the anomaly score of a given observation (Chandola et al., 2007). Standardized residuals were used for the selected models to determine the eventual anomalous nature of each data point. R packages used for modelling `qbGAM` (“`mgcv`”, Wood 2017) and `gamlss` for “`beinfGAMLSS`” (Rigby and Stasinopoulos, 2005) provide different types of residuals. In the first case scaled pearson residuals (PRS) while in the latter z-scores (ZSC) were used. The threshold for z-scores and PRS were:  $0 < \text{PRS|ZSC} < 1$  for inlier,  $2 < \text{PRS|ZSC} < 3$  for possible outlier,  $\text{PRS|ZSC} > 3$  for likely outlier.

In both IF (IFALL and IFBS) and RB (`qbGAM` and `beinfGAMLSS`) it is not possible to indicate *a priori* if the centre of gravity of the observations consists in non - misreported or misreported hauls. Consequently these scenarios (“few-misreport”, FM here-after; “most-misreport”, MM here-after) were explored assuming that misreport occur in the same direction in a given context. Moreover those observations classified as “possible misreporting” can be regarded as records “correctly reported” (PC here-after) or “misreported” (PM here-after). The combination of the four techniques (IFBS, IFALL, `qbGAM`, `beinfGAMLSS`) with the different scenarios (FM, MM) and the treatment of the possible misreporting (PC, PM) led to 12 classifications and thus 12 alternative time-series (see below).

The generation of alternative time series was performed after the classification by: i) calculating an expected proportion of the *C.harengus* species in the *C. harengus* + *S.sprattus* total catch for each context using the observations classified as normal, ii) compare the proportion expected with the one observed in the observations classified as anomalous, iv) multiply the difference between the two times the total catch in case of an anomalous observation and vi) use the algebraic sum between this quantity and the total catch to shift the amount between the species according to the models.

### 10.1.3. Main outcomes of the analysis done

The results show variability in the the predictions of the different models (Fig. 2). The predictions of corrected catch under the “few misreport” hypothesis are relatively consistent with the catch originally reported. Under the “most misreport” hypothesis the models predicted catches very different from the original reported ones. Indicating under-reporting of Herring (over-reporting of Sprat) in the past (2001 - 2011) and slight over-reporting of herring and under-reporting of sprat in recent years. However, depending on the treatment of possible misreporting (PC or PM), some approaches pointed in the opposite direction (e.g. “IFBS\_MM\_PC”), indicating a lack of unanimity in the predictions of the models and a possible pivotal role of the “possible misreporting” observations which interpretation can substantially change the interpretation of results.

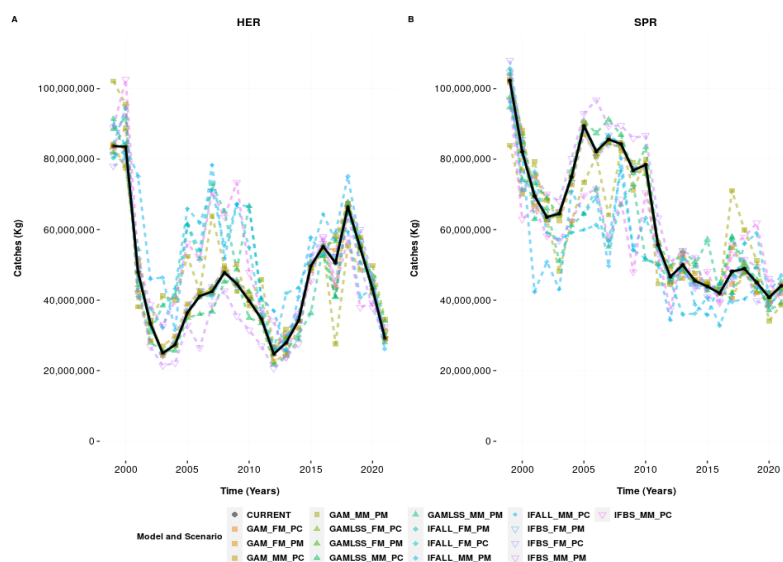


Fig.2: Outputs from the different models (GAMLSS = beinfGAMLSS, GAM = qbGAM, IFBS = IFBS, IFALL = IFALL) in the different scenarios (indicated by dashed lines in different color, as described in the legend, FM\_PC = few misreport and possible misreporting is not misreporting, FM\_PM = few misreport and possible misreporting is misreporting, MM\_PC = most misreport and possible misreporting is not misreporting, MM\_PM = most misreport and possible misreporting is misreporting), versus the reported catch (indicated by a solid black line), divided by species (A: Herring, B: Sprat). Black line corresponds to the reported catches.

When all models in the two main scenarios (FM and MM) are averaged the considerations above translate in an average prediction relatively in line with the reported catch in the “few misreport” hypothesis (Fig. 3 case A and B) and a predicted catch that diverge from the reported one in the “most misreport” hypothesis (Fig. 3 case A and B), (Tab. 2; Tab.3).

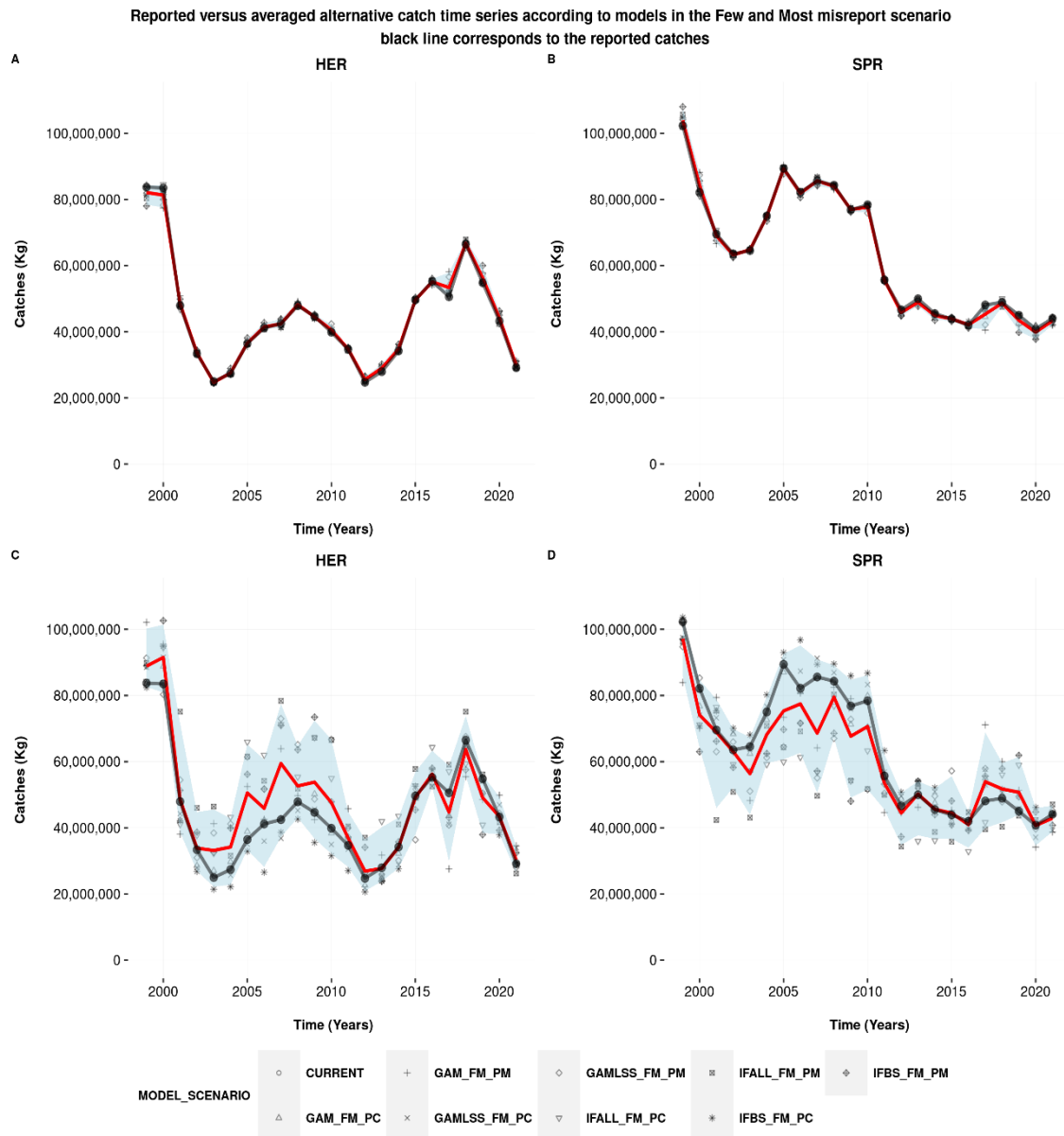


Fig.3: Outputs from the different models (GAMLSS = beinfGAMLSS, GAM = qbGAM, IFBS = IFBS, IFALL = IFALL) in the few (A, B) and most (C, D) misreport scenarios (indicated by dashed lines in different color, as described in the legend FM\_PC = few misreport and possible misreporting is not misreporting, FM\_PM = few misreport and possible misreporting is misreporting), MM\_PC = most misreport and possible misreporting is not misreporting, MM\_PM = most misreport and possible misreporting is misreporting) averaged, versus the reported catch (indicated by a solid black line), divided by species (A: Herring, B: Sprat). Light blue bands indicate the interval in which 95% of the prediction of the models are falling. ). Black line corresponds to the reported catches.

APPROACH	SCENARIO	QUANTITY	MEAN	SD
GAM	FM_PM	HER	3,40	3,43
		SPR	2,75	3,45
	FM_PC	HER	1,02	1,45
		SPR	0,85	1,52
	MM_PM	HER	19,29	18,42
		SPR	12,60	11,14
MM_PC	HER	5,24	4,90	
	SPR	3,35	3,51	
GAMLSS	FM_PM	HER	2,86	2,79
		SPR	2,21	2,74
		HER	0,23	0,26

	FM_PC	SPR	0,16	0,24
	MM_PM	HER	22,57	22,96
		SPR	12,77	10,77
	MM_PC	HER	8,61	5,34
SPR		6,03	4,30	
IFALL	FM_PM	HER	1,33	1,20
		SPR	0,95	0,95
	FM_PC	HER	1,10	1,44
		SPR	0,87	1,33
	MM_PM	HER	29,42	26,77
		SPR	16,90	12,32
	MM_PC	HER	25,69	22,93
		SPR	15,42	10,85
IFBS	FM_PM	HER	3,93	2,54
		SPR	2,91	2,58
	FM_PC	HER	1,98	1,74
		SPR	1,53	1,44
	MM_PM	HER	23,92	18,46
		SPR	15,04	9,94
	MM_PC	HER	13,18	7,84
		SPR	8,57	4,30

Tab.2: Mean and standard deviation of the absolute percentual difference between the reported and predicted catches according to the different models (beinfGAMLSS, qbGAM, IFBS, *IFALL*) in different scenarios (FM\_PC = few misreport and possible misreporting is not misreporting, FM\_PM = few misreport and possible misreporting is misreporting), MM\_PC = most misreport and possible misreporting is not misreporting, MM\_PM = most misreport and possible misreporting is misreporting), when data are aggregated by year and species. Since the information on year is omitted here, the absolute value of the predictions for each model and scenarios combination in each year has been calculated and used to compute the statistics shown.

Year	HER				SPR			
	FM		MM		FM		MM	
	mean	sd	mean	sd	mean	sd	mean	sd
1999	-2,71	2,29	4,59	3,76	2,22	1,87	-3,75	3,07
2000	-2,11	2,55	9,33	8,85	2,15	2,59	-9,48	9,00
2001	0,42	1,93	7,20	24,33	-0,29	1,33	-4,96	16,77
2002	0,77	1,52	4,03	20,29	-0,41	0,80	-2,12	10,65
2003	-1,09	0,84	31,44	32,57	0,42	0,33	-12,17	12,60
2004	1,24	2,24	24,02	29,32	-0,45	0,82	-8,77	10,70
2005	0,91	1,79	43,10	36,15	-0,37	0,73	-17,57	14,74
2006	0,78	1,59	14,24	29,10	-0,39	0,80	-7,13	14,58
2007	-0,31	2,48	44,86	40,02	0,15	1,23	-22,26	19,86
2008	0,45	0,74	12,79	17,66	-0,26	0,42	-7,26	10,02
2009	-0,31	1,01	26,05	30,84	0,18	0,58	-15,14	17,92
2010	1,61	2,21	26,50	34,69	-0,82	1,13	-13,47	17,64
2011	0,63	0,86	3,01	13,05	-0,39	0,54	-1,88	8,13
2012	3,29	2,92	11,70	24,31	-1,74	1,55	-6,20	12,89
2013	4,17	2,59	-1,45	23,28	-2,33	1,45	0,81	13,01
2014	2,35	2,08	0,76	17,87	-1,77	1,57	-0,57	13,43
2015	0,15	0,76	-0,79	13,71	-0,17	0,86	0,90	15,51
2016	-0,01	1,08	2,67	7,30	0,01	1,42	-3,52	9,61
2017	3,69	3,90	-5,52	14,76	-3,88	4,10	5,80	15,51
2018	0,09	1,15	-2,71	8,62	-0,13	1,57	3,69	11,71
2019	2,74	3,48	-11,89	12,14	-3,33	4,24	14,48	14,79
2020	2,53	3,32	-2,38	7,07	-2,69	3,52	2,53	7,50
2021	2,32	2,64	2,28	9,10	-1,53	1,74	-1,51	6,02

Tab.3: Mean and standard deviation of the percentual difference between the reported and predicted catches according to the different models (beinfGAMLSS, qbGAM, IFBS, *IFALL*) in different scenarios (FM\_PC = few misreport and possible misreporting is not misreporting, FM\_PM = few misreport and possible misreporting is misreporting), MM\_PC = most



misreport and possible misreporting is not misreporting, MM\_PM = most misreport and possible misreporting is misreporting), when data are aggregated by year and species.

#### 10.1.4. Advice to the benchmark

**The national data will not be updated in the present benchmark but might be updated in the future since there are some indications of possible misreporting. The country is ready to provide one or more new time series that the stock assessors can explore but these are not, for the time being, considered sufficiently reliable for a definitive inclusion in assessment.**

In this work the NBL was used to highlight the presence of possible anomalies and Isolation Forest algorithm and regressive approaches (GAM; GAMLSS) used in trying to estimate the quantities eventually misreported.

The NBL should not be interpreted as evidence of misreporting and alteration of data, but rather highlights the possible presence of anomalous activity and suggests further investigation on the processes originating the data (Nigrini, 2012). Swedish logbook data relative to this fishery did not, in general, conform to the NBL model. Conformity improved when landing declarations were used, but discrepancies were still observed at the F12T. The patterns shown namely those of multipliers of five characterizing a large extent of the records and the improvement of performance with landing declarations suggests that rounding of quantities may have had a role in explaining the discrepancies observed. Patterns observed in logbooks and landing declarations may be explained by misreporting but also by rounding or lack of accuracy in estimation of large catches. Misreporting is usually considered an intentional directional activity while rounding is conceptually distinct and likely more erratic and bi-directional. Overall, both aspects act to change the conformity of the underlying data and introduce inaccuracies in catch reports that may be worth studying more in detail.

IF and RB are unsupervised techniques and unsupervised techniques and, as such, are not able to distinguish between white noise and possible anomalies (Bolton and Hand, 2002, Nisbet, 2018). Consequently, a classification as outliers of the observations by these methods should not be regarded as proof of misreporting but rather as a description of the grade of difference between the classified observation and the others, as well as a possible indication of anomalies in the data that require further investigation (Nisbet et al., 2018). Both IF and RD models application showed substantial variability in results and different of performance in diagnostic analyses. Furthermore, RB models are known to be susceptible to the presence of outliers which renders them non - robust as outlier detection tools. Results may be further influenced by combinations of parameters in both frameworks (e.g. the threshold to be used in order to spot and outlier).

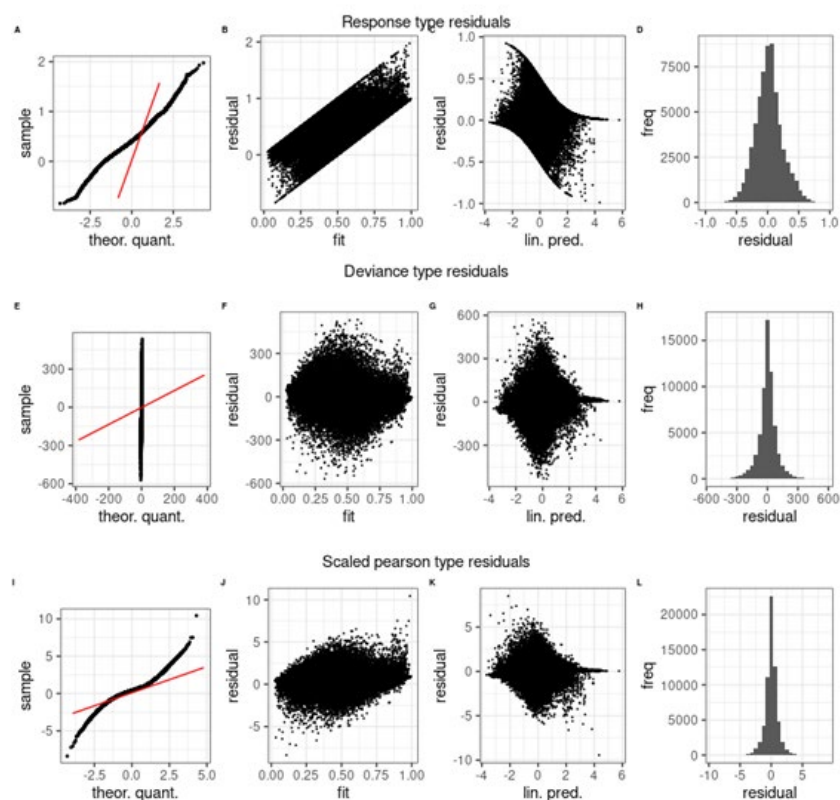
Under the variability observed in both predictions and diagnostics of the models tested in the present study, further research seems to be needed into the identification of a model that is a good descriptor of the expected proportion of Herring in different spatio-temporal and methodological contexts while being robust to different parametrizations and to the possible presence of outliers. Even if the modelling approaches seem to be consistent with a perception of historical misreporting of herring, these results require improvement and tests before strong conclusions can be drawn and a reliable alternative time-series of Swedish catches can be produced. As such, any application of the present results should be considered, for the time being, exploratory.

#### References:

1. Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical science*, 17(3), 235-255.
2. Consul, P. C., & Famoye, F. (2006). *Lagrangian probability distributions* (pp. 21-49). Birkhäuser Boston.
3. da Silva Azevedo, C., Gonçalves, R. F., Gava, V. L., & de Mesquita Spinola, M. (2021). A Benford's law based method for fraud detection using R Library. *MethodsX*, 8, 101575.
4. Drake, P. D., & Nigrini, M. J. (2000). Computer assisted analytical procedures using Benford's Law. *Journal of Accounting Education*, 18(2), 127-146.
5. Hastie, T., & Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398), 371-386.
- 6.
7. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In 2008 eighth IEEE international conference on data mining (pp. 413-422). IEEE.

8. Liu, Y., Axell, L., Jandt, S., Lorkowski, I., Lindenthal, A., Verjovkina, S., & Schwichtenberg, F. (2019). Baltic Sea Production Centre. BALTICSEA\_REANALYSIS\_PHY\_003\_011. COPERNICUS Marine Environment Monitoring Service, 10.
9. Nigrini, M. J. (2012). Benford's Law: Applications for forensic accounting, auditing, and fraud detection (Vol. 586). John Wiley & Sons.
10. NOAA National Centers for Environmental Information. 2022: ETOPO 2022 15 Arc-Second Global Relief Model. NOAA National Centers for Environmental Information. DOI: [10.25921/fd45-gt74](https://doi.org/10.25921/fd45-gt74)
11. Nisbet, R., Elder, J., & Miner, G. D. (2009). Handbook of statistical analysis and data mining applications. Academic press.
12. Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(3), 507-554.
13. Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., & De Bastiani, F. (2017). Flexible regression and smoothing: using GAMLSS in R. CRC Press.
14. Rigby, R. A., Stasinopoulos, M. D., Heller, G. Z., & De Bastiani, F. (2019). Distributions for modeling location, scale, and shape: Using GAMLSS in R. CRC press.
15. Wood, S. N. (2006). Generalized additive models: an introduction with R. chapman and hall/CRC.

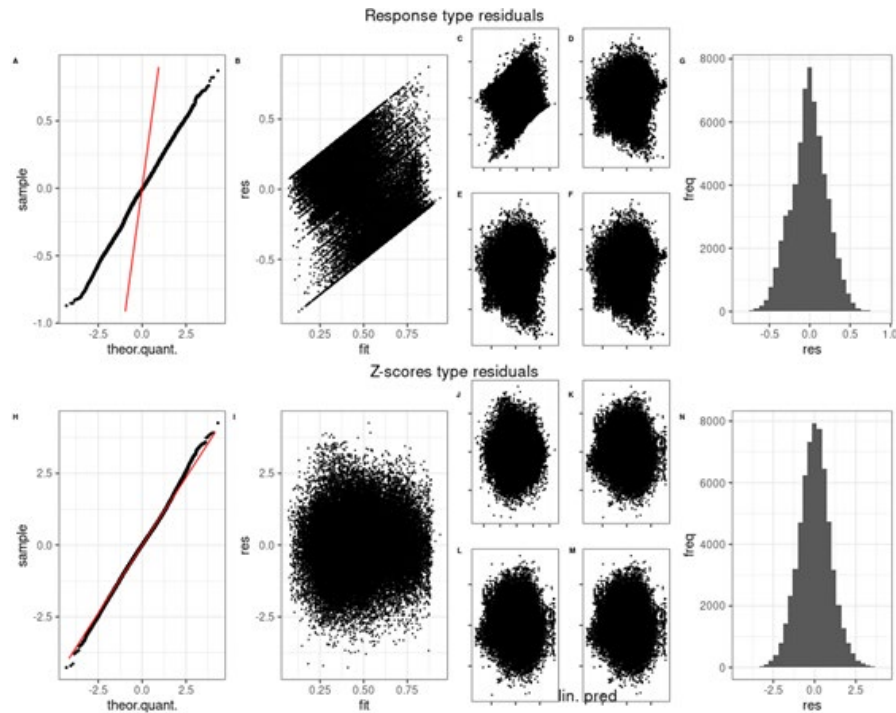
## Appendix



16.

17. Fig. 4: Selected quasi binomial Generalized Additive Model residuals: A = qq-plot of response residuals, B = response residual versus fitted values, C = response residual versus linear predictor, D = histogram of response residual, E = qq-plot of deviance residuals, F = deviance residual versus fitted values, G = deviance residual versus linear predictor, H = histogram of deviance residual, I = qq-plot of scaled - pearson residuals, J = scaled - pearson residual versus fitted values, K = scaled - pearson residual versus linear predictor, L = histogram of scaled - pearson residual

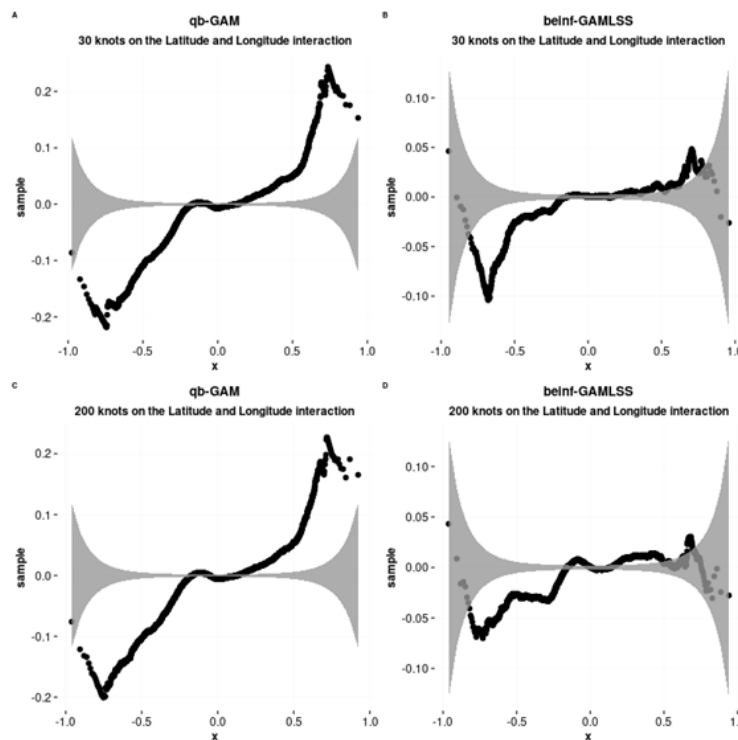
18.



19.

20. Fig. 5: Selected beta zero and one inflated Generalized Additive Model for Location Scale and Shape residuals: A = qq-plot of response residuals, B = response residual versus fitted values, C = response residuals versus linear predictor for mu model component, D = response residuals versus linear predictor for nu model component, E = response residuals versus linear predictor for tau model component, F = response residuals versus linear predictor for sigma model component, G = histogram of response residuals, H = qq-plot of z-scores residuals, I = z-scores residuals versus fitted values, J = z-scores residuals versus linear predictor for mu model component, K = z-scores residuals versus linear predictor for nu model component, L = z-scores residuals versus linear predictor for tau model component, M = z-scores residuals versus linear predictor for sigma model component, N = histogram of z-scores residuals

21.



22.

23. Fig. 6: Worm plot for the two models described in the text: (A) for quasi – binomial model (30 knots) and (B) for beta zero and one inflated gamlss model (30 knots). The models with the same parametrization but a lesser amount of knots in the interaction between latitude and longitude are also shown for comparison: (A) for quasi – binomial model (30 knots) and (B) for beta zero and one inflated gamlss model (30 knots).



## WD9: A proposal for FMSY reference points for Baltic sprat SDs 22-32

by Stefanie Haase (Thünen Institute of Baltic Sea Fisheries), Jan Horbowy (MIR) and Olavi Kaljuste (SLU)

### Summary

Reference Point	Value	Rationale
Blim	459 000 t	See explanation at 1.2
Bpa	541 071 t	$Blim * e^{\sigma_{SSB} * 1.645}$ ; $\sigma_{SSB} = 0.1$
MSY Btrigger	541 071 t	Bpa
Fmsy	0.34	Estimated by EqSim
FmsyUpper	0.35	Fp0.5
FmsyLower	0.25	Estimated by EqSim as the F at 95% of the landings of Fmsy
Flim	0.49	Estimated by EqSim as the F with 50% probability of SSB being less than Blim
Fpa	0.58	$Flim * e^{-\sigma_F * 1.645}$ ; $\sigma_F = 0.1$

## 1. FMSY

### 1.1 Choice of S-R relationship for the FMSY simulations

As a first step, it was analyzed which S-R relationship best explained the relationship between SSB and the recruitment of sprat. The analyses revealed that the Beverton and Holt function and segmented regression (segreg) had the highest contribution to the bootstrap model averaging procedure with 45% and 42%, respectively. The Ricker function had a contribution of 13% (Figure 1). Due to the low weight of the Ricker function, this S-R relationship was not included in the further estimation of FMSY. In the following analysis, the combination of the Beverton and Holt and the segmented regression is used.

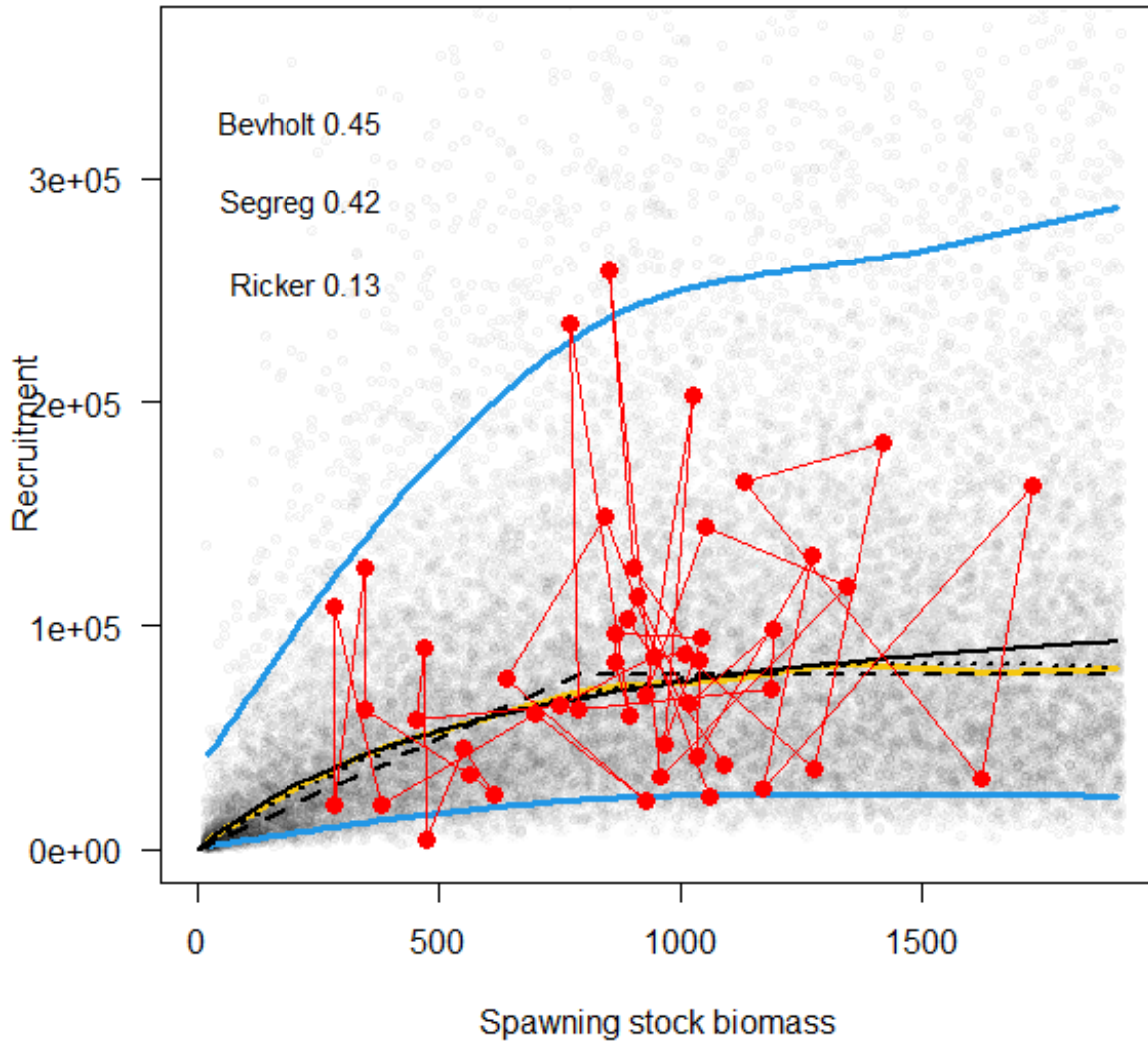


Figure 1. Explored S-R specifications. Spawning stock biomass is shown in k tonnes, Recruitment in Mio numbers. The numbers represent the weights of the different functions in explaining the S-R pattern of sprat.

### 1.2. Estimation of Blim and B0

For Blim estimation a few approaches were considered.

1. Following approach from the last benchmark (ICES, 2013), inter-benchmark assessment (ICES, 2020), and earlier WGBFAS estimations (ICES, 199xx), Blim was defined as the spawning stock biomass which produces 50% of maximal recruitment. Maximal recruitment was estimated using the approach of Horbowy & Luzeńczyk (2012) and Horbowy & Hommik (2022) where equilibrium recruitment ( $R_{eq}$ ), yield ( $Y_{eq}$ ), and spawning stock biomass ( $B_{eq}$ ) at fishing mortality  $F$  may be derived from the equations:

$$R_{eq}(F) = \frac{SPR(F)-a}{b*SPR(F)} \tag{1a}$$

$$Y_{eq}(F) = YPR(F) * R_{eq}(F) = YPR(F) \frac{SPR(F)-a}{b*SPR(F)} \tag{1b}$$

$$B_{eq}(F) = SPR(F) * R_{eq}(F) = \frac{SPR(F)-a}{b} \quad (1c)$$

The SPR and YPR denote stock-per-recruit, and yield-per-recruit, respectively; a and b are parameters of the B&H S-R relationship of the form  $R=B/(a+b*B)$ .

The R0 was estimated at 105 billions and the biomass which produces 50% R0 is 490 kt.

## 1.2. Proposed FMSY, ranges and Fpa reference points

For the FMSY simulations with EqSim the following year-ranges were used for biological parameters (weights, natural mortality) and fishing pattern:

Biological parameters: 2012-2021 (the last 10 years)

Fishing pattern: 2017-2021 (the last 5 years)

Blim was set to 459 000 t

Bpa was set  $Blim * e^{\sigma_{SSB} * 1.645} = 541\,071$  with  $\sigma_{SSB} = 0.1$  based on the last assessment year

Btrigger = Bpa

As described above, the FMSY simulations were run using the combined Beverton and Holt and segmented regression S-R function. When allowing the program to use the full range, and combinations of, bootstrap simulated a and b parameters in the S-R function, the results presented unrealistically high catches at low fishing mortalities (Figure 2). The resulting FMSY values were therefore not considered reliable. The extreme values of the parameters were thus removed and only values within the 5<sup>th</sup> and 95<sup>th</sup> percentile kept. The simulations was rerun with this trimmed set of a and b parameters. A FMSY simulation using the trimmed set of a and b parameters for the Beverton and Holt function resulted in a FMSY of 0.34, with a range of 0.26 - 0.45 (Table 1a and b; Figure 3a and b).

Because Fmsy has been restricted by Fp05,  $FMSY_{upper} = Fmsy$ , thus the range was set to 0.26 – 0.35.

Flim was estimated as F corresponding to 50% probability for  $SSB > B_{lim}$ . This resulted in a Flim of 0.58, which corresponds to Fpa of 0.49 ( $F_{lim} * (\exp(-1.645 * 0.1))$ ).

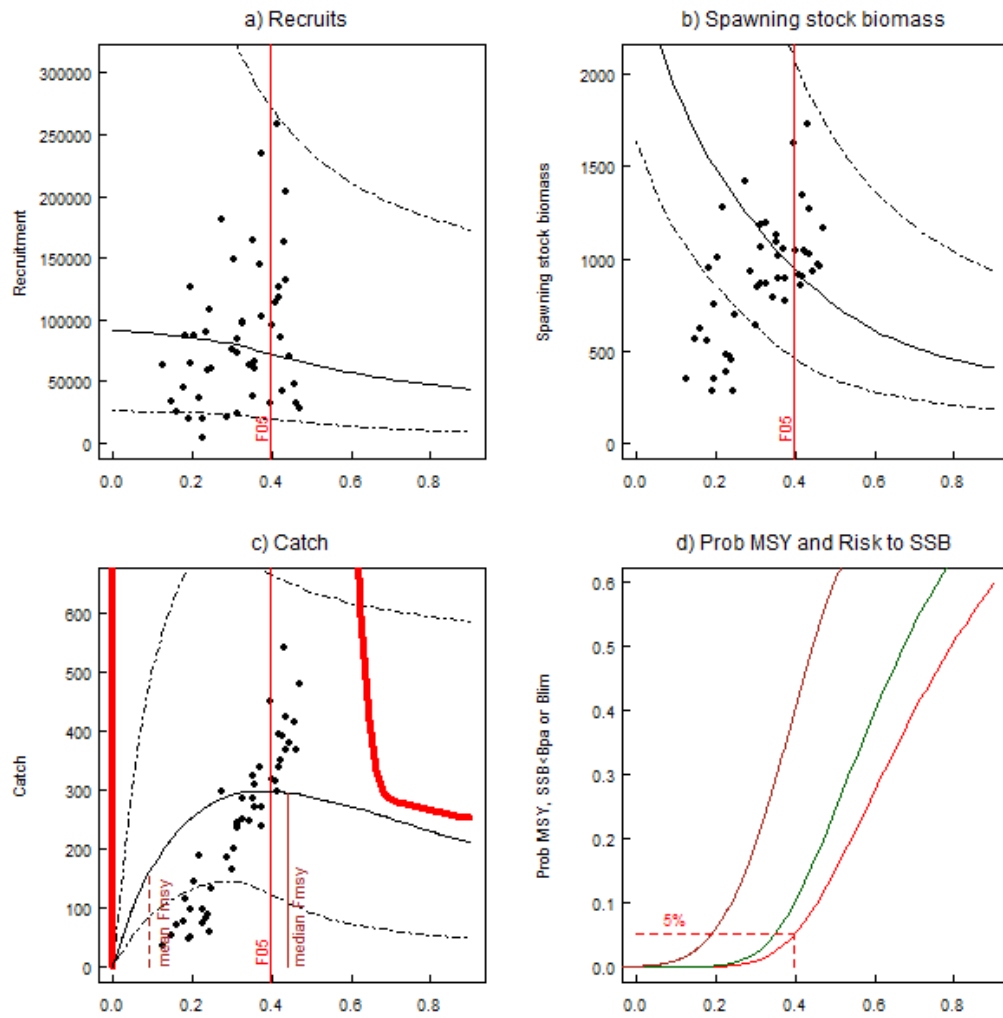


Figure 2. Results of the FMSY simulation using the full range of bootstrap simulated a and b parameters in the Beverton and Holt function.

Table 1a. Results of Fmsy simulations without the advice rule

FmsyMedianC	0.3391960
FmsyLowerMedianC	0.2577889
FmsyUpperMedianC	0.4432161
FmsyMedianL	0.3391960
FmsyLowerMedianL	0.2577889
FmsyUpperMedianL	0.4432161
F5percRiskBlim	0.3451884
Btrigger	0.0000000

Table 1b. Results of Fmsy simulations with the advice rule

FmsyMedianC	0.3798995
FmsyLowerMedianC	0.2713568
FmsyUpperMedianC	0.5291457
FmsyMedianL	0.3798995
FmsyLowerMedianL	0.2713568



FmsyupperMedianL 0.5291457  
 F5percRiskBlm 0.3901007  
 Btrigger 541.0708384

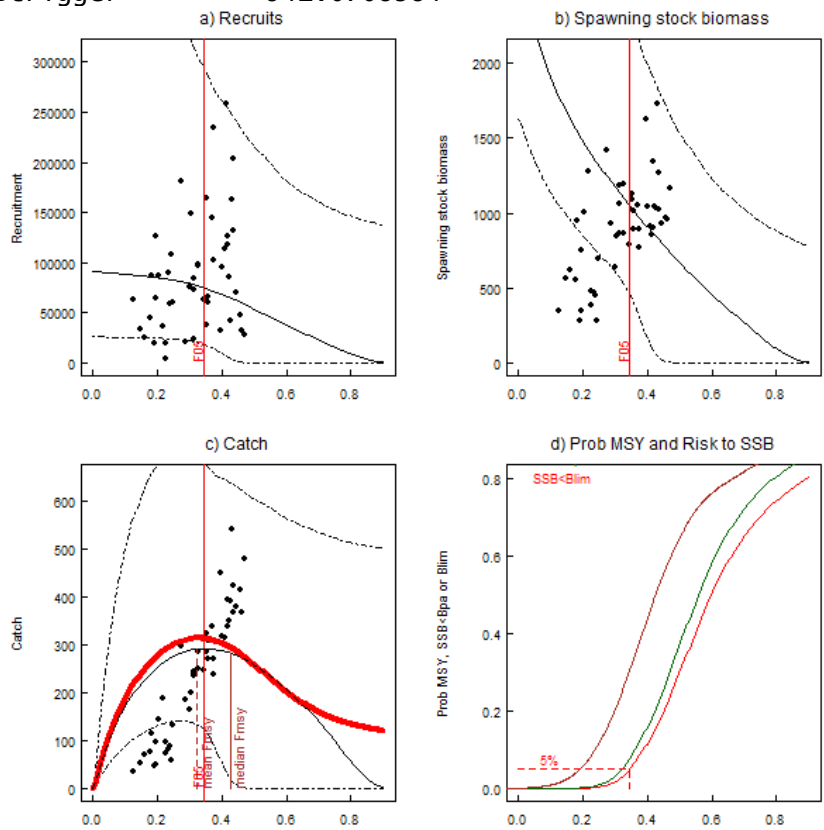


Figure 3a. FMSY simulation without the advice rule.

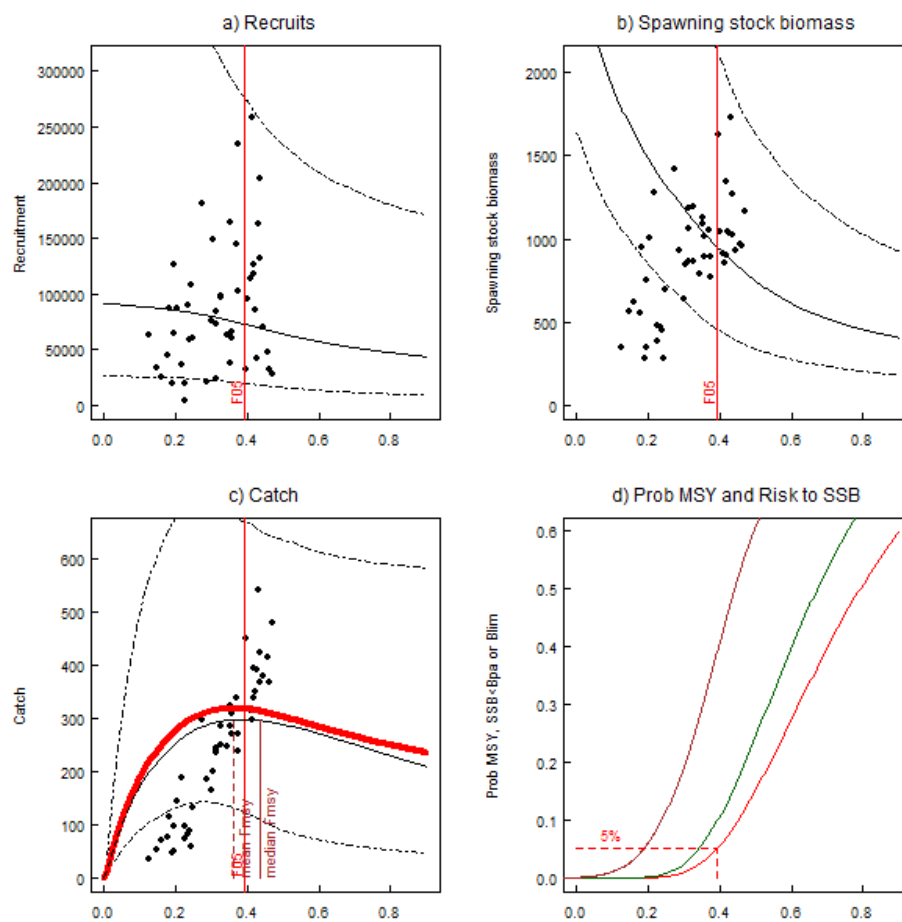


Figure 3b. FMSY simulation with the advice rule.

## References

Horbowy, J., Luzeńczyk, A. 2012. The estimation and robustness of FMSY and alternative fishing mortality reference points associated with high long-term yield. *Can. J. Fish. Aquat. Sci.* 69: 1468–1480

Horbowy, J., & Hommik, K. 2022. Analysis of Fmsy in light of life-history traits—Effects on its proxies and length-based indicators. *Fish and Fisheries*, 00, 1–17. <https://doi.org/10.1111/faf.12640>

## WD-10: Maturity-at-age analysis for Gulf of Riga herring (Subdivision 28.1)

### 1. Introduction

Gulf of Riga herring is a local natural population of Baltic herring (*Clupea harengus*) that occurs mainly in the Gulf of Riga (ICES Subdivision 28.1). It is a slow-growing herring with one of the smallest length and weight at age in the Baltic and thus differs considerably from the neighbouring herring stocks in the Baltic Proper (Subdivisions 25-29). The differences in otolith structure serve as a basis for discrimination of Baltic herring populations (ICES, 2005).

#### 1.1 Maturity Ogive

A constant maturity ogive has been used for the whole time series (1977-2021). It has been assumed that the Gulf of Riga herring starts to spawn at age of 2, when 93% of the fish is mature and by the age of 5 it is considered that all fish are mature. There has been no special survey directed to determine the proportion of mature fish.

Age	1	2	3	4	5	6	7	8+
Proportion mature	0	0.93	0.98	0.98	1	1	1	1

#### 1.2 Purpose

The purpose of this Working Document is to provide an updated maturity ogive for Gulf of Riga herring. Maturity estimates have been studied separately by country and as combined together.

## 2. Material and methods

### 2.1 Available data

New maturity ogives were calculated using maturity data collected from commercial trawls in months January-April. Raw sample data from commercial trawls was available starting from 1995 for Estonia. To conform with Estonian data availability, Latvia also provided its sampling data for the same time period.

R code for maturity ogive estimation for Gulf of Riga herring was adopted from scripts used for haddock in Subarea4, Division 6.a and Subdivision 20 (ICES, 2022).

### 2.2. Data manipulation

Maturity stages were reassigned to a binary response variable: immature (0) and mature (1) (Table 1). Samples age  $\geq 8$  were considered a plus group (8+) and given the age value of 8. Total number of samples per country, year, and age are tabulated in Appendix (Tables 4-5). Individuals marked as spent (6 in national code) in Estonian dataset were removed. This was done, as these markings could be an error or in some cases it could have been most probably the autumn spawning individuals not the spring spawning population individuals.

**Table 1.** Reassignment of maturity stages from national scale to binary response variable.

National code	Description	New code
1	Juvenile/immature	0
2	Resting	0
3	Maturing	1
4	Maturing (very close to spawning)	1
5	Spawning	1

---

6	Spent	1
---	-------	---

---

Estonia does not use length stratified sampling collection procedure. However, in some years, Latvia has taken length stratified samples. To accommodate this, raising procedure was done for those years. After that raw Estonian data was combined with raised Latvian data. No area weighting was used in the calculation.

### 2.3 Maturity ogive estimation

As standard practice WKMOG (ICES, 2008), the maturity ogive was produced by modelling maturity data as a binomial GLM with a logit link:

$$\text{logit}(M) = \log(M / 1 - M)$$

Where  $M$  is the probability of being mature.

With this log transformation, a linear model was applied:

$$\text{logit}(M) = \alpha + \beta X$$

Where  $\alpha$  is the intercept,  $\beta$  is the slope, and  $X$  being the variable(s) of interest. In this instance, age and year, alongside their interactions were included in the full model. Year was treated as factor.

A50 (age-at-50%-maturity), the midpoint of the modelled ogive, was used as an indicator for time-related changes in maturation and was produced as:

$$A50 = -\alpha / \beta$$

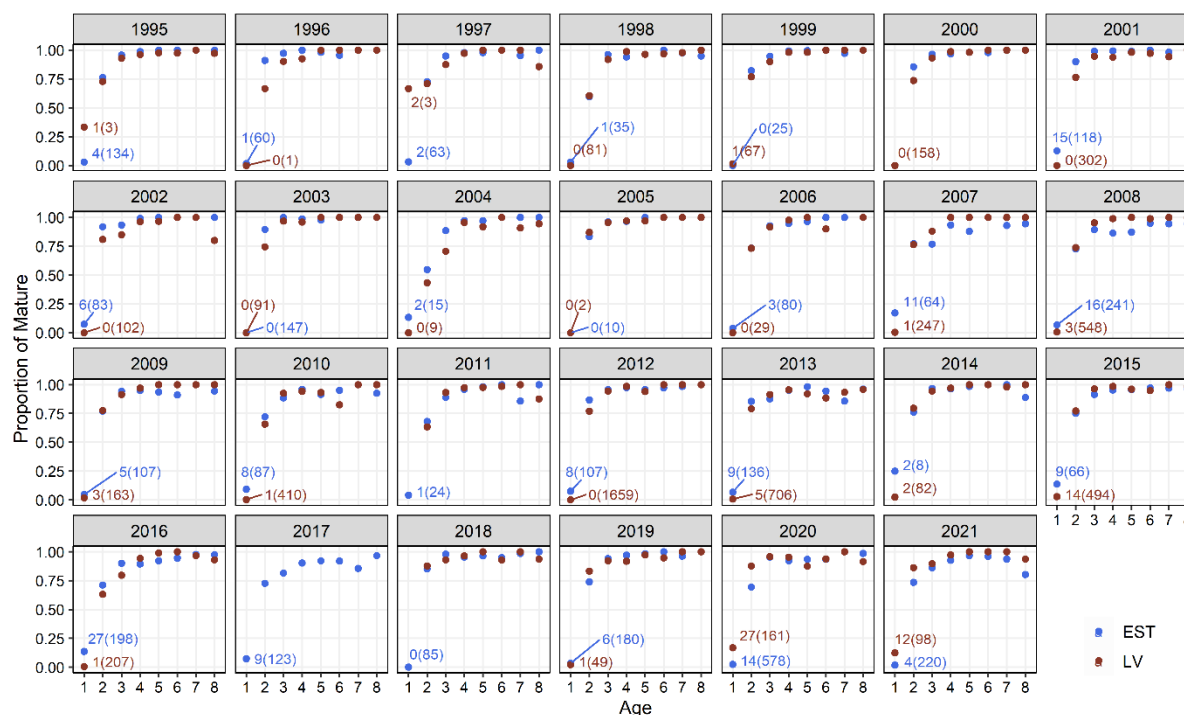
Maturity ogives were produced as predictions from the fitted models.

### 2.4 Smoothing of time series

To reduce the effect of interannual variability, the raw estimated time series was smoothed for age classes 1-4, as it was assumed that from age 5 all individuals are mature. The function *gam* (R package: 'gam'; Hastie, 2020) was used to fit a generalised additive model separately for each age class. Smoothing spline was applied using Gaussian error and respective degrees of freedom (hereby: 'df'), where df=1 implies a linear fit. Degrees of freedom were selected by minimising AIC.

### 3. Results

Proportion mature of fish ages 1-8+ by year and country are shown in Figure 1. For 2017 only Estonian data is shown, as Latvian data was deemed erroneous for this year. Proportion of mature individuals by age classes is in good accordance between Estonia and Latvia. This is expected, as both Latvian and Estonian trawls are fishing in same region, with little spatial difference in main trawling locations. As seen in Appendix Tables 4 -5 and Figure 1, in some years numbers of age 1 individuals is very low. It is more pronounced in Latvian data, that in samples there are no mature individuals in age class 1.



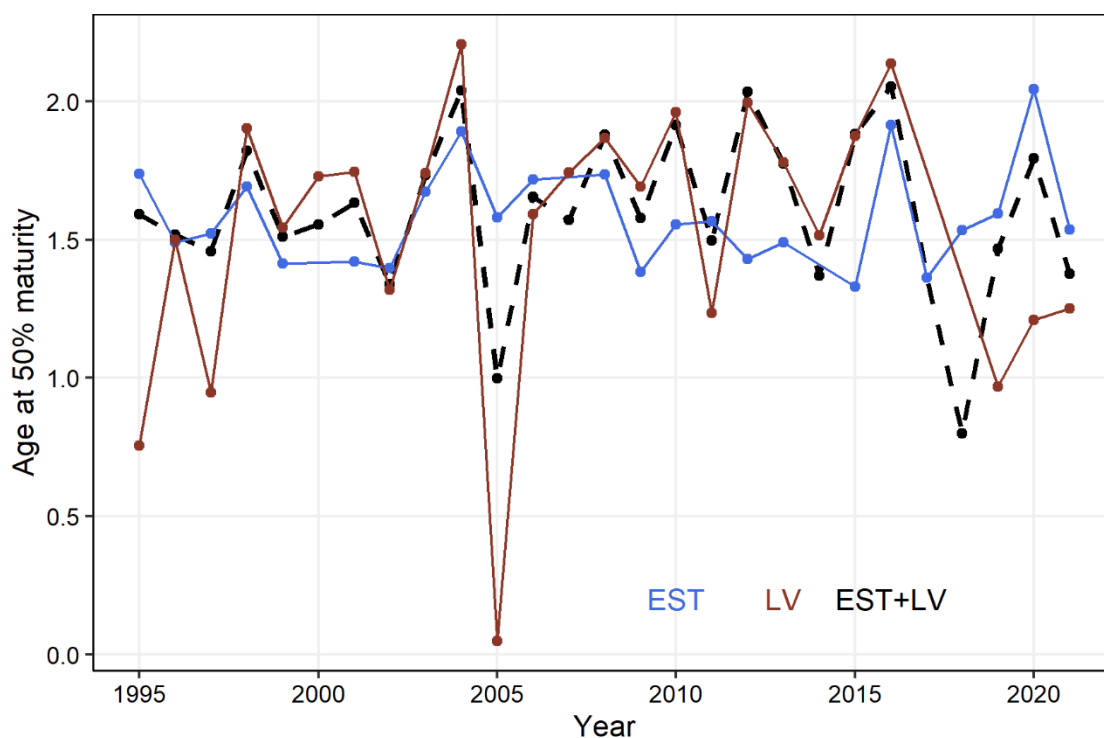
**Figure 1.** Proportion mature of fish ages 1-8+ (8+ represents a plus group) by year and country (blue dots is based on Estonian data, red dots based on Latvian data). Numbers shown in plot are the number of age 1 individuals who are mature, and in brackets is the total amount of age 1 individuals.

In first stage, GLM model with age and year effect with all interactions was fitted separately to Estonian and Latvian data. Years 2017 and 2018 were removed from Latvian data, as in 2017 the data was erroneous and in 2018 there were no individuals aged 1 in the dataset. For Estonian data years 2000, 2007 and 2014 were removed, as these caused problems with model fitting. In second step, Estonian and Latvian data was combined and fitted to GLM model. Age at 50% maturity (a50) estimates by country and with combined data set is shown in Table 2 and Figure 2.

The final model used to calculate an updated maturity ogive for Gulf of Riga herring uses age and year effect and all interactions and is based on Latvian and Estonian combined data (all years included) (Table 3). In Figure 3 final GLM models fits to raw data are shown. In Figure 4 and 5, predicted estimates of proportion mature of ages 1-8+ are shown.

The predicted maturity ogive estimates display high annual variability. To reduce the effect of interannual variability, smoothed values were estimated for ages 1-4 (from age 5 it is assumed full maturity) (Figure 6). Tabulated smoothed values of each age class are listed in Appendix (Table 7).

In Figure 7 comparison in SSB estimates are shown using old maturity ogive versus new estimated smoothed maturity ogive. Overall, the SSB estimates are similar, with new maturity ogive the SSB estimates are smoother. The SSB comparison is based on XSA analysis.



**Figure 2.** Age at 50% maturity (a50) of the Gulf of Riga herring based on Estonian (blue line) and Latvian (red line) data separately, and with combined data (black line).

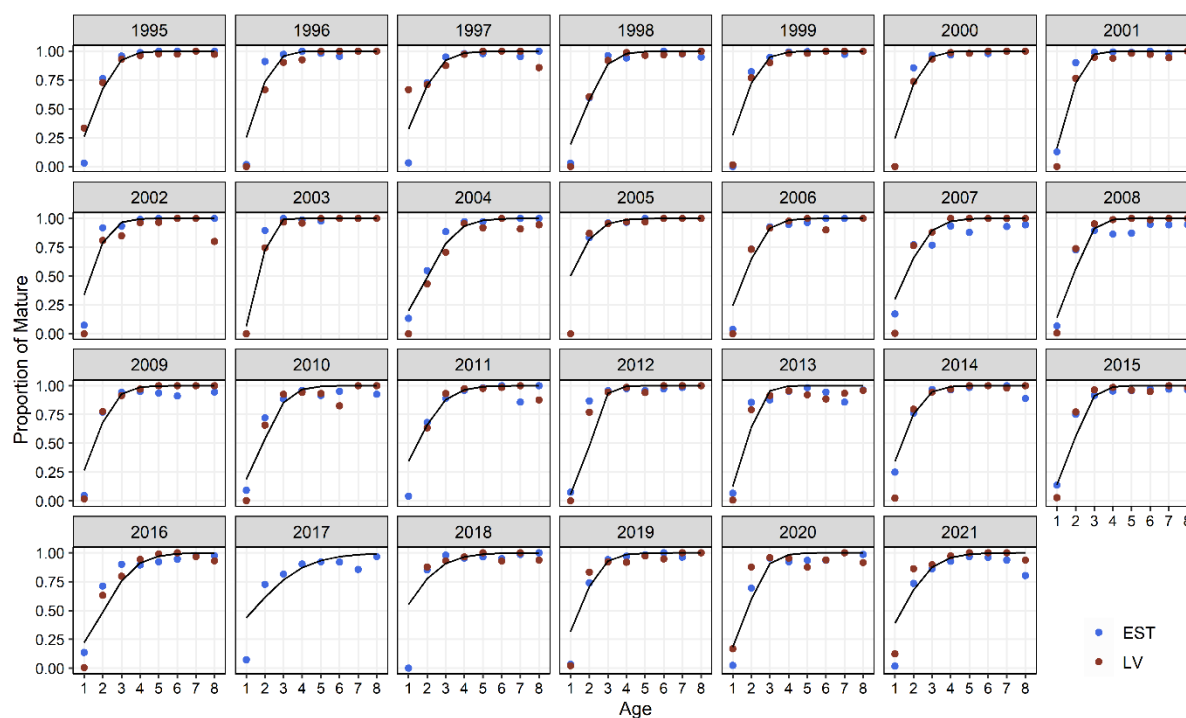
**Table 2.** Age at 50% maturity (a50) estimates of the Gulf of Riga herring by country and by combined Estonian and Latvian data.

Year	a50	a50_LV	a50_EST
1995	1.592981	0.75534154	1.73787
1996	1.518069	1.49831345	1.488927
1997	1.458596	0.9479354	1.521878
1998	1.821223	1.90233047	1.692097
1999	1.511382	1.5432661	1.412584
2000	1.554541	1.72787858	
2001	1.634378	1.74552963	1.42145
2002	1.339891	1.31873601	1.398354
2003	1.734389	1.7392529	1.672813
2004	2.04027	2.20652876	1.890098
2005	0.998311	0.04796103	1.581345
2006	1.655106	1.59204163	1.716271
2007	1.571912	1.74367882	
2008	1.878968	1.86848161	1.73494
2009	1.577738	1.69084999	1.382976
2010	1.913101	1.96130477	1.554536
2011	1.496105	1.23475073	1.567473
2012	2.035238	1.99553435	1.430045

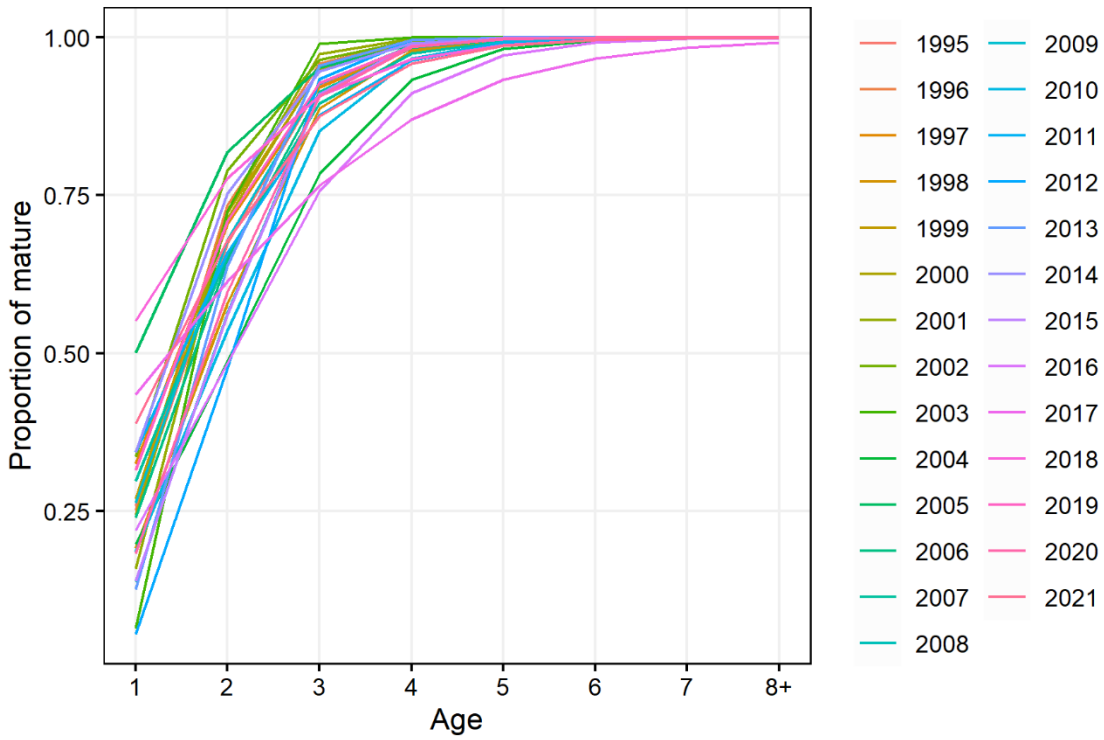
2013	1.773979	1.78047681	1.489789
2014	1.368466	1.51449757	
2015	1.881513	1.874658	1.330681
2016	2.054392	2.13716845	1.915168
2017	1.362622		1.362622
2018	0.799324		1.533998
2019	1.466055	0.96897849	1.594972
2020	1.79292	1.21021224	2.044851
2021	1.375754	1.25216408	1.535736

**Table 3.** 2-way analysis of variance (anova) of the glm fit of maturity data with combined Estonia and Latvian data with an age and year effect alongside their interactions.

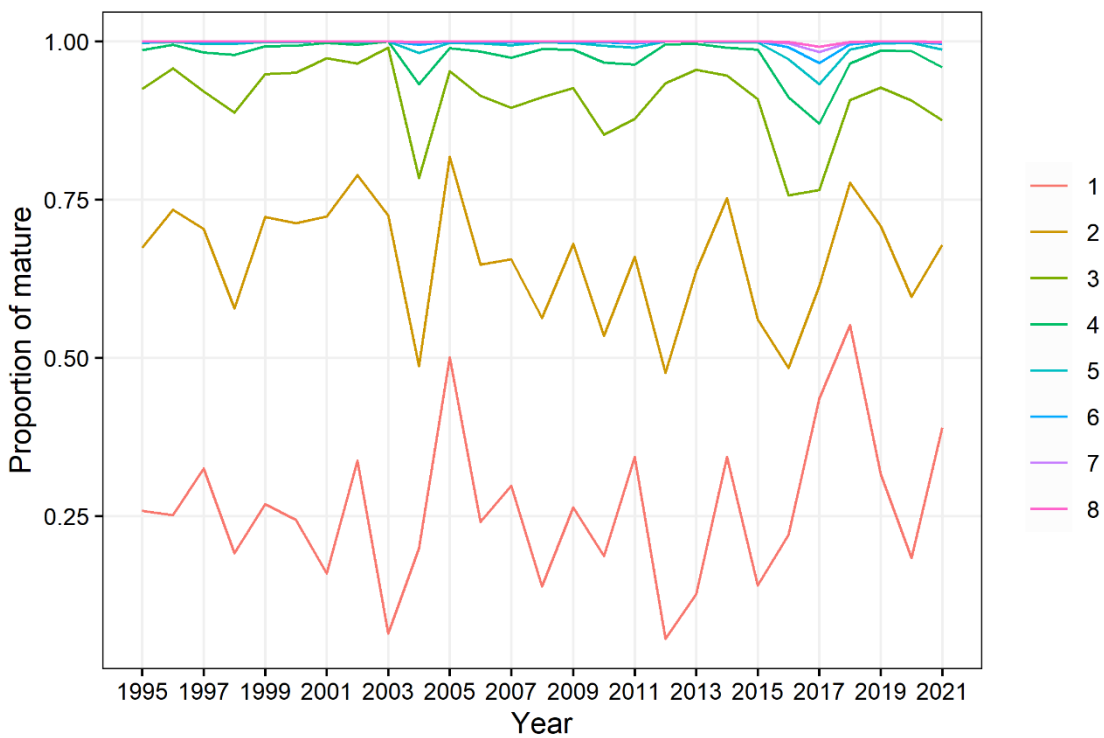
	Df	Sum sq.	Mean sq.	F value	p value
<b>Age</b>	1	2620	2620.1	20493.22	< 0.05
<b>Year</b>	26	375	14.4	112.96	< 0.05
<b>Age x Year</b>	26	264	10.2	79.56	< 0.05



**Figure 3.** Proportion mature of fish ages 1-8+ (8+ is considers as plus group) in Gulf of Riga. Points (blue=EST, red=LV) represent raw estimates and line represents glm model estimates.

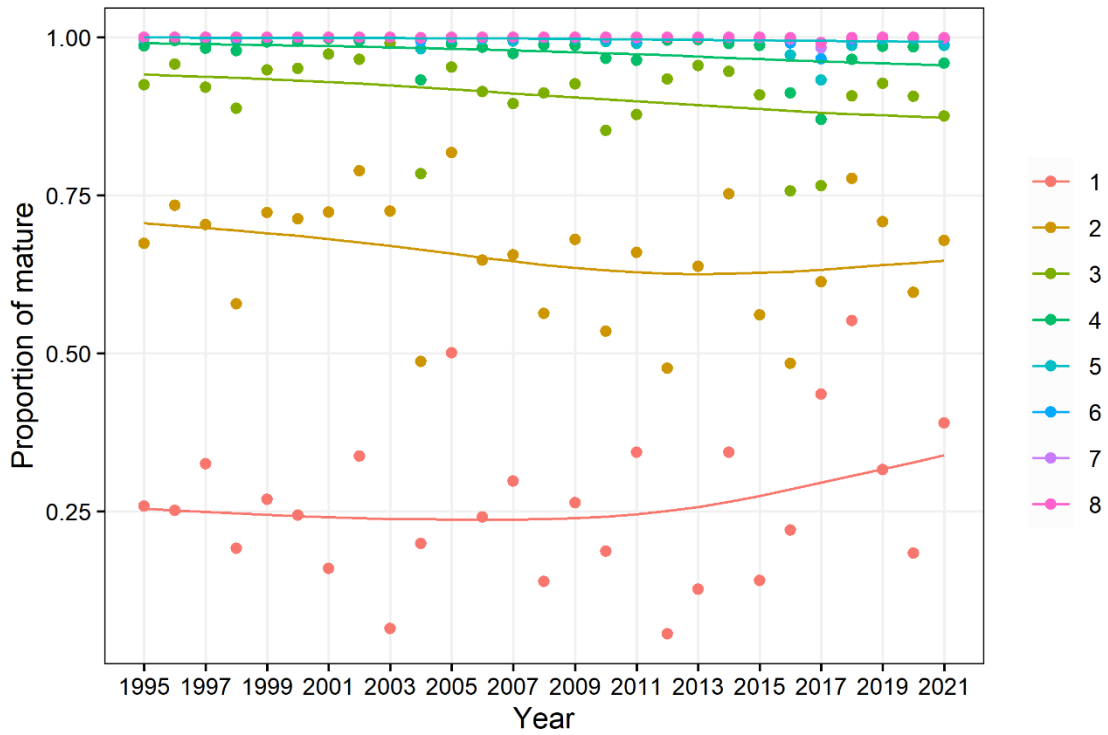


**Figure 4.** Proportion mature of fish ages 1-8+ (8+ represents a plus group) by year in the Gulf of Riga (Estonian and Latvian data combined).

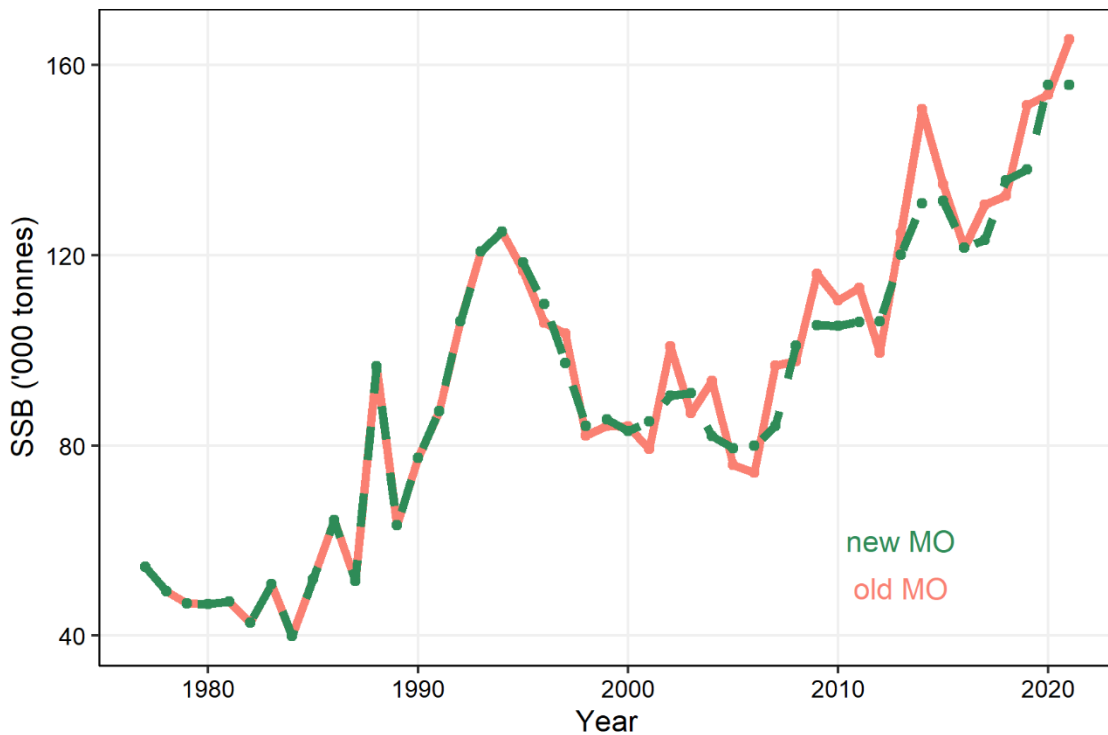


**Figure 5.** Proportion mature of fish ages 1-8+ (8+ represents a plus group) in the Gulf of Riga (Estonian and Latvian data combined).





**Figure 6.** Proportion mature of fish ages 1-8+ (8+ represents a plus group) in the Gulf of Riga (Estonian and Latvian data combined). Dots represent raw estimates with lines representing the smoothed values. Age=>5 is assumed fully mature.



**Figure 7.** SSB estimates from XSA analysis using old maturity ogive (pink) and new estimated maturity ogive (green dashed line).

#### 4. Discussion

This document highlights that there is higher variability in maturity ogive than previously considered. Even though the analysis results showed that adopting the new smoother maturity ogive estimates for years 1995-2021 does not affect very strongly the SSB estimates compared to the old maturity ogive. Still, the estimated new maturity ogive produces smoother SSB estimates, and that there is seen slight trend that more age 1 individuals obtain maturity in the latest years. Moving to time-varying, data derived maturity ogive could give more accurate depiction of SSB changes over times, and ability to spot changes in maturity ogive in real time.

Raw estimates of proportion mature in ages 1-8+ match between Estonia and Latvia. When comparing the estimates of age at 50% maturity by country the differences between countries can be observed but no clear trend. However, these differences are probably mostly due to the effect of small sample sizes for age 1 (Tables 4-5). On average, the age at 50% maturity is close to 1.5 years, and no clear increasing or decreasing trend is seen for time period 1995-2021.

When combining Estonian and Latvian data no country specific weighting of the data was implemented. In some years the sample sizes for age 1 was rather low, and combination of two data sets made the estimation partly better compared to country specific estimates (e.g.,  $a_{50}$  estimates, Table 2). In Appendix Table 6 country sample proportions by age class for ages 1-4 are shown. On average, Estonian data contribution by age class varies from 35-55% compared to Estonian commercial catches contribution which varies from 35-45%. For both Latvia and Estonia, the number of samples taken from commercial trawl fishery is dependent on the fishing intensity. Hence, the total amount of samples taken by country should rather well correspond to the proportions of catch taken by country.

There are no data derived maturity estimates from 1977-1994, hence for this period it is suggested to continue using the previously agreed on maturity ogive.

#### 5. References

xxx

## 6. Appendix

**Table 4.** Number of herring individuals sampled by Estonia from 1995-2021. Age 8+ represents a plus group.

Year	Age1	Age2	Age3	Age4	Age5	Age6	Age7	Age8+
1995	134	216	348	313	241	71	30	41
1996	60	136	112	70	55	44	10	6
1997	63	633	444	237	138	94	42	19
1998	35	166	260	188	84	61	40	20
1999	25	156	76	152	225	51	35	14
2000		282	252	93	330	95	13	33
2001	118	280	359	293	96	130	62	42
2002	83	602	151	121	56	33	29	21
2003	147	115	261	77	43	19	11	20
2004	15	260	61	110	32	15	5	6
2005	10	30	102	29	18	7	1	3
2006	80	184	55	165	56	16	8	11
2007	64	427	194	30	124	47	14	18
2008	241	295	615	124	31	153	54	37
2009	107	540	101	330	77	11	101	35
2010	87	150	249	24	128	20	2	40
2011	24	203	169	213	54	110	7	22
2012	107	135	338	262	366	34	119	33
2013	136	234	40	118	54	107	7	51
2014	8	217	182	55	64	30	54	27
2015	66	88	414	352	47	72	67	79
2016	198	159	60	313	217	36	46	82
2017	123	395	240	31	233	164	14	90
2018	85	148	199	129	28	98	68	32
2019	180	869	336	374	123	14	108	69
2020	578	200	258	103	91	92	9	68
2021	220	730	194	273	122	77	32	51

**Table 5.** Number of herring individuals sampled by Latvia from 1995-2021. Age 8+ represents a plus group.

Year	Age1	Age2	Age3	Age4	Age5	Age6	Age7	Age8+
1995	3	246	256	343	169	120	23	36
1996	1	72	41	27	33	15	8	3
1997	3	186	88	37	27	27	11	7
1998	81	152	336	154	81	62	49	44
1999	67	456	129	208	109	47	28	38
2000	158	575	291	83	99	31	11	26
2001	302	520	287	165	50	70	35	41
2002	102	863	224	102	57	10	20	10
2003	91	258	315	72	30	16	3	12
2004	9	266	123	223	37	19	11	18
2005	2	130	447	32	64	15	10	20
2006	29	197	24	127	12	10		1
2007	247	680	118	14	53	7	7	1
2008	548	343	634	176	15	94	5	25
2009	163	397	92	232	43	7	28	16
2010	410	431	444	119	205	40	8	36
2011		82	129	150	42	66	17	8
2012	1659	208	251	131	131	31	72	29
2013	706	464	47	44	25	26	15	24
2014	82	697	440	33	116	59	51	43
2015	494	166	535	240	25	39	29	67
2016	207	190	79	251	184	30	30	72
2017	38	562	155	51	144	87	17	68
2018		311	273	111	37	111	55	32
2019	49	476	155	146	72	38	60	38
2020	161	294	332	132	64	33	22	47
2021	98	491	164	183	75	42	12	32

**Table 6.** Country sample data proportions by age group for ages 1-4. EST – Estonia, LV – Latvia.

Year	Age 1		Age 2		Age 3		Age 4	
	EST	LV	EST	LV	EST	LV	EST	LV
1995	0.98	0.02	0.47	0.53	0.58	0.42	0.48	0.52
1996	0.98	0.02	0.65	0.35	0.73	0.27	0.72	0.28
1997	0.95	0.05	0.77	0.23	0.83	0.17	0.86	0.14
1998	0.30	0.70	0.52	0.48	0.44	0.56	0.55	0.45
1999	0.27	0.73	0.25	0.75	0.37	0.63	0.42	0.58
2000	0.00	1.00	0.33	0.67	0.46	0.54	0.53	0.47
2001	0.28	0.72	0.35	0.65	0.56	0.44	0.64	0.36
2002	0.45	0.55	0.41	0.59	0.40	0.60	0.54	0.46
2003	0.62	0.38	0.31	0.69	0.45	0.55	0.52	0.48
2004	0.63	0.38	0.49	0.51	0.33	0.67	0.33	0.67
2005	0.83	0.17	0.19	0.81	0.19	0.81	0.48	0.52
2006	0.73	0.27	0.48	0.52	0.70	0.30	0.57	0.43
2007	0.21	0.79	0.39	0.61	0.62	0.38	0.68	0.32
2008	0.31	0.69	0.46	0.54	0.49	0.51	0.41	0.59

---

2009	0.40	0.60	0.58	0.42	0.52	0.48	0.59	0.41
2010	0.18	0.82	0.26	0.74	0.36	0.64	0.17	0.83
2011	1.00	0.00	0.71	0.29	0.57	0.43	0.59	0.41
2012	0.06	0.94	0.39	0.61	0.57	0.43	0.67	0.33
2013	0.16	0.84	0.34	0.66	0.46	0.54	0.73	0.27
2014	0.09	0.91	0.24	0.76	0.29	0.71	0.63	0.38
2015	0.12	0.88	0.35	0.65	0.44	0.56	0.59	0.41
2016	0.49	0.51	0.46	0.54	0.43	0.57	0.55	0.45
2017	0.76	0.24	0.41	0.59	0.61	0.39	0.38	0.62
2018	1.00	0.00	0.32	0.68	0.42	0.58	0.54	0.46
2019	0.79	0.21	0.65	0.35	0.68	0.32	0.72	0.28
2020	0.78	0.22	0.40	0.60	0.44	0.56	0.44	0.56
2021	0.69	0.31	0.60	0.40	0.54	0.46	0.60	0.40
<b>Mean</b>	0.34	0.66	0.45	0.55	0.49	0.51	0.56	0.44

---

**Table 7.** Proportion mature at age smoothed estimates for Gulf of Riga herring between 1995-2021.

<b>Year</b>	<b>Age1</b>	<b>Age2</b>	<b>Age3</b>	<b>Age4</b>	<b>Age5</b>	<b>Age6</b>	<b>Age7</b>	<b>Age8+</b>
1995	0.254	0.706	0.941	0.991	1	1	1	1
1996	0.251	0.702	0.939	0.990	1	1	1	1
1997	0.249	0.698	0.938	0.990	1	1	1	1
1998	0.246	0.694	0.936	0.989	1	1	1	1
1999	0.244	0.690	0.934	0.988	1	1	1	1
2000	0.242	0.686	0.932	0.987	1	1	1	1
2001	0.240	0.681	0.930	0.986	1	1	1	1
2002	0.239	0.676	0.927	0.985	1	1	1	1
2003	0.238	0.670	0.924	0.984	1	1	1	1
2004	0.237	0.664	0.921	0.983	1	1	1	1
2005	0.237	0.658	0.918	0.982	1	1	1	1
2006	0.237	0.651	0.915	0.981	1	1	1	1
2007	0.237	0.645	0.911	0.979	1	1	1	1
2008	0.238	0.640	0.908	0.978	1	1	1	1
2009	0.239	0.635	0.905	0.977	1	1	1	1
2010	0.241	0.631	0.902	0.975	1	1	1	1
2011	0.245	0.628	0.899	0.973	1	1	1	1
2012	0.250	0.626	0.896	0.972	1	1	1	1
2013	0.257	0.626	0.893	0.970	1	1	1	1
2014	0.265	0.626	0.890	0.968	1	1	1	1
2015	0.274	0.627	0.886	0.965	1	1	1	1
2016	0.284	0.629	0.883	0.963	1	1	1	1
2017	0.295	0.632	0.881	0.962	1	1	1	1
2018	0.306	0.636	0.879	0.960	1	1	1	1
2019	0.317	0.639	0.877	0.959	1	1	1	1
2020	0.328	0.643	0.875	0.958	1	1	1	1
2021	0.338	0.647	0.873	0.956	1	1	1	1

## WD-11: Analysis of Gulf of Riga herring trap net tuning series

### 1. Introduction

Gulf of Riga herring is a separated population of Baltic herring (*Clupea harengus*) that occurs mainly in the Gulf of Riga (ICES Subdivision 28.1).

Up to 2002 Gulf of Riga herring was considered as part of the whole Baltic herring stock area (SD 25-29,32). The assessment of herring in Subdivision 25-29, 32 (including Gulf of Riga) was mainly used for determination of TAC in the Baltic Sea, and it didn't consider the state of herring stocks in different regions in the Baltic Sea. Separate XSA assessment for GoR herring has been conducted in 1980-1991 and again since 1993. The separate ICES advice has been provided since 1997 and since 2003 a separate TAC value based on Gulf of Riga herring assessment is set for GoR.

#### 1.1 Tuning series

Since 1993 the XSA for the Gulf of Riga herring was tuned using data on the effort (number of trapnets) directed to catch the Gulf of Riga herring in the Estonian and Latvian trapnet fishery and the corresponding abundance (catch in numbers-at-age) of gulf herring in the trapnet catches. The trapnet tuning data series starts from 1980, however in 2008 the trapnet tuning series was shortened to start from 1996 due to positive trend in log-catchability residuals (ref to 2008 benchmark?). It was noted that the trapnet tuning series could be very sensitive to changes in market demand, therefore, the joint Estonian-Latvian hydroacoustic survey was commenced in 1999 in order to obtain additional tuning data. Since 2004 the hydroacoustic survey has been included in the assessment as second tuning series. Specificities of the two tuning series are shown in Table 1.

**Table 1.** Tuning data for Gulf of Riga herring.

Type	Name	Year range	Duration	Age range
Tuning fleet 1	Trapnets	1996-last data year	April-June	2-8
Tuning fleet 2	Acoustics	1999-last data year	End of July/beginning of August	1-8

#### 1.2 Trapnet tuning series

The trapnet tuning series starts from 1980. The effort in the tuning series corresponds to the number of trapnets directed towards Gulf of Riga herring in the trapnet fishery. The whole available trapnet tuning series is shown in Figure 1. The number of trap-nets used in herring fishery (in the tuning series) has been variable up to the beginning of 2000s, however, has shown decreasing trend since 2003, and since 2010 is at 50% lower level. The number of trapnets used as effort in the tuning series has been constant since 2015 (value of 43).

At this stage, there is no clear documentation available on the procedure how the trapnet effort (number of trapnets, directed to catching the Gulf of Riga herring) was calculated when first introduced into the XSA assessment in 1993. The stock was assessed by the same assessor for 24 years, and when in 2018 a new team of people took the assessment over there was no clear perception how these numbers have been calculated. By 2018 the effort number had been constant for 3 years, and the new team worked on the assumption that the number of trapnets used in herring fishery by Estonia and Latvia has been rather stable, hence continuing using the same effort value seemed reasonable.

However, it was noted that the trapnet tuning series needs thorough investigation and would be done when the stock will be benchmarked.

### 1.3 Purpose

Purpose of this Working Paper is to:

- a) Provide overview of the importance of the trapnet tuning series in the Gulf of Riga herring assessment.
- b) Try to investigate and understand how the “effort” in trapnet tuning series was calculated.
- c) Test if the assumption of constant effort since 2015 holds.

## 2. Material and methods

### 2.1 Available data

To investigate into the assumption that the effort has been stable since 2015, detailed information about how the trapnet fishery is conducted is needed. Getting detailed information from the trapnet fishery is difficult for both fishing countries, Estonia and Latvia. The main issue for both countries is that the number of trapnets used in the fisheries and their soaking time is unknown. For Estonia, the detailed information on effort is available starting from 2018. Knowing the number of trapnets and soaking time it is possible to calculate effort as  $\text{gear} \times \text{DAS}$ , where DAS=days at sea.

For Latvia, the digitalised information from coastal fishing logbooks has been available since 1995 however, information on gear count is not always available as this field up to 2021 was occasionally left blank, meaning that precise effort calculation for historical data is not possible. Since 2021 new electronic coastal logbooks have been introduced in Latvia, where information on gear count is mandatory.

### 2.2 Data manipulation

The quality of the Estonian effort data is not uniform between years. There are some presentational differences between 2018-2019 vs 2020-2021. Herring trapnet fishery is conducted mainly only during the spawning time near coastal areas. It is usually that in the beginning of the fishing season the fishermen cast their trapnets to the sea and regularly go check the trapnets and retrieve the catch without transporting the trapnet itself to the shore. In an essence, the trapnets are casted to the sea in the beginning of the season and taken out when the season ends. To be accurate in the effort calculation it is important to know the first time when the trapnets were cast to the sea to start counting the DAS. In years 2018 and 2019 the “coastal” logbooks used by the fishermen in trapnet fishery, did not have the record on when the trapnets were first cast, the records start with the first time when the catch was collected. For years 2020 and 2021 however this information is recorded and available. From the data in years 2020 and 2021 it is seen that sometimes the trapnets are casted rather early and it could be up to 20 days when the first catch is collected. This behaviour is seen in situations when fishermen are fixing their fishing positions before the season, i.e. saving the spot for near future. Hence, the differences how the starting date for the casting of trapnets is set causes differences in the total calculated effort. To make the effort data more uniform and comparable between years, procedure was adopted, where the effort data was recalculated using the assumption that the first casting day equals to the first landing event. This action eliminates the first landing event in all years, making the starting point (end date of first landing event) same for each year.



### 3. Results

#### 3.1 Overview of trapnet tuning series in the assessment

The trapnet tuning series has high weight in the XSA assessment. This can be seen from Figure 2, where the scaled weights per data source are shown for survivor estimates in the XSA model. Trapnets have the highest weight in survivor estimates. Same can be seen from the leave-one-out analysis (Figure 3), where the current assessment estimates for SSB and Fbar are closer to the estimates when assessment is run with only trapnets as tuning series. Assessment with only acoustics tuning series provides considerably higher estimates in SSB and lower estimates in Fbar. Similar trend is also seen in exploratory SAM assessment leave-one-out runs (Figure 4), although the absolute difference in SSB and Fbar estimates are lower in case of SAM compared to XSA. From exploratory SAM assessment, it is also seen that the model trusts trapnet tuning series more compared to acoustics tuning series, and the highest trust goes to catch data (Figure 5).

In addition to the lack of knowledge how the trapnet effort number has been calculated, the tuning series is inherently part of the catch matrix, hence it is expected that the model has more trust in this tuning series. The correlation between trapnet tuning series and commercial catch matrix is strong (Figure 6, Table 2), especially for ages 3-6, while the correlation between commercial catch matrix and acoustics tuning series is lower (Figure 7, Table 2). Consistency between the surveys is at similar level as for commercial catch matrix and the acoustics tuning series (Figure 8, Table 2).

#### 3.2 'Effort' calculation in the trapnet tuning series

It has been defined, that the effort in the trapnet tuning series corresponds to the number of trapnets directed towards Gulf of Riga herring in the trapnet fishery. In addition to catching Gulf of Riga herring stock, the trapnet fishery also exploits open sea herring from the Central Baltic herring (CBH) stock, which constitutes roughly 30-50% of the trapnet catches, varying by years.

Based on this knowledge, it is hypothesised that for the effort calculation, the total number of trapnet used in the herring fishery is adjusted with a coefficient which corresponds to the proportion of gulf herring in trapnet catches. Estimating the total number of trapnets used in the herring trapnet fishery is difficult. The number of licences issued for trapnet usage is known for both Estonia and Latvia, and these values have been stable since the beginning of 2000s. However, the number of issued licenses does not directly correspond to the number of trapnet used, as fishermen can use fewer trapnets than written in their licences. In trapnet fishery, coastal logbooks are used to report the catch. In Latvia, it has not been mandatory for the fishermen to fill out the 'gear count' field in the coastal logbooks, meaning that not all necessary information is available to determine the number of trapnets actually used by Latvian fishermen. In Estonia, 'gear count' field is mandatory, however it still might not always be filled out, and another problem is that until 2018 this information, even if present, was not transported from the paper coastal logbooks to electronic systems/databases. Meaning, that this information is out there but not available for analysis.

Through the investigation of old excel spreadsheets owned by the previous GoR herring stock assessor, remanent partial calculation processes for trapnet effort were discovered. From this, it was deduced that there is an assumption that total number of trapnets used in herring fishery is 120. And this number is adjusted based on the proportion of gulf herring in the trapnet catches. It was also found that the number of 120 as total number of trapnets could potentially correspond to the number

of how many trapnets are licensed to fishermen in Latvia. However, it is still not understood why exactly this number was used, as Latvia has smaller trapnet catches compared to Estonia.

Starting from 2004 the effort in the tuning series has been decreasing, and if assumption of total of 120 trapnets is used, it would lead to assumption that the proportion of gulf herring in trapnet catches has been decreasing, which is false. It might be, that sometime in the beginning of 2000s different total number of trapnets assumption was used, but this is unknown.

At this stage there is no clear understanding how the trapnet effort was calculated for the tuning series. It is not possible to reproduce these estimates or try to adjust the effort values starting from 2015 which have been kept constant.

### 3.3 Effort estimation for Estonian trapnet fishery 2018-2021

The raw and uniformed effort estimates for years 2018-2020 are shown in Figure 9. The trend between years is the same for raw and uniformed effort estimates, while the absolute scale differs. With the uniformed effort estimates comparisons between years can be made.

The effort in 2018 and 2019 has been rather stable, while substantial increase in effort is seen for 2020, and in 2021 effort decreased but was higher compared to 2018 and 2019. The effort in 2020 was 100% and 38% higher compared to years 2019 and 2021, respectively. The trapnet fishery directed to herring can be quite complex and is affected by many factors. It is important to note that the trapnet herring fishery does not catch only the gulf herring, but it is also directed towards the sea herring that also spawns in the same coastal areas as the gulf herring. On average the proportion of sea herring the gulf trapnet catches is 30-35%, in weight. In 2020 the sea herring proportion in the trapnet catches was very low, only 12%, which can be explained by the sharp decline in the Central Baltic herring stock level. The quota for 2020 was close to the quota for 2019, however, with very low open sea herring (CBH), catches only 85.1% of the quota was caught, compared to 2019 when 91.6% of quota was taken. As the quota uptake was slow in 2020, on average the trapnets were deployed longer in water in 2020 compared to 2019. Out of the fishermen that were active both in 2019 and 2020 (n=61), 70.5% of them kept trapnets more than 15 days longer in the water in 2020 compared to 2019. 21% kept then over 30 days longer in 2020 compared to 2019, and only 11.5% of the fishermen kept their trapnets longer in 2019 compared to 2020. It is assumed that the total amount of trapnets used in the fishery since 2015 is rather stable throughout the years, as the number of licensed trapnets has been very stable since 2005 and starting from 2015 the fishermen have individual quotas. However, when comparing number of trapnets used by fishermen in 2020 and 2019, it was noticed that in 2020 more trapnets were used. Even though the number of licenced trapnets is constant, this does not automatically mean that the fishermen have to actively use all their trapnets. In addition, the fishermen are allowed to swap the trapnets between themselves, leading to situation where are fishermen can use more trapnets that are historically licenced to them.

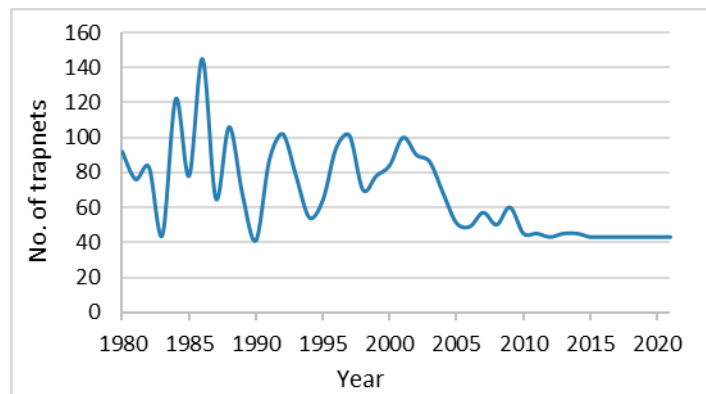
The effort attributed to catching herring with trapnets in the Gulf of Riga is highly dependent on the open sea herring and its stock status. While the proportions of sea herring caught in the gulf is taken into account when the quotas are set, these values represent the overall proportions in herring fishery (trapnets+trawls). The mixing of the two herring stocks is higher in trapnet fishery compared to trawl fishery. In a situation where one stock (GoR herring) is increasing while other stock decreases (CBH), it becomes more difficult to fill the trapnet fishery quotas.

## 4. Discussion

The effort described in the trapnet tuning series corresponds to the trapnet directed towards Gulf of Riga herring fishing. However, the number of trapnets used in the fishery alone does not directly indicate the true effort of the fisheries, as for passive gears the soaking time is an important factor. The comparison of fishing effort in Estonian trapnet fishery in years 2019 and 2020 provides a good example how the difference in soaking time affects the total effort. In addition, the trapnet herring fishery is highly dependent on the open sea herring, as both open sea and gulf herring are targeted in this fishery. In situations when Central Baltic herring stock level is low and the migration into Gulf of Riga is therefore impaired, the fishermen will increase the effort to fill their herring quotas. The changes in the effort of the herring trapnet fisheries are not directly linked to the changes in GoR herring stock. Neither the number of trapnet used in herring directed trapnet fishery nor the effort (gear\*DAS) show the true effort directed towards Gulf of Riga herring.

After extensive investigation into the possible procedures how the number of trapnets used in the trapnet tuning series was achieved, no clear or transparent description emerged. The 'effort' number in the trapnet tuning series still remains somewhat a "black box", even when some of the steps in this calculation procedure were revealed, the reasoning behind those steps is not fully understood or found plausible.

In the upcoming benchmark meeting different scenarios should be explored where the trapnet tuning series is excluded from the assessment or the time series length of the tuning series is shortened to exclude the most recent years, e.g., starting from 2015 when the assumption of constant effort was introduced or limit the time series to beginning of 2000s, when decrease in number of trapnets is seen but no known reasoning behind it.



**Figure 1.** Effort (number of trapnets) value used in the trapnet tuning series, years 1980-2021.

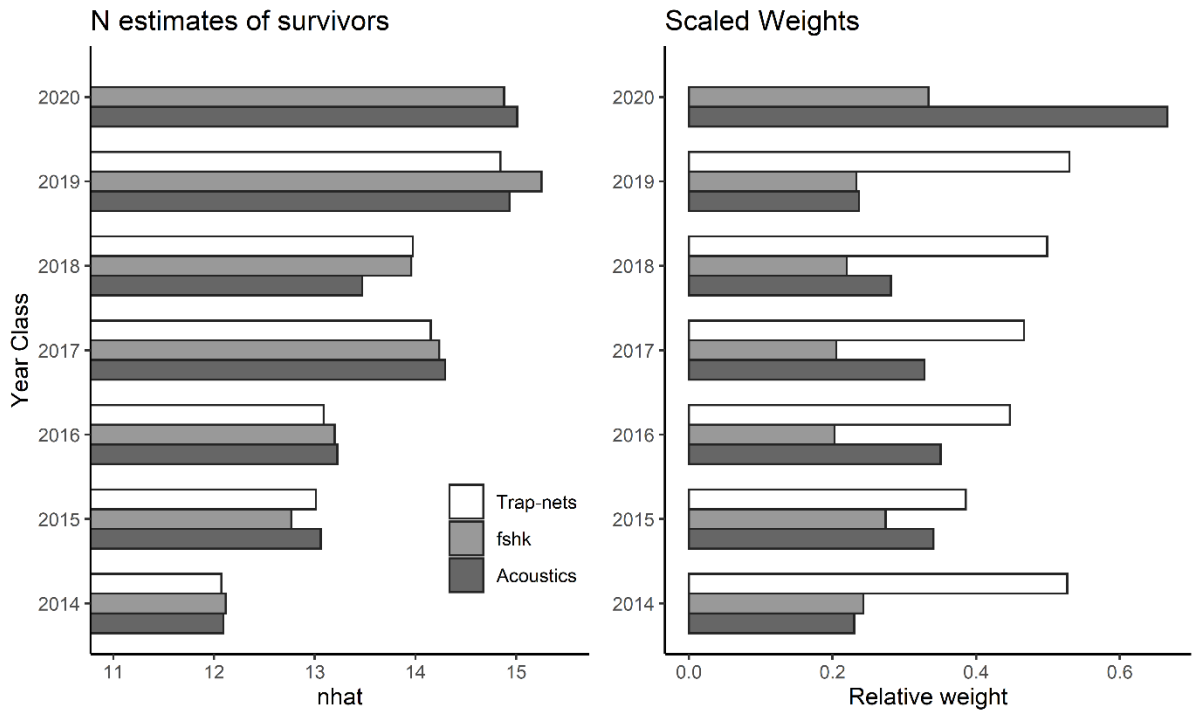


Figure 2. Survivor estimates and scaled weights.

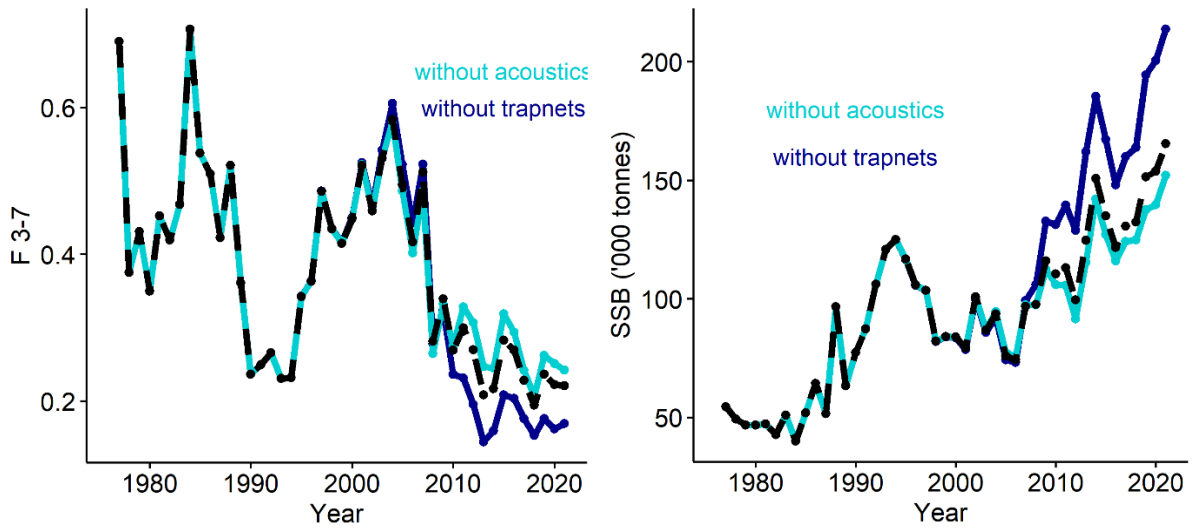
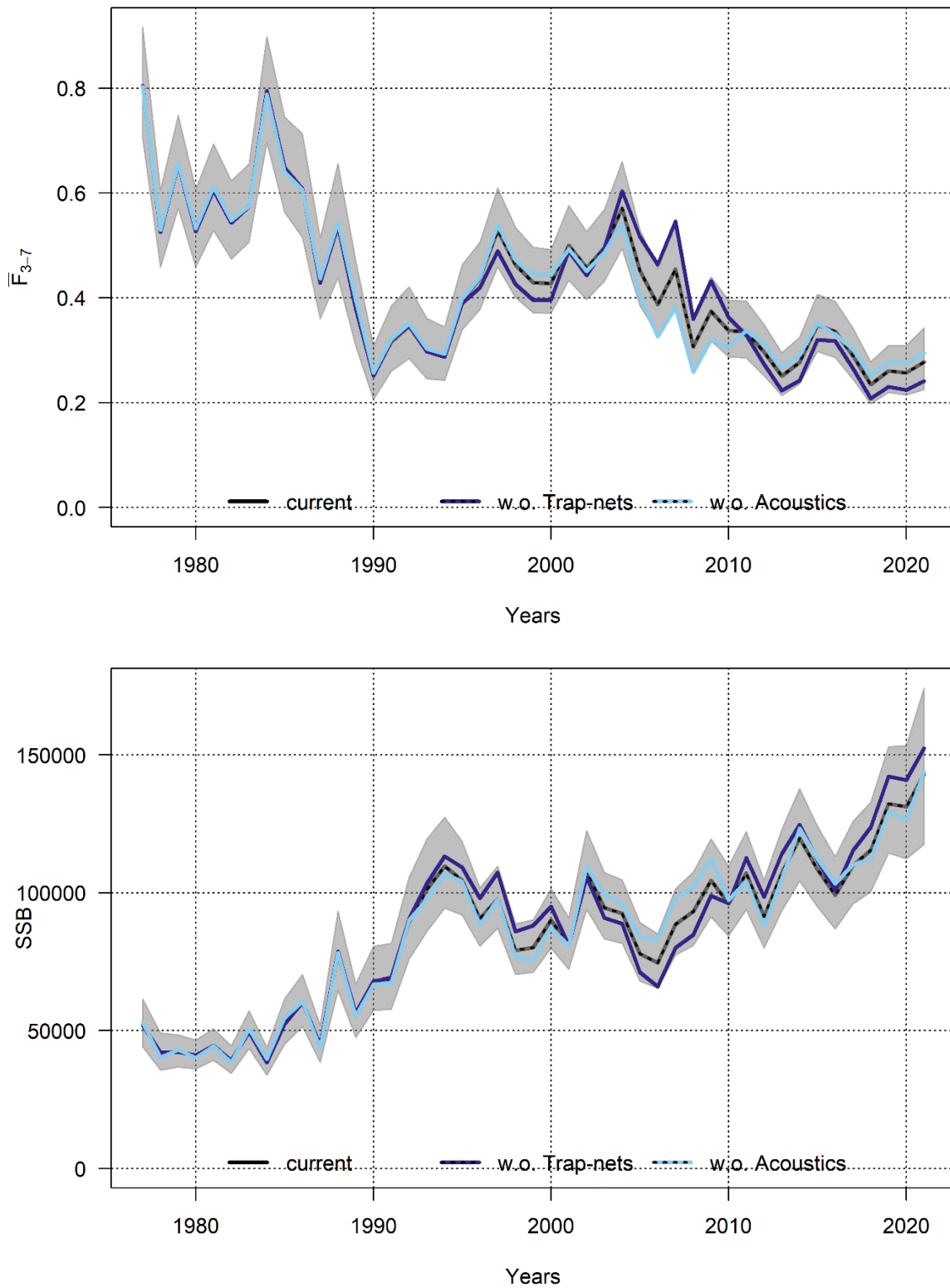
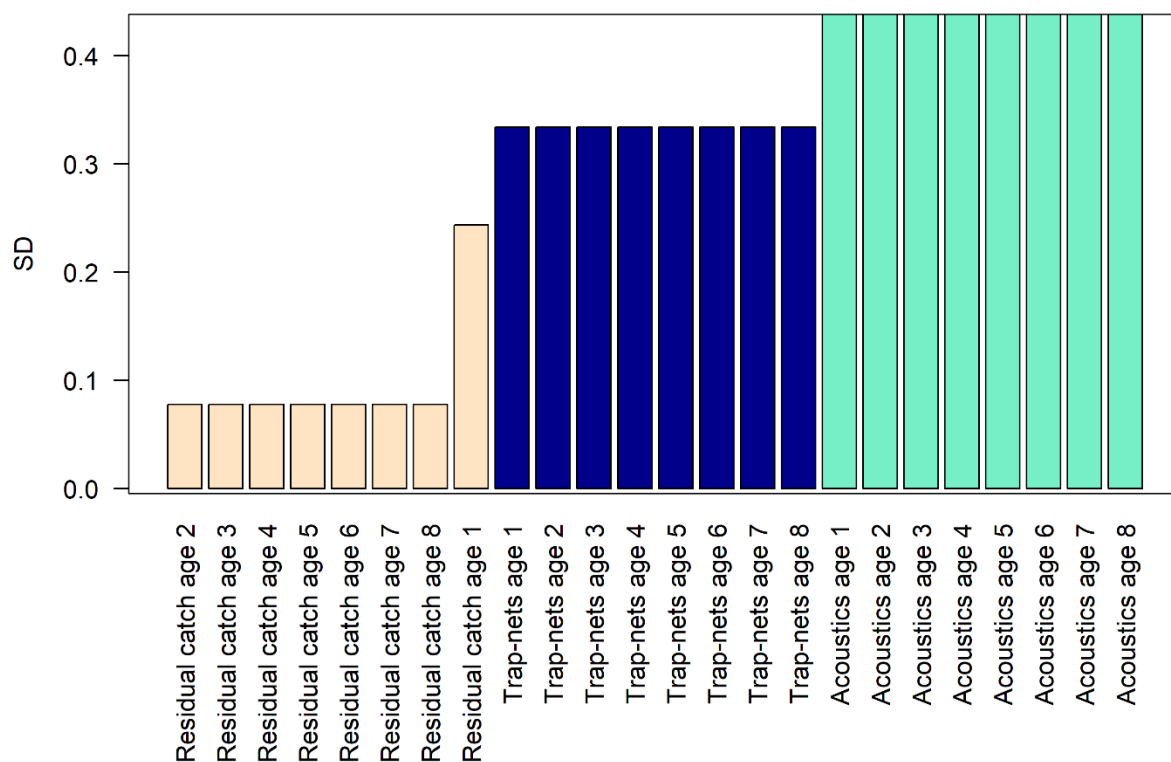


Figure 3. Leave-one-out runs for XSA assessment.



**Figure 4.** Leave-one-out runs from exploratory SAM assessment.



**Figure 5.** Standard deviation plot from exploratory SAM assessment.

**Table 2.** Consistencies by age between surveys and between surveys and commercial catch data. Values in the table are R<sup>2</sup> values.

Age	Trapnet-acoustics	Trapnet-commercial	Acoustics-commercial
1	-		0.56
2	0.41	0.55	0.63
3	0.53	0.73	0.62
4	0.70	0.80	0.60
5	0.55	0.78	0.51
6	0.60	0.77	0.76
7	0.62	0.77	0.72
8(+)	0.21	0.22	0.29

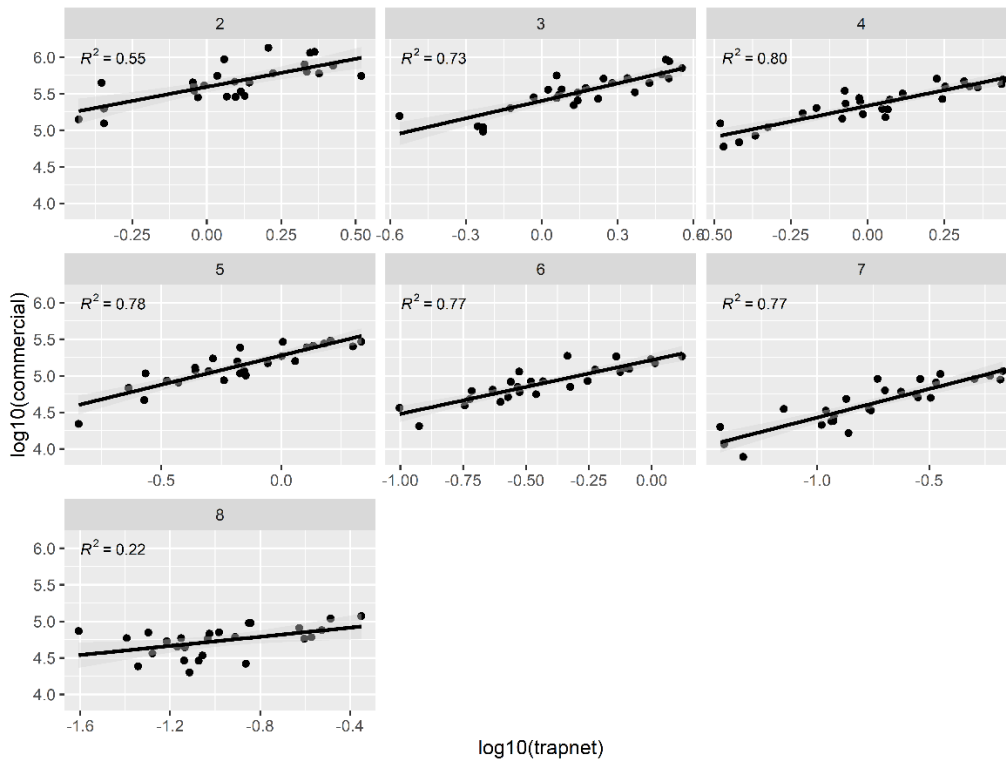


Figure 6. Gulf of Riga herring trapnet survey and commercial data correlation by age.

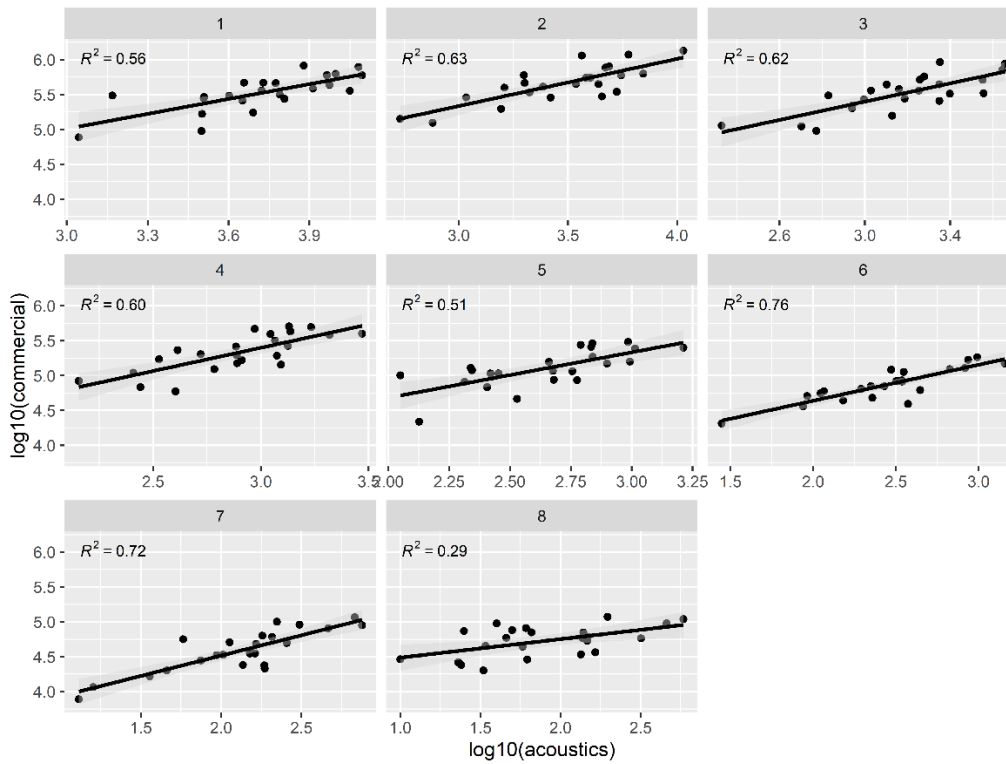


Figure 7. Gulf of Riga herring acoustics survey and commercial data correlation by age.

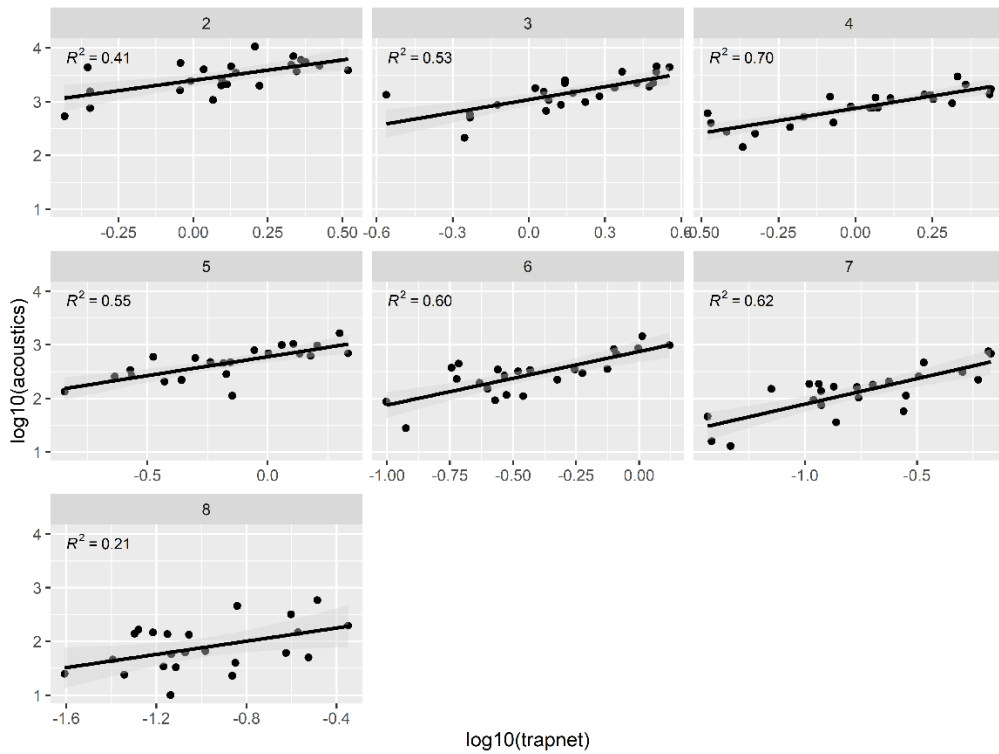


Figure 8. Gulf of Riga herring acoustics survey and commercial trapnet survey correlation by age.

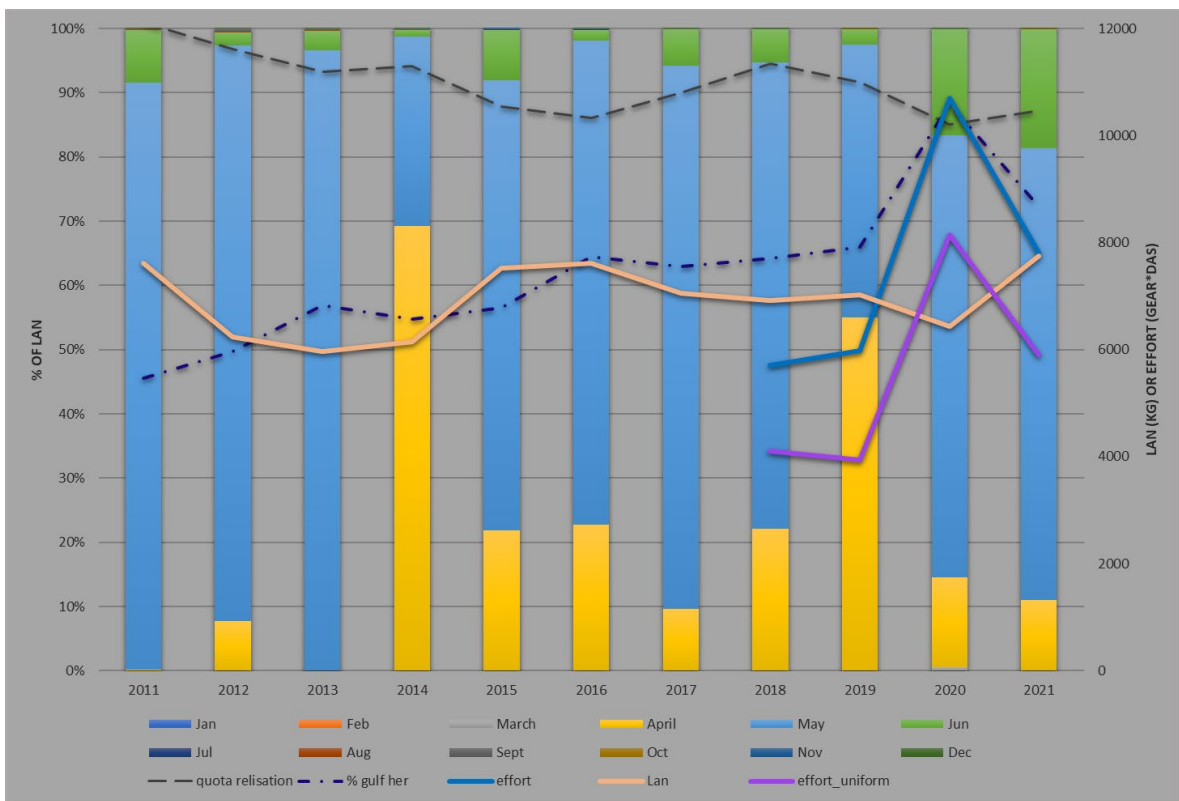


Figure 9. Gulf of Riga herring catches in Estonia with trapnets per month and year. Effort (gear\*das) is shown with dark blue (raw estimates) and purple (uniform estimates) lines. Dark grey dashed line



represents the quota realisation percentage. Blue dashed line represents % of gulf herring in trapnet catches. Note the presence of two different axes.

## **WD-12. Gulf of Riga herring assessment input data and assessment model settings**

### **1. Data**

#### **1.1. CANUM**

There was no specific data call made for this benchmark. Same catch-at-age data is used as previously. Only addition that is made to the catch-at-age data is that the age 0 is included into the catch-at-age matrix which has been previously left out. As there was no specific data call made, the original catch-at-age exchange files of the two fishing countries, Estonia and Latvia, were located and the information was gathered from there of age 0 for years 2003-2021. Age 0 estimates for year 1977-2001 were gathered from old WGBFAS reports (report years 1996-2002). We were not able to locate the data for year 2002 neither from old reports nor old original exchange files. However, we had the information of catch-at-age and weight-at-age for age classes 1-8+ and we were able to calculate the total catch of these age classes (1-8+) and we also had information on the total catch (including all age classes). Subtracting the total catch of age classes 1-8+ from the whole total catch we get rough estimate of how much in weight age 0 constituted in the catch. Average weight of age 0 was calculated based on the average weight of age 0 in years 2001 and 2003. Having the information of weight-at-age and total catch of age 0 in 2002, after dividing the total catch of age 0 with the average weight of age 0 we get the estimate of age 0 in numbers. This approximation was found to be the best solution to get the age 0 estimate for 2002.

Catch-at-age values were corrected for years 2003 and 2008. In 2003 the age 1 numbers were corrected as previously this number also included age 0. For 2008 inconsistencies were found between previously shown catch-at-age values and values that were calculated based on the original exchange files. Hence small changes in catch-at-age for all ages is seen. Updated catch-at-age estimates are in Table A1.

#### **1.2. WECA, WEST**

Same CANUM and WECA that has previously been used. One addition is that age 0 is included to WECA and CANUM. For 2002 age 0 estimates were not available. Catch at age for age 0 was retrieved as average based in years 2001 and 2003. To get weight at age estimate for age 0 the difference of catch was found based on total reported catch and catch estimate based on age classes 1-8+. This catch difference was then divided by the estimated number at age to retrieve weight at age for age 0. Updated weight-at-age estimates are in Table A2. Weight-at-age in stock is assumed same as in catch (WEST=WECA).

#### **1.3. Maturity ogive**

New maturity ogive estimates for years 1995-2021 were used (WD-1). For the older time period (1977-1994) previously used maturity ogive was used. Updated maturity ogive is seen in Table A3.

#### **1.4. Tuning series**

Join Estonian-Latvian hydroacoustic survey abundance estimates are used as fisheries independent tuning series. In the assessment, latest version of WGBIFS approved tuning

series values were used. The main differences in WGBIFS approved and previously used tuning series values are the slight difference in year 2010, 2020 and 2021. In addition, in the updated tuning series age 8 is a plusgroup, while before it was a true age class. Lates tuning series values are in Table A4.

**1.5. Natural mortality**

Constant natural mortality  $M = 0.20$  is used for all the years except for the period 1979–1983 when a value of  $M = 0.25$  is used due to presence of cod in the Gulf of Riga.

**2. SAM configuration**

SAM model was set up with the new updated data and default configuration by the model were taken as the first step. One addition had to be included for the model convergence, the process variance parameters for log(F) process were decoupled for ages 0, 1 and 2+. The initial SAM configuration is seen in Table 1, and is labelled as “GoR\_BP\_base”.

Table 1. Gulf of Riga herring SAM model different runs and configuration settings.

Model run	Parameter	setting	AIC	par
GoR_BP_base	Coupling of F state process for each age \$keyLogFsta	0 1 2 3 4 5 6 7 7 -1 -1 -1 -1 -1 -1 -1 -1 -1	619.92	17
	Correlation of F across ages \$corFlag	2		
	Coupling of the survey catchability \$keyLogFpar	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 0 1 2 3 4 5 6 7		
	Coupling of process variance params for log(F) process \$keyVarF	0 1 2 2 2 2 2 2 2 -1 -1 -1 -1 -1 -1 -1 -1 -1		
	Coupling of R and survival process variance \$keyVarLogN	0 1 2 2 2 2 2 2		
	Coupling of the variance param. for the observations \$keyVarObs	0 0 0 0 0 0 0 0 0 -1 1 1 1 1 1 1 1 1		
	Covariance structure \$obsCorStruct	ID ID		
	\$stockRecruitmentModelCode	0 (plain random walk)		
	\$fbarRange	3-7		
	\$matureModel	0 (MO is used as known)		
GoR_BP_vol1	\$keyVarObs	0 0 0 0 0 0 0 0 0 -1 1 2 2 2 2 2 2 3	602.39	19
GoR_BP_v2	\$obsCorStruct	ID AR	550.0	21
	\$keyCorObs	NA NA NA NA NA NA NA NA -1 0 0 0 0 1 1 1		
GoR_BP_v2.2	\$keyVarF	0 1 2 2 2 3 4 5 5 -1 -1 -1 -1 -1 -1 -1 -1 -1	532.9	24
GoR_BP_v2.2.1q	\$keyLogFpar	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 0 0 0 0 0 0 0 0	543.27	17
GoR_BP_v2.2.multq	\$keyLogFpar	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 0 0 1 1 1 2 3 3	536.23	20

GoR_BP_v2.2.3q	\$keyLogFpar	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 0 1 2 2 2 2 2 2	532.43	19
GoR_BP_v2.2.3qF_s	\$Fbar	2 6	558.27	19
	\$KeyVarLogN	0 1 1 1 1 1 1 1 1		

The one-step-ahead (OSA) residuals for run “GoR\_BP\_base” are shown in Figure 1. From the residual plot (Figure 1) it can be seen that there are some larger positive residuals in the beginning of the catch timeseries while the residuals are more even sized the closer we move nowadays. In the acoustics survey residual plot, there is certain year effect seen (e.g., years 2001,2011,2017, 2018). In addition, the residuals are larger for two older age groups, age 7 and 8. Next step is to resolve the problem with larger residuals for ages 7 and 8+ in the survey. For this the decoupling of variance parameters for the observations for ages 1, 2-6 and 7, 8+ is tried (Table 1, run “GoR\_BP\_vol1”). The corresponding OSA residuals for this fit are shown in comparison in Figure 1. This modification led to some improvement in the OSA residuals, and this is also confirmed with lower AIC value (Table 1).

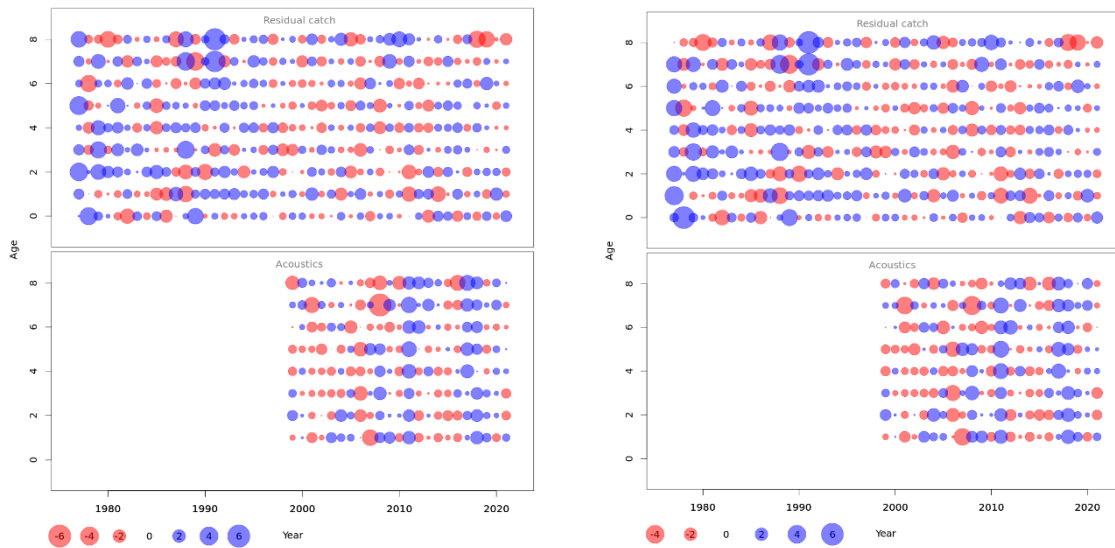


Figure 1. OSA residuals base (left) and vol1 (right).

Next step is to deal with the year-effects that are seen in the survey residuals. For this we assumed an AR structure for survey covariance matrix. And the correlation parameters were coupled for ages 1-4 and 5-8+. This led to significant improvement in the OSA residuals (Figure 2) and is supported also by the decreasing AIC value (Table 1).

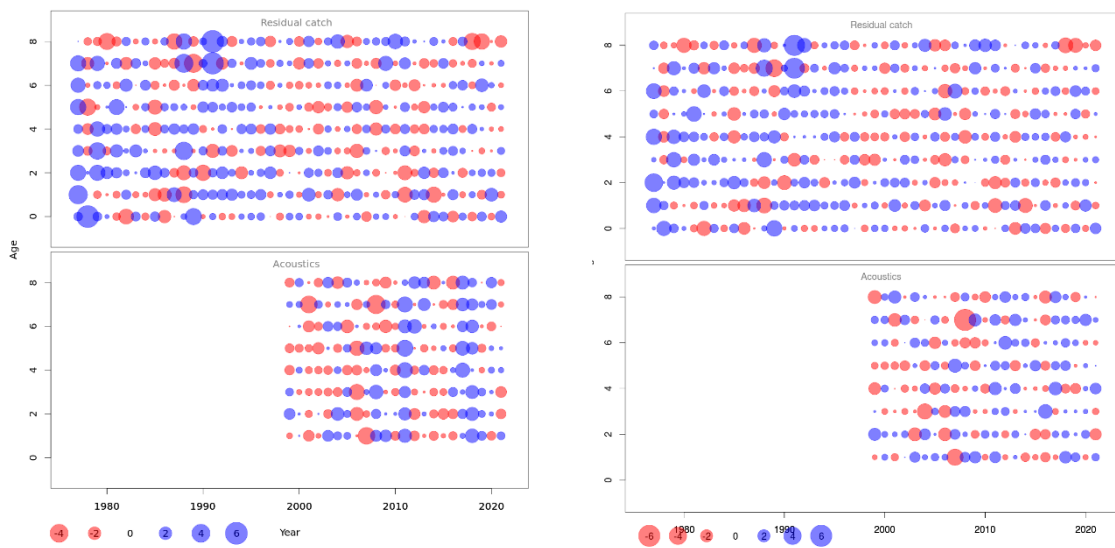


Figure 2. OSA residuals vol1 (left) and v2 (right).

There are still present some larger positive residuals in the beginning of catch timeseries, and especially for two older age groups. For the next step we investigated into the coupling of process variance parameters for log(F) process. By decoupling parameters for all ages and comparing the model parameter estimates, following coupling by age was found suitable: age 0 and 1 separate, ages 2-4 coupled, ages 5 and 6 separate, ages 7-8+ coupled. This configuration led to lower AIC (Table 1) and is named as “GoR\_BP\_v2.2”. OSA residual comparison is seen in Figure 3.

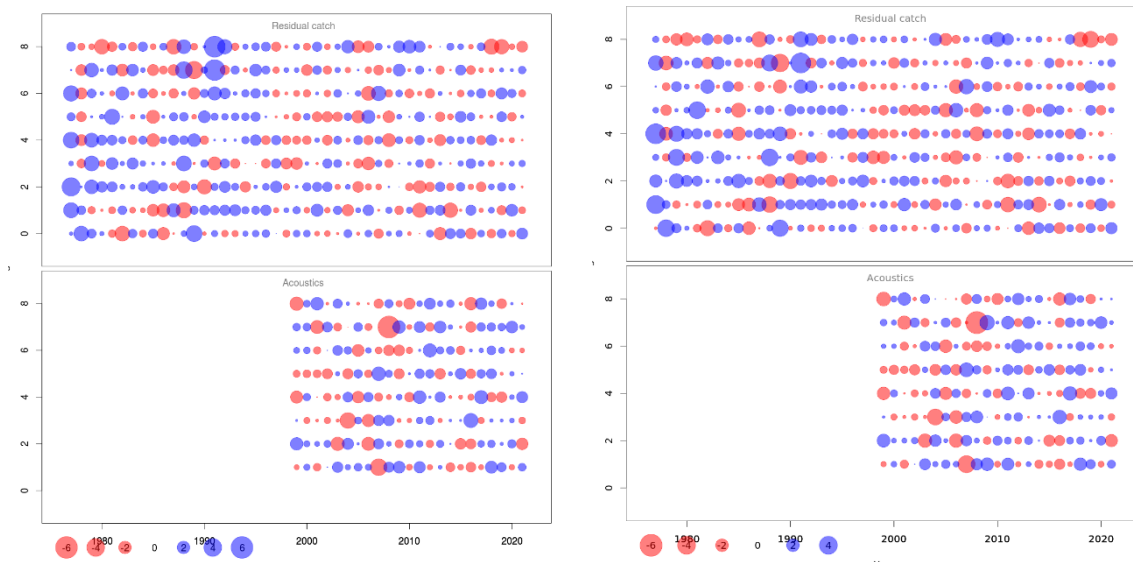


Figure 3. OSA residuals for v2 (left) and v2.2 (right).

Next step of investigation looked into the survey catchability. Thus far the survey catchability was set separate for each age, following the procedure that was adopted already in XSA. However, when looking at the model parameter estimated on survey catchability then the estimated values are very similar. Therefore, we explored different assumptions on

the survey catchability. Run “GoR\_BP\_v2.2.1q” assumes same catchability for all ages, run “GoR\_BP\_v2.2.multq” assumes four different catchabilities (coupled for ages 1-2, 3-5, 6 and 7-8+). Decreasing number of catchability parameters to estimate didn’t improve the model AIC values (Table 1), although having less catchability parameters to estimate did have an effect on the retrospective analysis, fewer catchability parameters to estimate led to lower Mohn’s rho value (Figure 4).

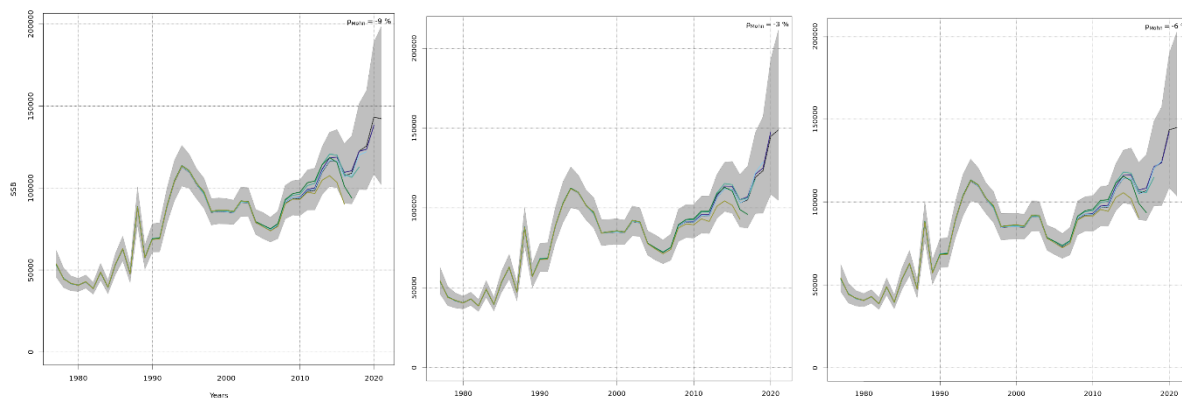


Figure 4. Retrospective analysis of SSB estimates for runs GoR\_BP\_v2.2 (left), GoR\_BP\_v2.2.1q (middle) and GoR\_BP\_v2.2.multq (right). Mohn’s Rho values: -9%, -3% and -6%, respectively.

Having separate catchability parameters for all ages doesn’t seem feasible, as the model estimates these to be almost the same, while having only one catchability for all ages might not be biologically correct. Even though based on the model estimates the catchabilities by age could roughly couple as in run “GoR\_BP\_v2.2.multq”, it is difficult to reason why would the catchability separately different for age 6 compared to younger and older ages. Therefore, another compromise was tried, assuming separate catchability for ages 1 and 2, and for ages 3-8+ coupled. This run is named as “GoR\_BP\_v2.2.3q” and led to same AIC as run “GoR\_BP\_v2.2”, which assumes different catchability parameter for all ages. As these two runs produce same AIC (Table 1), but in case of “GoR\_BP\_v2.2.3q” the model has to estimate five less parameters, making a model in some sense more robust, this run was preferred.

The  $F_{bar}$  for Gulf of Riga herring has previously been set by ages 3-7. Applicability of continuing with the same  $F_{bar}$  was tested, as there are indications that younger fisher are caught compared to the beginning of time series. To determine the best range of  $F_{bar}$ , cohort analysis was conducted. Cohort analysis by cohort year is shown in Figure 5. It is seen that in the beginning of time series the catches (in number) were highest at age 3, however this has moved more to age 2 in current times. Based on this analysis it is suggested that  $F_{bar}$  value starts from age 2. To determine the end age in  $F_{bar}$  we looked into how much of catch is proportionally represented by different  $F_{bar}$  options (Figure 6). We tested the old  $F_{bar}$  (3-7) and two new  $F_{bar}$  options, F2-5 and F2-6. From Figure 6 it can be seen that having a  $F_{bar}$  at ages 2-5 would lead to almost 80% catch included in most of years. Having  $F_{bar}$  include

ages 2-6 will lead to slightly higher inclusion of catch compared to F2-5 but will also add more stability, which is seen in retrospective patterns. The new suggested Fbar is ages 2-6.

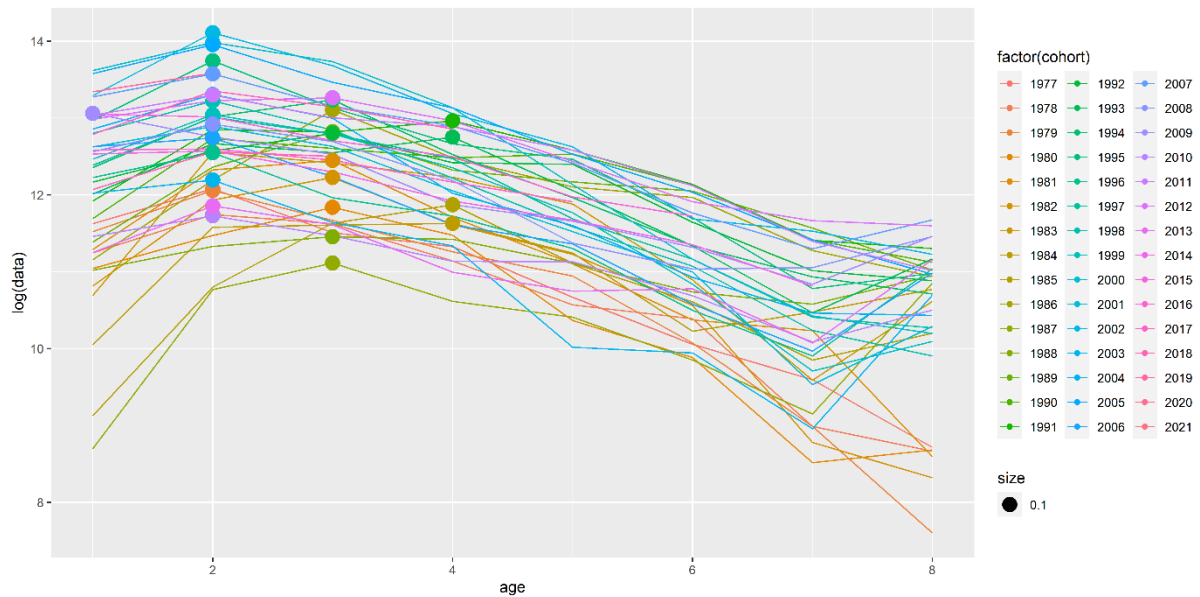


Figure 5. Cohort analysis for Gulf of Riga herring. Colours represent cohort years, and the points show the highest abundance by age in that cohort.

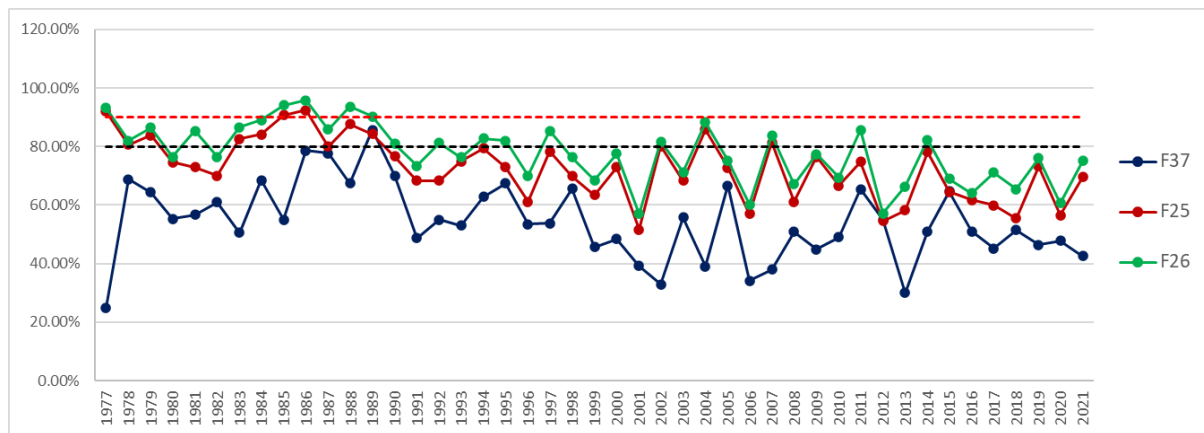


Figure 6. Proportion of catch assuming Fbar=F3-7 (dark blue), Fbar=F2-5 (red), Fbar=F2-6 (green). Blue dashed line is 80% and red dashed line 90% mark.

In the final steps of model configuration, it was discovered that the recruitment estimates at age 0 have very little variability and do not follow similar drastic changes as seen for age 1. After investigation it was discovered that one crucial setting in model configuration was wrongly specified. This is the coupling of recruitment and survival process variance. In the beginning of configuration this was accidentally set separate for ages 0, 1 and 2+. This configuration set a separate variance between age 0 and age 1. This allowed the model to ignore the connection down the cohort, meaning that the signal about recruitment strength didn't move back in time trough cohort. A more standard setting was introduced (run

“GoR\_BP\_v2.2.3qF\_s”), where same survival variance between all ages was set, and this led to much better estimates of age 0 (Figure 7).

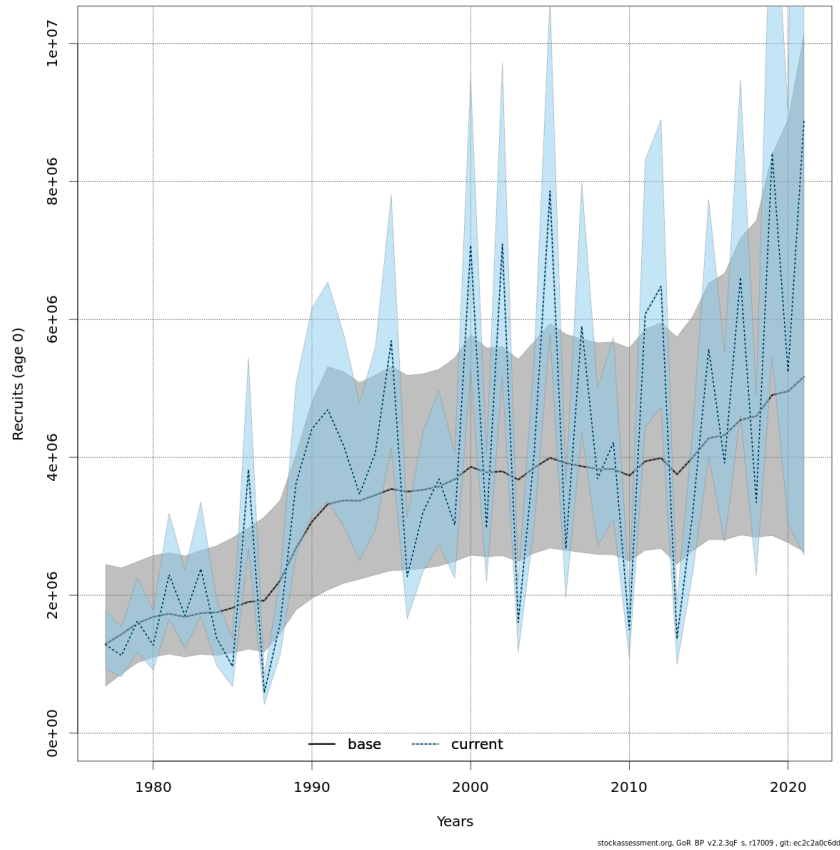


Figure 7. Comparison of recruitment (age 0) estimates for run “GoR\_BP\_v2.2.3q” (black line) and “GoR\_BP\_v2.2.3qF\_s” (blue line).

Run “GoR\_BP\_v2.2.3qF\_s” was the final agreed model configuration. Results on SSB, F and recruitment are shown in Figure 8.

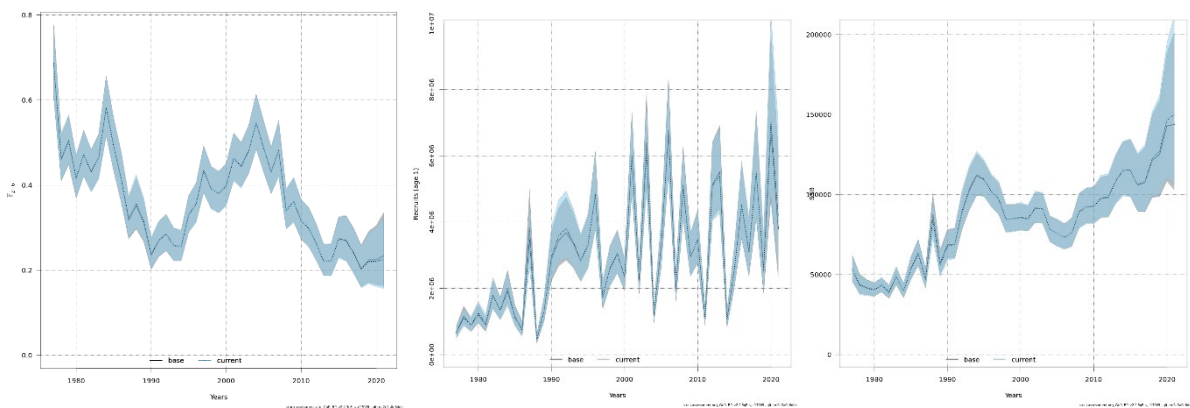


Figure 8. Summary output for the final model configuration (run “GoR\_BP\_v2.2.3qF\_s”).



As previously mentioned, there is year-effect seen in the residuals of the acoustics survey. Those year-effects are most probably caused by the fact that in certain years the cohort is not tracked very well (i.e, look Table 3) in certain age-classes.

Table A1. Gulf of Riga herring catch-at-age numbers ( $10^3$ ).

Year/age	0	1	2	3	4	5	6	7	8+
1977	800	69500	885100	141400	109700	35300	15700	16000	600
1978	7600	112000	97300	403900	39200	35900	9300	3200	5700
1979	15400	76700	176500	103800	342500	22100	19300	6800	5500
1980	18500	101000	125900	99600	55400	133100	10500	8600	2500
1981	10700	62500	172500	112000	83000	51400	71700	7400	3500
1982	1400	80000	96000	116900	68800	43000	29900	24500	3300
1983	3100	49700	225300	138300	77700	38900	23300	15500	9600
1984	1900	44000	152100	255100	96300	56700	32500	14700	11900
1985	4400	23200	283900	203900	121700	31800	23700	8000	6100
1986	1000	9200	106700	246900	110600	66500	19600	8000	5800
1987	1000	70000	49000	110000	205000	75000	32000	5000	2000
1988	1400	6000	197700	112700	112400	144600	38700	27800	5900
1989	15100	61100	47400	492700	143000	76300	53900	6500	5400
1990	12500	88100	83100	67100	263500	66800	27600	14600	4100
1991	18500	119500	234000	94500	40800	180500	40500	35400	40800
1992	12100	150300	339100	369300	91300	33200	157400	19000	47600
1993	8600	192200	381400	298100	224400	66800	19000	78800	26900
1994	11760	164230	288440	368870	263500	192700	46080	9410	56150
1995	18100	232400	316900	363000	426900	277200	170900	39300	51500
1996	31700	428800	450100	281400	247600	291000	183800	105600	57000
1997	31700	204200	930700	559700	345400	242800	186700	90600	61100
1998	19600	239360	282060	505410	274890	172470	114020	90230	67650
1999	31400	361890	446500	157050	316480	157200	83650	60670	81050
2000	49700	259030	552300	359430	123730	258070	83980	35120	53370
2001	38700	819480	461570	378160	261040	81170	120980	56040	70710
2002	29057	304160	1182680	360540	202120	118950	36310	48060	44940
2003	5930	591660	396178	922839	231178	107441	70509	19995	58637
2004	50863	166756	1342017	306214	505774	129160	64392	33204	73423
2005	44630	384871	205390	833206	213430	171555	55243	27450	28925
2006	70251	787870	600122	113606	467376	100900	70418	16470	20007
2007	28897	305069	1145972	441269	83886	305940	59687	33710	24165
2008	40183	583363	341051	703895	165817	22389	119082	13798	26776
2009	55660	274301	765448	200530	494726	107356	20478	100014	28994
2010	48129	469192	407892	515483	109991	275715	55632	7764	75734
2011	48443	88964	327256	391007	278589	170847	128611	31572	63420
2012	76397	458920	123970	276010	196090	245430	39330	90650	33980
2013	17708	435220	596630	95600	143650	86850	128500	21350	57920
2014	50932	76960	553760	443440	68530	115750	62060	80660	58830
2015	108856	277380	141080	575230	394950	68160	82500	63190	117450
2016	36183	467310	287890	110350	427240	291430	43770	50850	94760
2017	61159	291780	449000	219830	59410	251400	183300	24030	94910
2018	29515	357867	295664	329437	150533	46463	149032	88866	36412
2019	64518	174379	629505	255381	267814	117162	48007	116436	60657
2020	41046	623754	285022	512507	192367	158621	85216	23743	109093
2021	136985	314882	794199	268629	384044	148641	123598	49741	70121

TableA 2. Gulf of Riga weight-at-age in catch and stock (kg).

Year/age	0	1	2	3	4	5	6	7	8+
1977	0.00290	0.01320	0.01600	0.02270	0.02690	0.02950	0.03120	0.02940	0.05080
1978	0.00530	0.00980	0.01770	0.02190	0.02730	0.03110	0.03040	0.03810	0.05040
1979	0.00630	0.01220	0.01620	0.02340	0.02760	0.02980	0.03400	0.03680	0.03600
1980	0.00710	0.01450	0.02010	0.02410	0.03210	0.03930	0.04560	0.05330	0.07110
1981	0.00760	0.01210	0.02160	0.02880	0.03340	0.03900	0.04390	0.04990	0.05950
1982	0.00540	0.01410	0.02140	0.02870	0.03570	0.03720	0.04510	0.05030	0.06837
1983	0.00570	0.01380	0.01930	0.02760	0.03790	0.04160	0.05090	0.06100	0.09130
1984	0.00540	0.01000	0.01500	0.02150	0.02810	0.03430	0.03910	0.04910	0.05590
1985	0.00600	0.01290	0.01720	0.02080	0.02780	0.03580	0.04870	0.05310	0.06650
1986	0.00600	0.01260	0.01980	0.02560	0.03140	0.04020	0.04620	0.06390	0.07090
1987	0.00600	0.01010	0.01540	0.01970	0.02630	0.03030	0.03790	0.04310	0.09050
1988	0.00660	0.01170	0.01860	0.02100	0.02730	0.03680	0.04340	0.05860	0.07500
1989	0.00670	0.01200	0.01480	0.01660	0.01960	0.02300	0.03150	0.03820	0.03640
1990	0.01140	0.01460	0.01780	0.01980	0.02690	0.03060	0.03310	0.05220	0.05540
1991	0.00690	0.01190	0.01540	0.01780	0.01990	0.02140	0.02250	0.02690	0.03360
1992	0.00630	0.01120	0.01360	0.01770	0.02150	0.02360	0.02500	0.02640	0.03590
1993	0.00640	0.01250	0.01360	0.01610	0.02010	0.02470	0.02630	0.02750	0.03520
1994	0.00410	0.01120	0.01460	0.01620	0.01880	0.02150	0.02520	0.02630	0.03000
1995	0.00540	0.01040	0.01360	0.01640	0.01790	0.02090	0.02290	0.02630	0.02910
1996	0.00390	0.01050	0.01250	0.01570	0.01770	0.01890	0.02150	0.02350	0.02800
1997	0.00490	0.00970	0.01240	0.01490	0.01780	0.01910	0.01960	0.02120	0.02420
1998	0.00660	0.01010	0.01330	0.01690	0.01820	0.02030	0.02130	0.02250	0.02400
1999	0.00490	0.01310	0.01550	0.01890	0.02210	0.02310	0.02450	0.02650	0.02890
2000	0.00631	0.01250	0.01650	0.02010	0.02290	0.02540	0.02640	0.02820	0.02960
2001	0.00523	0.01020	0.01600	0.02050	0.02300	0.02450	0.02770	0.02830	0.03070
2002	0.00495	0.01000	0.01530	0.01930	0.02360	0.02500	0.02710	0.02800	0.03090
2003	0.00468	0.00758	0.01530	0.01995	0.02226	0.02476	0.02632	0.02678	0.02760
2004	0.00445	0.00863	0.01012	0.01651	0.02103	0.02422	0.02676	0.02709	0.03310
2005	0.00525	0.01198	0.01393	0.01583	0.01930	0.02411	0.02536	0.02871	0.03080
2006	0.00541	0.00857	0.01319	0.01776	0.01913	0.02284	0.02656	0.02752	0.02960
2007	0.00562	0.00891	0.01166	0.01544	0.02020	0.01957	0.02369	0.02715	0.02780
2008	0.00541	0.00976	0.01493	0.01728	0.02047	0.02389	0.02331	0.02845	0.03270
2009	0.00584	0.00916	0.01399	0.01755	0.01907	0.02177	0.02068	0.02441	0.02940
2010	0.00452	0.00913	0.01380	0.01685	0.01942	0.02089	0.02369	0.02307	0.02600
2011	0.00448	0.01232	0.01586	0.01838	0.02152	0.02377	0.02540	0.02568	0.02877
2012	0.00545	0.00940	0.01593	0.02026	0.02317	0.02581	0.02771	0.02994	0.03340
2013	0.00582	0.00965	0.01465	0.01966	0.02266	0.02572	0.02820	0.02952	0.03190
2014	0.00562	0.00981	0.01384	0.01760	0.02158	0.02356	0.02534	0.02709	0.03020
2015	0.00576	0.00892	0.01502	0.01822	0.02108	0.02297	0.02516	0.02723	0.02950
2016	0.00599	0.00864	0.01516	0.01810	0.02039	0.02227	0.02388	0.02596	0.02830
2017	0.00514	0.00866	0.01473	0.01852	0.02093	0.02251	0.02412	0.02481	0.02760
2018	0.00649	0.00965	0.01532	0.01909	0.02159	0.02298	0.02452	0.02561	0.02840
2019	0.00592	0.00871	0.01357	0.01809	0.02066	0.02320	0.02366	0.02477	0.02620
2020	0.00602	0.00899	0.01535	0.01890	0.02123	0.02310	0.02499	0.02473	0.02600
2021	0.00539	0.00862	0.01379	0.01775	0.01963	0.02148	0.02310	0.02470	0.02530

Table A3. Gulf of Riga herring maturity ogive

Year/age	0	1	2	3	4	5	6	7	8+
1977-1994	0	0	0.93	0.98	0.98	1.0	1.0	1.0	1.0
1995	0	0.254	0.706	0.941	0.991	1.0	1.0	1.0	1.0
1996	0	0.251	0.702	0.939	0.990	1.0	1.0	1.0	1.0
1997	0	0.249	0.698	0.938	0.990	1.0	1.0	1.0	1.0
1998	0	0.246	0.694	0.936	0.989	1.0	1.0	1.0	1.0
1999	0	0.244	0.690	0.934	0.988	1.0	1.0	1.0	1.0
2000	0	0.242	0.686	0.932	0.987	1.0	1.0	1.0	1.0
2001	0	0.240	0.681	0.930	0.986	1.0	1.0	1.0	1.0
2002	0	0.239	0.676	0.927	0.985	1.0	1.0	1.0	1.0
2003	0	0.238	0.670	0.924	0.984	1.0	1.0	1.0	1.0
2004	0	0.237	0.664	0.921	0.983	1.0	1.0	1.0	1.0
2005	0	0.237	0.658	0.918	0.982	1.0	1.0	1.0	1.0
2006	0	0.237	0.651	0.915	0.981	1.0	1.0	1.0	1.0
2007	0	0.237	0.645	0.911	0.979	1.0	1.0	1.0	1.0
2008	0	0.238	0.640	0.908	0.978	1.0	1.0	1.0	1.0
2009	0	0.239	0.635	0.905	0.977	1.0	1.0	1.0	1.0
2010	0	0.241	0.631	0.902	0.975	1.0	1.0	1.0	1.0
2011	0	0.245	0.628	0.899	0.973	1.0	1.0	1.0	1.0
2012	0	0.250	0.626	0.896	0.972	1.0	1.0	1.0	1.0
2013	0	0.257	0.626	0.893	0.970	1.0	1.0	1.0	1.0
2014	0	0.265	0.626	0.890	0.968	1.0	1.0	1.0	1.0
2015	0	0.274	0.627	0.886	0.965	1.0	1.0	1.0	1.0
2016	0	0.284	0.629	0.883	0.963	1.0	1.0	1.0	1.0
2017	0	0.295	0.632	0.881	0.962	1.0	1.0	1.0	1.0
2018	0	0.306	0.636	0.879	0.960	1.0	1.0	1.0	1.0
2019	0	0.317	0.639	0.877	0.959	1.0	1.0	1.0	1.0
2020	0	0.328	0.643	0.875	0.958	1.0	1.0	1.0	1.0
2021	0	0.338	0.647	0.873	0.956	1.0	1.0	1.0	1.0

Table A4. Gulf of Riga herring hydroacoustic survey.

Year/age		1	2	3	4	5	6	7	8+
1999	1	5292	4363	1343	1165	457	319	208	98
2000	1	4486	4012	1791	609	682	336	151	243
2001	1	7567	2004	1447	767	206	296	56	173
2002	1	3998	5994	1068	526	221	87	165	128
2003	1	12441	1621	2251	411	263	269	46	193
2004	1	3177	10694	675	1352	218	195	94	137
2005	1	8190	1564	4532	337	691	92	75	83
2006	1	12082	1986	213	937	112	223	36	49
2007	1	1478	3662	1265	143	968	116	103	39
2008	1	9231	2109	4398	816	134	353	6	23
2009	1	6422	4703	870	1713	284	28	223	44
2010	1	5077	2311	1730	244	593	107	12	50
2011	1	3162	5289	2503	2949	597	865	163	162
2012	1	5957	758	1537	774	1035	374	308	193
2013	1	9435	5552	592	1240	479	827	187	427
2014	1	1109	3832	2237	276	570	443	466	370
2015	1	3221	539	1899	1110	255	346	181	325
2016	1	4542	1081	504	1375	690	152	113	103
2017	1	3231	3442	874	402	1632	982	137	752
2018	1	11216	4529	3607	776	338	1439	755	381
2019	1	4912	7007	2237	1335	475	228	681	265
2020	1	9947	2659	3641	1234	1131	403	201	805

### WD-13. Reference point calculation for Gulf of Riga herring in subdivision 28.1

EqSim was run on the new SAM assessment results to determine new reference points for Gulf of Riga herring which were compared to the previous reference points. EqSim was run with the average of the last 5 years of biological data and the last 5 years of fishing selectivity data, the values of  $\sigma_{SSB}$  (0.177),  $F_{cv}$  and  $F_{phi}$  (0.25, 0.30), and without autocorrelation in recruitment. The suggested new reference point  $F_{MSY}$  is 0.28.

#### 1. Estimation of new reference points

New reference points for Gulf of Riga herring were estimated for the new SAM assessment model, which was agreed upon at WKBBALPEL 2023. This was done in a stepwise process, using the EqSim analysis (standardized ICES code) and ICES technical guidelines, detailed in the sections below. These new reference points were then compared to the current ones.

#### 2. Methods

##### 2.2. Estimating $B_{lim}$ and PA reference points

$B_{lim}$  is an important reference point from which other precautionary reference points are derived. To determine  $B_{lim}$ , the full assessment data series should be used to determine stock type in terms of the SSB-recruitment relationship.  $B_{lim}$  can then be defined accordingly following the ICES Technical Guidelines (ICES, 2021).

$B_{pa}$  can be estimated based on  $B_{lim}$  as follows, where  $\sigma_{SSB}$  value will be taken from the SAM assessment:

$$B_{pa} = B_{lim} * \exp(1.645 * \sigma_{SSB})$$

In cases where  $B_{pa}$  can be estimated but  $B_{lim}$  cannot, a proxy for  $B_{lim}$  is considered based on the inverse of the standard factor for calculating  $B_{pa}$  from  $B_{lim}$  (i.e. a  $B_{lim}$  proxy equal to  $B_{pa}/1.4$ ). It should be noted that when calculating the  $B_{lim}$  proxy the factor 1.4 (equivalent to  $\sigma = 0.20$ ) should be applied, instead of  $\exp(1.645 * \sigma)$  with  $\sigma$  from the assessment uncertainty in SSB in the terminal year.

To estimate  $F_{lim}$ , EqSim should be run without assessment/advice error and without advice rule (without MSY  $B_{trigger}$ ), using a segmented regression with breakpoint fixed at  $B_{lim}$ , as per the ICES guidelines (ICES, 2021), to model the spawning stock recruitment relationship, in order to get the  $F$  ( $F_{50}$ ) that ensures a 50% probability for SSB to remain above  $B_{lim}$ .

According to the latest ICES Technical Guidelines (ICES, 2021),  $F_{pa}$  is no longer estimated from  $F_{lim}$  but instead should be at  $F_{p.05}$ .

Table 1. Categorization of stock types as presented in the ICES Technical guidelines (ICES, 2021).

Stock characteristics			Limit point estimation options dependent on data and specific stock information	
Stock type	S–R plot characteristics	Sample S–R plot	$B_{lim}$ estimation possible according to standard method	$B_{lim}$ estimation possible on the basis of stock-specific method or judgement
Type 1	Spasmodic stocks – stocks with occasional large year classes.		$B_{lim}$ is based on the lowest SSB, where large recruitment is observed – unless F has been low throughout the observed history, in which case $B_{loss} = B_{pa}$ .	
Type 2	Stocks with a wide dynamic range of SSB, and evidence that recruitment is or has been impaired.		$B_{lim} =$ segmented regression change point.	
Type 3	Stocks with a wide dynamic range of SSB, and evidence that recruitment is or has been impaired, with no clear asymptote in recruitment at high SSB.			$B_{lim}$ may be close to the highest SSB observed. The estimate depends on an evaluation of the historical fishing mortality.
Type 4	Stocks with a wide dynamic range of SSB, and evidence that recruitment increases as SSB decreases.			No $B_{lim}$ from this data, only the PA reference point. ( $B_{loss}$ would be a candidate for $B_{pa}$ ).
Type 5	Stocks showing no evidence of impaired recruitment or with no clear relation between stock and recruitment (no apparent S–R signal).		$B_{lim} = B_{loss}$	
Type 6	Stocks with a narrow dynamic range of SSB and showing no evidence of past or present impaired recruitment.			No $B_{lim}$ from this data, only the PA reference point ( $B_{loss}$ could be a candidate for $B_{pa}$ , however, this is dependent on considerations involving historical fishing mortality).

### 2.3. Estimating $F_{msy}$ , MSY $B_{trigger}$

$F_{MSY}$  should be initially calculated based on an EqSim with assessment/advice error, which should give maximum yield, and without advice rule (without MSY  $B_{trigger}$ ). For the spawning stock recruitment relationship, a suitable SRR should be used.

To include assessment and advice error, the values  $(F_{cv}, F_{phi}) = (0.25, 0.30)$  were used, the default values suggested by WKMSYREF3 (ICES, 2015).

To ensure consistency between the precautionary and the MSY frameworks,  $F_{MSY}$  is not allowed to be above  $F_{p.05}$ ; therefore, if the initial  $F_{MSY}$  value is above  $F_{p.05}$ ,  $F_{MSY}$  is reduced to  $F_{p.05}$ .

MSY  $B_{trigger}$  is a lower bound of the SSB distribution when the stock is fished at  $F_{MSY}$  (ICES, 2021). To set MSY  $B_{trigger}$  the flowchart in Figure 1 is followed together with recent fishing mortality estimates.

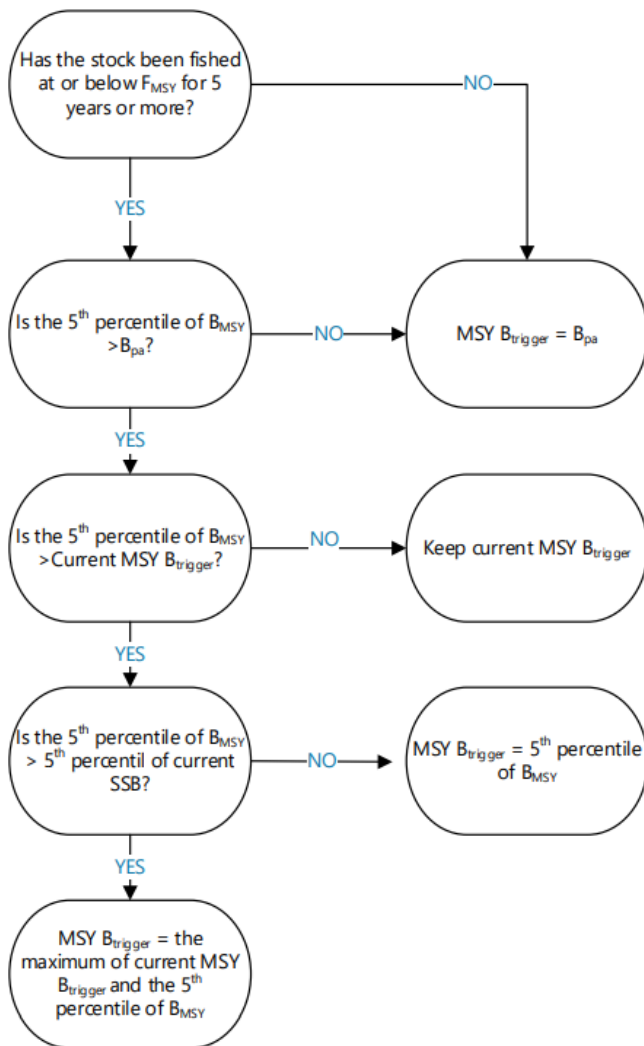


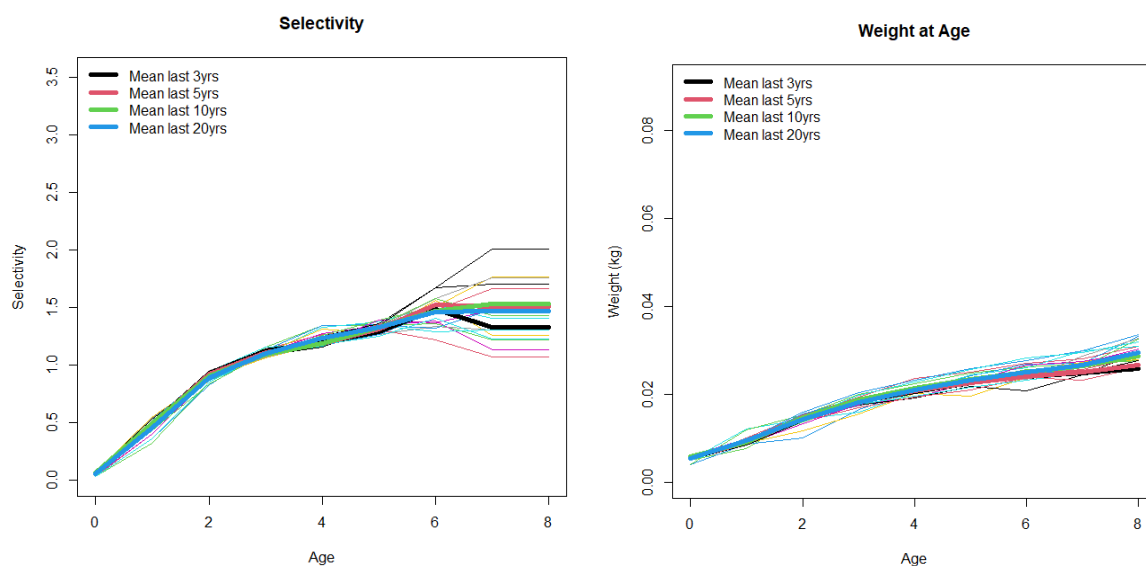
Figure 1. Flowchart to set MSY  $B_{trigger}$  as given by ICES Advice Technical guidelines (ICES, 2021).

Calculations for MSY  $B_{\text{trigger}}$  were based on EqSim runs without assessment/advice error and without advice rule, using the most suitable SRR.

When applying the advice rule (AR),  $F$  was reduced when SSB falls below this threshold. Using the advice rule, it should be checked that when fishing at  $F_{\text{MSY}}$  the probability of falling below  $B_{\text{lim}}$  remains smaller than 5%. Therefore, it should be ensured that the initially calculated  $F_{\text{MSY}}$  was at or below  $F_{p.05}$ .

#### 2.4. EqSim settings

SigmaSSB from the new SAM run was 0.172 in 2021, and this value will be used in further calculation is needed. For fisheries selectivity, the most recent 5 years were found to be representative (Fig. 2), and re-sampling from the last 5 years was used in EqSim. Likewise for weights-at-age, the most recent 5 years were found to be representative (Fig. 2), and re-sampling from the last 5 years was also used in EqSim. Autocorrelation in recruitment was not significant and therefore not included (Fig. 3).



**Figure 2.** Fisheries selectivity (left) and weight-at-age (right) at age by year and averages for recent 3, 5, 10 and 20 years.



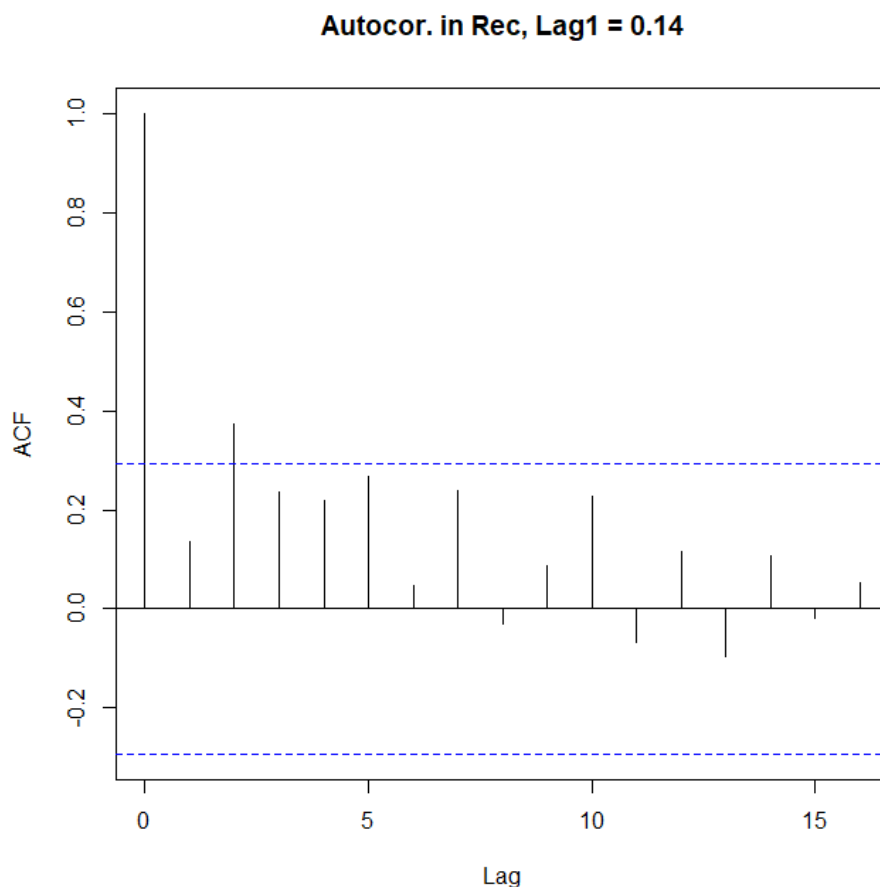


Figure 3. Autocorrelation in recruitment.

### 3. Results

#### 3.1. Blim, Flim, and Bpa

Stock-recruitment pairs for Gulf of Riga herring stock are seen in Figure 5. When fitting Beverton-Holt SRR function to the data then this produces a straight line, same problem was exhibited during the last workshop when reference points were calculated (ICES, 2015). There is no apparent SRR for this stock. In Figure 5 different SRR functions were fitted to the data (e.g., BH, smooth hockey-stick). The smooth hockey-stick function estimates the breakpoint at highest SSB, which is considered not plausible for this stock.

Based on ICES guidelines, Table 1, (ICES, 2021) stock type based on SRR should be found. Gulf of Riga herring stock is not well classified by these *types*. Based on the SR pairs indication for stock *Type 3* are strong, however for GoR herring stock the SSB is highest observed currently and recruitment is above average, hence not classifying as “*evidence that recruitment is or has been impaired*”.

As there is no clear SRR and choosing a certain SRR and choosing a certain *Type* and corresponding SRR will be subjective. Following options were tested to determine suitable Blim and/or Bpa reference points, and corresponding SRR.

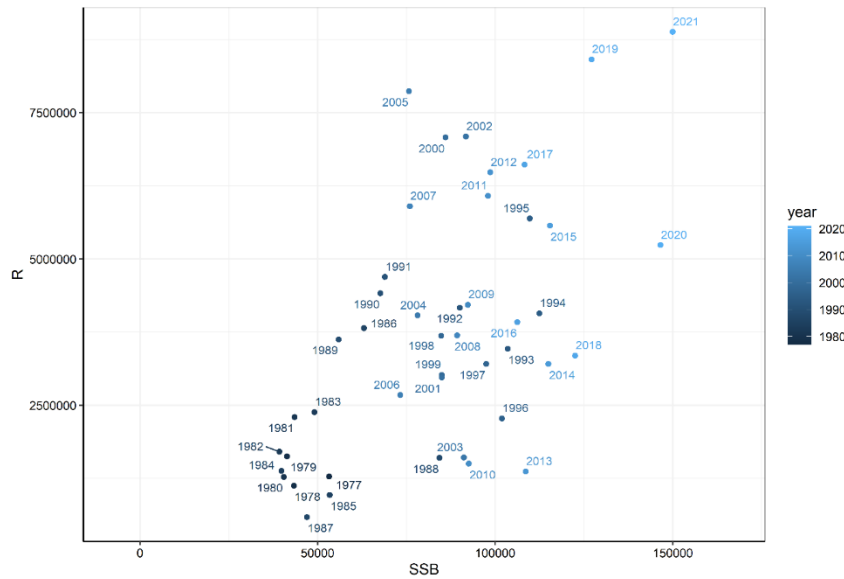


Figure 4. Gulf of Riga herring stock-recruitment relationship.

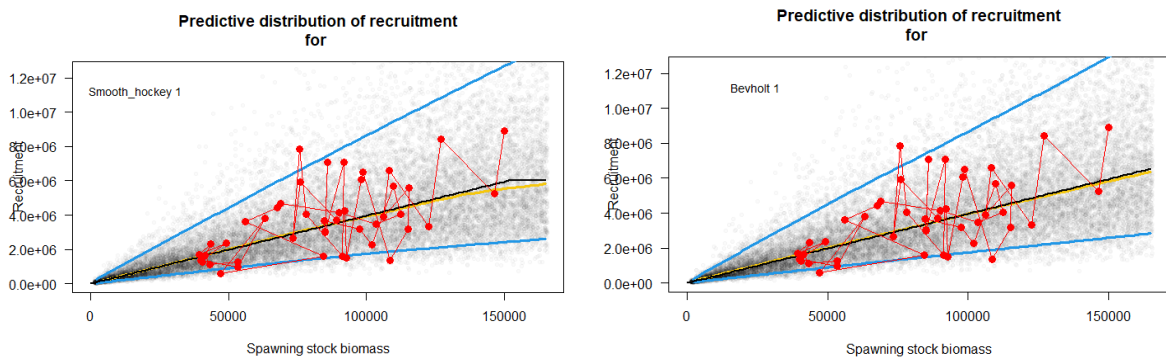


Figure 5. Fitted “smooth hokey-stik” and “Beverton-Holt” SRR functions.

### 3.1.1. Option 1 – defining Bpa

Gulf of Riga herring stock does not categorize very well under the stock *Types* defined in Table 1. Defining Blim is complicated, even though it can be seen that there are low SSB-R pairs in the beginning of time series (Figure 4), it is also seen that low level of R values can be paired with 2x higher SSB values (e.g., years 2003, 2010, 2013). Therefore, we tested alternative option which was also used during previous workshop (ICES, 2015). Blim can't be reliably estimated; however we can more reliably define Bpa, and then infer Blim from Bpa. Bpa is a stock status reference point above which the stock is considered to have full reproductive capacity.

Following steps were introduced to determine Bpa:

- i) Calculate median SSB and R based on the whole time-series
- ii) Calculate average SSB based on data points that are  $\leq$  medianSSB and  $\geq$  medianR.

The median SSB and R, and the average SSB are shown in Figure 6. The corresponding values are:

MedianSSB = 85994 tonnes

MedianR = 3626236 indiv.

averageSSB<sub>ssb=<medianSSB&R>=median</sub> = 72907 tonnes

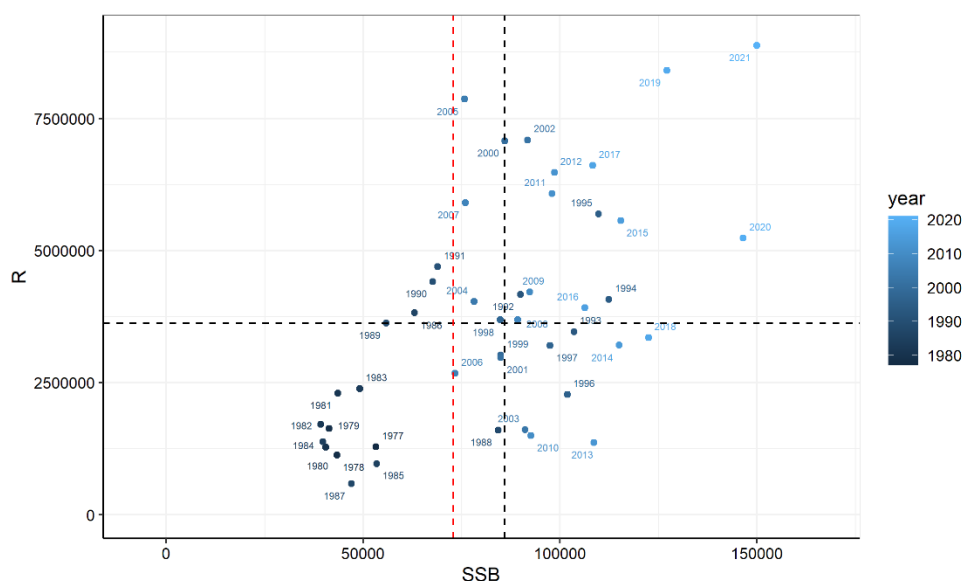


Figure 6. Gulf of Riga herring stock-recruitment relationship. Black dashed vertical line = median SSB, horizontal black dashed line = median R, red dashed line = average SSB.

Based on these assumptions:

$B_{pa} = \text{averageSSB} = 72907$  tonnes, and

$B_{lim} = B_{pa}/1.4 = 52076$  tonnes.

### 3.1.2. Option 2 – Type 2, defining Blim

Potential Blim value can be defined as the lowest SSB that still produces median recruitment. S-R pair in 1989 conforms to this assumption. Hence, we can define:

$B_{lim} = SSB_{1989} = 55892$  tonnes

$B_{pa} = B_{lim} \cdot \exp(1.645 \cdot \sigma_{SSB}) = 74783$  tonnes

### 3.1.3. Option 3 - Type 5, Blim=Bloss

Assuming that the stock conforms to under Type 5, then  $B_{lim} = B_{loss}$ .

Blim = 39228 tonnes

Bpa = Blim\*exp(1.645\*sigmaSSB) = 52056 tonnes.

### 3.2. Unconstrained F<sub>MSY</sub>

#### 3.3.1. Option 1

To estimate the unconstrained F<sub>MSY</sub>, the EqSim was run without the advice rule (i.e. no MSY B<sub>trigger</sub>), with assessment and advice error using values (F<sub>cv</sub>, F<sub>phi</sub>) = (0.25, 0.3) as suggested by WKMSYREF3 (ICES, 2015), and with a segmented relationship with a breakpoint fixed at B<sub>pa</sub> (Fig. 7). The resulting unconstrained F<sub>MSY</sub> obtained (median MSY for lanF) was F<sub>MSY</sub> = 0.277 (Table 2). The corresponding equilibrium plots are shown in Figure 8.

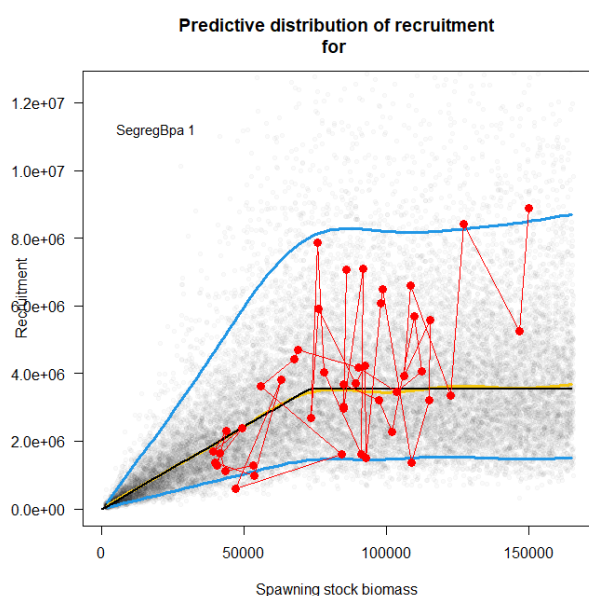


Figure 7. Segmented regression using a fixed breakpoint at B<sub>pa</sub> to fit stock-recruitment relationship.

Table 2. EqSim run with advice/assessment error and without advice rule, to determine unconstrained F<sub>MSY</sub> (segmented regression with breakpoint fixed at B<sub>pa</sub>).

	catF	lanF	catch	landings	catB	lanB
F05	0.284754	NA	27366.41	NA	94455.74	NA
F10	0.304198	NA	27079.87	NA	88271.24	NA
F50	0.372896	NA	19098.14	NA	52065.36	NA
medianMSY	NA	<b>0.278879</b>	NA	27413.59	NA	96355.66
meanMSY	0.274372	0.274372	27376.13	27376.13	97667.1	97667.1
Medlower	NA	0.211612	NA	26019.66	NA	117840.2
Meanlower	NA	0.209186	NA	27229.73	NA	NA
Medupper	NA	0.327928	NA	26018.35	NA	79688.12
Meanupper	NA	0.320037	NA	27219.94	NA	NA

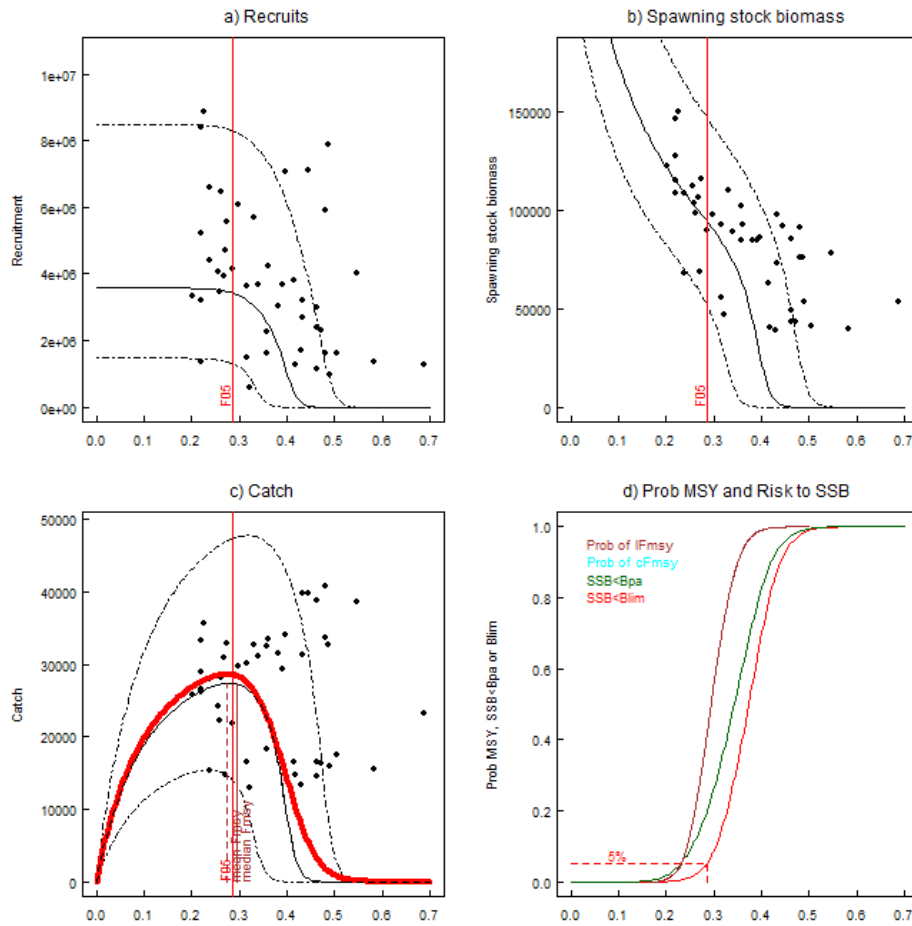


Figure 8. Equilibrium plots for the estimation of the initial (unconstrained)  $F_{MSY}$  (EqSim with assessment/advice error, and without advice rule, and with a segmented regression with breakpoint fixed at  $B_{pa}$ ).

### 3.3.2. Option 2

To estimate the unconstrained  $F_{MSY}$ , the EqSim was run without the advice rule (i.e. no MSY  $B_{trigger}$ ), with assessment and advice error using values  $(F_{cv}, F_{phi}) = (0.25, 0.3)$  as suggested by WKMSYREF3 (ICES, 2015), and with a segmented relationship with a breakpoint fixed at  $B_{lim}$  (Fig. 9). The resulting unconstrained  $F_{MSY}$  obtained (median MSY for lanF) was  $F_{MSY} = 0.338$  (Table 3). The corresponding equilibrium plots are shown in Figure 10.

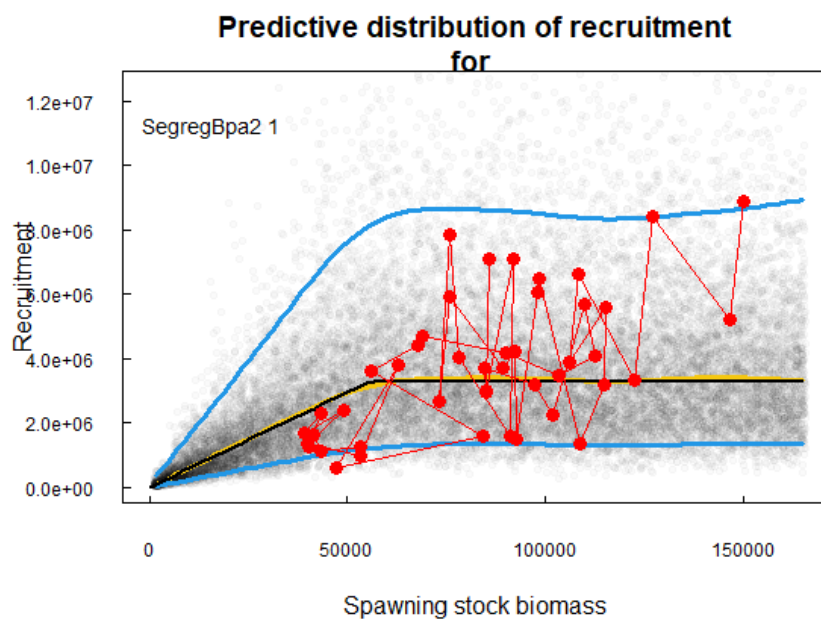


Figure 9. Segmented regression using a fixed breakpoint at  $B_{lim}$  to fit stock-recruitment relationship.

Table 3. EqSim run with advice/assessment error and without advice rule, to determine unconstrained  $F_{MSY}$  (segmented regression with breakpoint fixed at  $B_{lim}$ ).

	catF	lanF	catch	landings	catB	lanB
F05	0.287103	NA	26764.36	NA	91415.96	NA
F10	0.319491	NA	27151.19	NA	84059.61	NA
F50	0.438069	NA	23823.91	NA	55874.59	NA
medianMSY	NA	0.337538	NA	27221.8	NA	80168.65
meanMSY	0.333333	0.333333	27211.61	27211.61	81077.81	81077.81
Medlower	NA	0.245045	NA	25866.21	NA	102317.2
Meanlower	NA	0.24328	NA	27297.03	NA	NA
Medupper	NA	0.407207	NA	25867.28	NA	64632.04
Meanupper	NA	0.397264	NA	27294.54	NA	NA

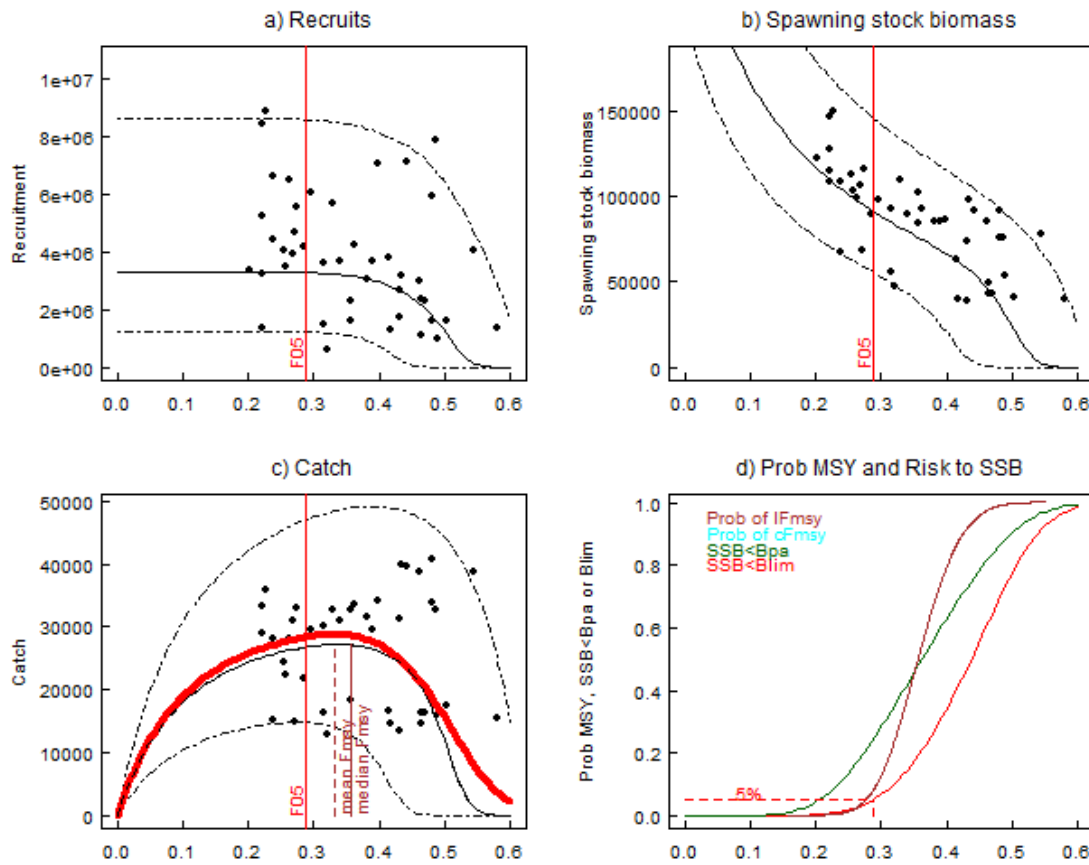


Figure 10. Equilibrium plots for the estimation of the initial (unconstrained)  $F_{MSY}$  (EqSim with assessment/advice error, and without advice rule, and with a segmented regression with breakpoint fixed at  $B_{lim}$ ).

### 3.3.3. Option 3

To estimate the unconstrained  $F_{MSY}$ , the EqSim was run without the advice rule (i.e. no MSY  $B_{trigger}$ ), with assessment and advice error using values  $(F_{cv}, F_{phi}) = (0.25, 0.3)$  as suggested by WKMSYREF3 (ICES, 2015), and with a segmented relationship with a breakpoint fixed at  $B_{lim}$  (Fig. 11). The resulting unconstrained  $F_{MSY}$  obtained (median MSY for lanF) was  $F_{MSY} = 0.439$  (Table 4). The corresponding equilibrium plots are shown in Figure 12.

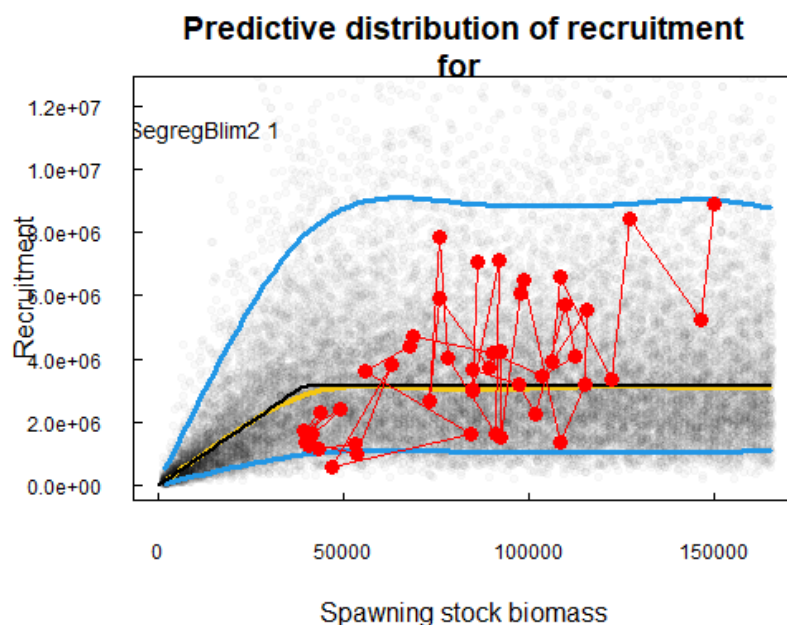


Figure 11. Segmented regression using a fixed breakpoint at  $B_{lim} = B_{loss}$  to fit the stock-recruitment relationship.

Table 4. EqSim run with advice/assessment error and without advice rule, to determine unconstrained  $F_{MSY}$  (segmented regression with breakpoint fixed at  $B_{loss}$ ).

	catF	lanF	catch	landings	catB	lanB
F05	0.401336	NA	27611.51	NA	68787.95	NA
F10	0.447694	NA	27695.38	NA	62386.15	NA
F50	NA	NA	NA	NA	NA	NA
medianMSY	NA	0.439039	NA	27704.08	NA	63534.2
meanMSY	0.436364	0.436364	27702.46	27702.46	63897.02	63897.02
Medlower	NA	0.291291	NA	26319.38	NA	88267.25
Meanlower	NA	0.295599	NA	28166.29	NA	NA
Medupper	NA	0.55015	NA	26327.91	NA	49179.45
Meanupper	NA	0.541736	NA	28172.4	NA	NA



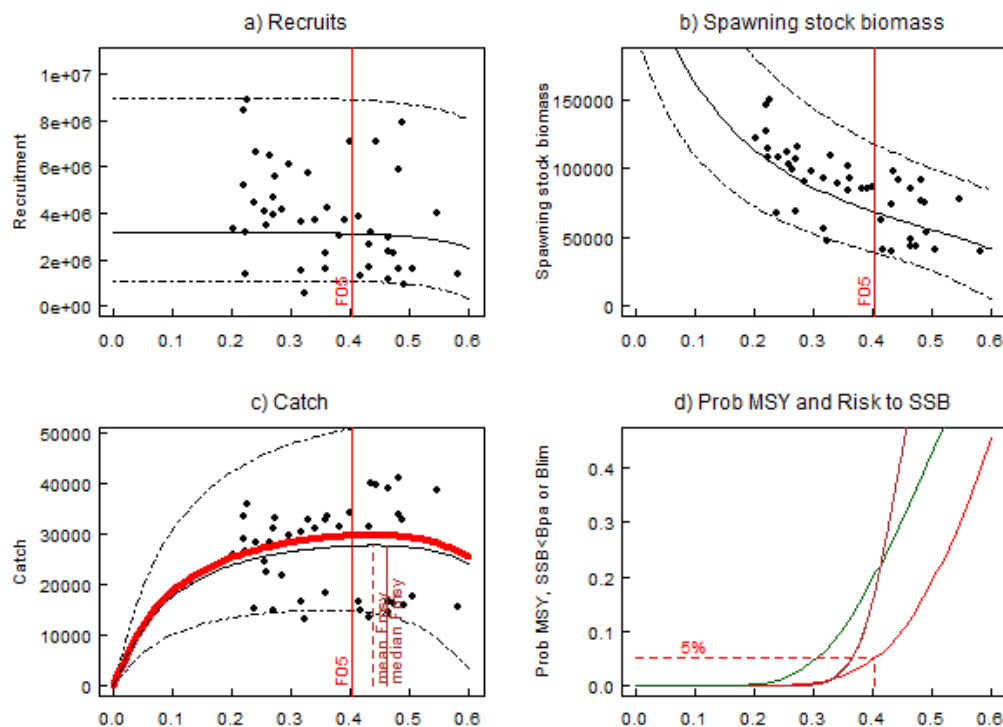


Figure 12. Equilibrium plots for the estimation of the initial (unconstrained)  $F_{MSY}$  (EqSim with assessment/advice error, and without advice rule, and with a segmented regression with breakpoint fixed at  $B_{loss}$ ).

### 3.4. Deterministic YPR and SPR reference points

In cases where there is no apparent SRR YPR and SPR reference points can be used alternatively, as determining a SRR is crucial for  $F_{MSY}$  reference points calculation. Therefore, for comparison and indication deterministic YPR and SPR reference points were calculated using the *deterministicReferencepoints* function (Albertsen and Trijoulet, 2020) in *stockassessment* package (Nielsen and Berg, 2014, 2016) in R, assuming biological input and selectivity average over last five years (2017-2021). We estimated four different reference points:  $F_{max}$ ,  $F_{30\%SPR}$ ,  $F_{35\%SPR}$ , and  $F_{40\%SPR}$ . Corresponding estimated are shown in Table 5.

Table 5. Deterministic YPR and SPR reference points.

Reference point	Median	Low	High
$F_{30\%SPR}$	0.332	0.297	0.372
$F_{35\%SPR}$	0.294	0.244	0.297
$F_{40\%SPR}$	0.221	0.203	0.241
$F_{max}$	0.752	0.646	0.876

### 3.5. Impact of different SRR assumptions

In chapter 3.3 three different descriptions were provided how to define biomass reference points (Option 1-3) and based on those assumption SRR was defined as segmented regression with a fixed breakpoint.

As the Beverton-Holt SRR function led to a straight line, and smooth-hockey stick produced also straight line with a breakpoint at higher SSB value, the only realistic option for a SRR is segmented regression with a fixed breakpoint. Option 1-3 showed three different possibilities of defining that fixed breakpoint in the segmented regression. Option 1 had the highest breakpoint and with each option the breakpoint value was lowered. From this exercise it can be seen that lowering the breakpoint value will lead to flatter simulated distribution of recruits vs F and flatter Fmsy curve. If the breakpoint would be even further lowered than in Option 3, i.e., to such a low value which would lead no SRR (testing with a value of breakpoint=5000), then the estimated unconstrained Fmsy=0.75 is same as Fmax value from a YPR analysis (Fmax=0.75) (Figure 13). This exercise again shows how important is the location of the breakpoint in the segmented regression in calculating the F<sub>MSY</sub>.

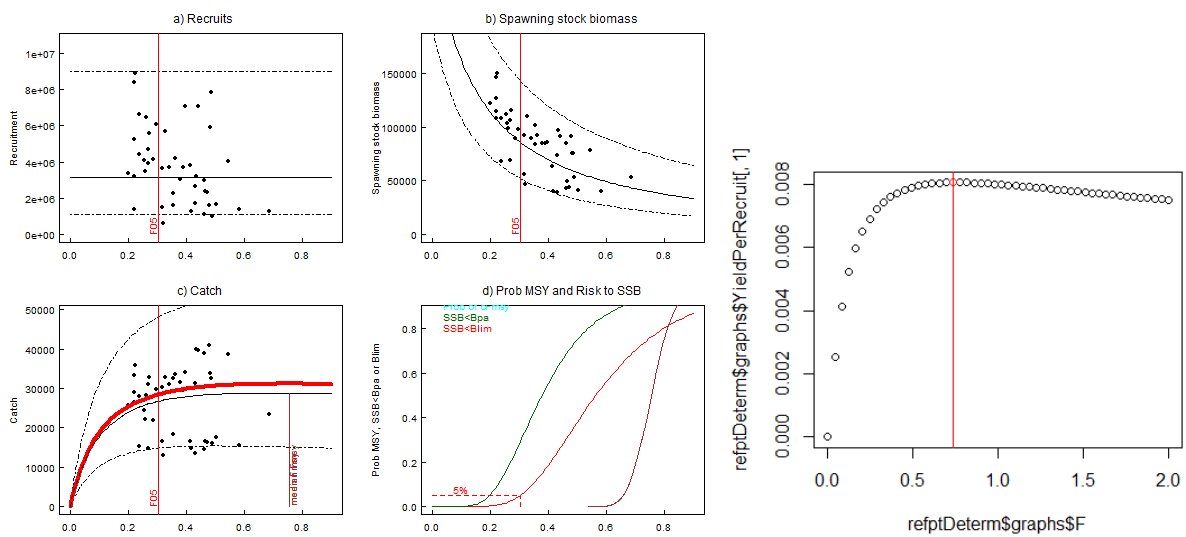


Figure 13. Equilibrium plots (left) for the estimation of the initial (unconstrained) F<sub>MSY</sub> (EqSim with assessment/advice error, and without advice rule, and with a segmented regression with breakpoint fixed at 5000 tonnes). YPR analysis (right) and F<sub>max</sub> (red).

When comparing the unconstrained F<sub>MSY</sub> estimates from Option 1-3 with the deterministic YPR and SPR reference points, it is seen that Option 1 would lead to a F<sub>MSY</sub> estimate which is between F<sub>35%SPR</sub> and F<sub>40%SPR</sub>. However, Option 2 would lead to and F<sub>MSY</sub> estimate which is slightly higher (0.34) compared to F<sub>30%SPR</sub> (0.33) estimate. F<sub>35%SPR</sub> and F<sub>40%SPR</sub> are commonly used F<sub>MSY</sub> proxies in the The North Pacific Fishery Management Council (Geromont and Butterworth, 2015).

Based on the equilibrium estimations and the comparison with deterministic YPR and SPR reference points, and we recommend using Option 1 for further calculations. This option will lead to unconstrained F<sub>MSY</sub> estimate which lies between F<sub>MSY</sub> proxies F<sub>35%SPR</sub> and F<sub>40%SPR</sub>, and therefore, is considered to be more precautionary compared to Option 2, under the current situation where no apparent SRR is present for the stock. Option 1 also conforms with the procedure that was applied to this stock during the previous ICES reference point

calculation workshop (ICES, 2015). In addition, as the  $F_{bar}$  was changed to ages 2-6 from previously used ages 3-7, we ran a comparison simulation with the old  $F_{bar}$ . With old  $F_{bar}$  Option 1 scenario will lead to unconstrained  $F_{MSY}$  estimate of 0.303 which is rather close to the old  $F_{MSY}$  estimate of 0.32.

### 3.6. MSY $B_{trigger}$

Further calculations are done based on the results from “Option 1”.

Following the flowchart (Figure 1) to set MSY  $B_{trigger}$ , next step is to investigate if the stock has been fished at or below  $F_{MSY}$  for the past five years. Based on the results in “Option 1” the unconstrained  $F_{MSY}=0.279$ , and the estimated  $F_{2-6}$  for past five years (Table 6) has been below unconstrained  $F_{MSY}$  estimate. Next step is to look if the 5<sup>th</sup> percentile of  $B_{MSY} > B_{pa}$ . This is not the case and following the flowchart (Fig.1) MSY  $B_{trigger}$  will be set at  $B_{pa}$ .

Table 6.  $F_{2-6}$  estimates from the SAM assessment.

Year	$F_{2-6}$	$F_{2-6}$ low	$F_{2-6}$ high
2017	0.238987	0.192038	0.297414
2018	0.20243	0.159056	0.257633
2019	0.220174	0.168345	0.287959
2020	0.220314	0.1617	0.300175
2021	0.225792	0.156777	0.32519

### 3.7. $F_{P0.05}$ and $F_{pa}$

$F_{P0.05}$  was calculated by running EqSim with assessment/advice error, with advice rule, and with a segmented regression with breakpoint fixed at  $B_{pa}$  to ensure that the long term risk of  $SSB < B_{lim}$  of any  $F$  used does not exceed 5% when applying the advice rule (Fig. 14).

$F_{P0.05}$  was estimated to be 0.353 (Table 6). Therefore, as explained above,  $F_{pa}=F_{P0.05}=0.353$ .

Table 6. EqSim run with assessment/advice error, with advice rule to test whether  $F_{MSY}$  was at or below  $F_{P0.05}$  (segmented regression with breakpoint fixed at  $B_{pa}$ ).

	catF	lanF	catch	landings	catB	lanB
F05	<b>0.352781</b>	NA	27818.14	NA	80257.53	NA
F10	0.390062	NA	27061.91	NA	73045.6	NA
F50	0.552833	NA	20660.43	NA	52076.35	NA
medianMSY	NA	0.319069	NA	28045.59	NA	87721.67
meanMSY	0.316583	0.316583	28040.36	28040.36	88307.79	88307.79
Medlower	NA	0.231231	NA	26652.43	NA	111607.6
Meanlower	NA	0.226368	NA	27828.22	NA	NA
Medupper	NA	0.403153	NA	26658.38	NA	70776.06
Meanupper	NA	0.41386	NA	27835.14	NA	NA

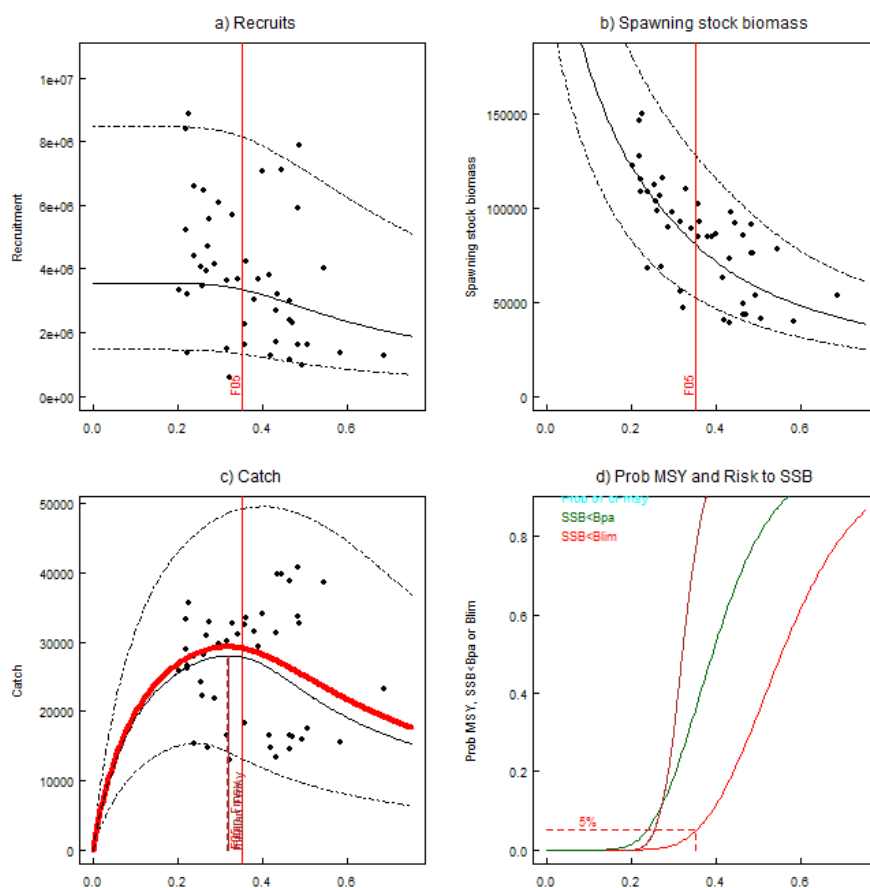


Figure 14. Equilibrium plots for the estimation of  $F_{P0.05}$  (EqSim with assessment/advice error, with advice rule, and with a segmented regression with breakpoint fixed at  $B_{pa}$ ).

### 3.8. Reference points summary table

Table 7. reference points from final EqSim settings.

	MSY $B_{trigger}$	$B_{pa}$	$B_{lim}$	$F_{pa}$	$F_{lim}$	$F_{P0.05}$	$F_{MSY\_unconstrained}$	$F_{MSY}$
Value	72907	72907	52076	0.353	0.491	0.353	0.279	0.279

### 3.9. MSY ranges

Table 8. MSY ranges.

Reference point	Value	Technical basis
$F_{MSY_{lower}}$	0.21	$F_{MSY_{lower}}$ (EqSim)
$F_{MSY}$	0.28	$F_{MSY}$ (EqSim)
$F_{MSY_{upper}}$	0.33	$F_{MSY_{upper}}$ (EqSim)

#### 4. Comparison with previous reference points

Table 9. New reference points obtained with the new assessment compared to the previous one. Note that new  $F_{\text{bar}} = F_{2-6}$  while the old  $F_{\text{bar}} = F_{3-7}$ .

Reference point	New value	Values from 2015 WK
$F_{\text{MSYlower}}$	0.21	0.24
$F_{\text{MSY}}$	0.28	0.32
$F_{\text{MSYupper}}$	0.33	0.38
MSY $B_{\text{trigger}}$	72907	60000
$B_{\text{pa}}$	72907	57000
$B_{\text{lim}}$	52076	40800
$F_{\text{pa}}$	0.35	0.38
$F_{\text{lim}}$	0.49	0.88
$F_{\text{p},05}$	0.35	0.38
$F_{\text{MSY\_unconstr}}$	0.28	0.32

## References

- Albertsen, C. M. and Trijoulet, V. 2020. Model-based estimates of reference points in an age-based state-space stock assessment model. *Fisheries Research*, 230, 105618.  
doi:10.1016/j.fishres.2020.105618
- Anders Nielsen, Casper W. Berg. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research* 158, 96-101.  
doi:10.1016/j.fishres.2014.01.014
- Casper W. Berg, Anders Nielsen. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES Journal of Marine Science* 73, 1788-1797.  
doi:10.1093/icesjms/fsw046
- Geromont H, Butterworth DS. 2015. A review of assessment methods and the development of management procedures for data-poor fisheries. FAO Report. The Marine Resource Assessment and Management Group (MARAM), University of Cape Town, South Africa. pp. 219
- ICES. 2008. Report of the Baltic Fisheries Assessment Working Group (WGBFAS), 8 - 17 April 2008, ICES Headquarters, Copenhagen. ICES CM 2008\ACOM:06. 692 pp
- ICES. 2015. Report of the Joint ICES-MYFISH Workshop to consider the basis for FMSY ranges for all stocks (WKMSYREF3), 17–21 November 2014, Charlottenlund, Denmark. ICES CM 2014/ACOM:64. 156 pp. <https://doi.org/10.17895/ices.pub.5661>
- ICES. 2021. ICES fisheries management reference points for category 1 and 2 stocks; Technical Guidelines. In Report of the ICES Advisory Committee, 2021. ICES Advice 2021, Section 16.4.3.1. <https://doi.org/10.17895/ices.advice.7891>

## WD14: Workshop on Ecosystem-Based Fisheries Advice for the Baltic II- (WKEBFABII) – preliminary conclusions and perspective.

WKEBFABII participants led by Maciej T. Tomczak, Mikaela Bergenius-Nord, Stefan Neuenfeldt

Some of the work developed at WKEBFABII was presented at WKBBALPEL in relation to Term of Reference 3b): *Estimating fisheries and biomass reference points that are in line with ICES principles and objectives when deriving reference points, including the calculation of Feco\** .

Please note that work is still ongoing and the information given here may change.

The specific aims of this work were to: (i) develop an F scaling factor (Feco) to tune the long-term Fmsy and, in this way, account for medium-term ecosystem-driven variability in productivity in the ICES advice on fishing opportunities for pelagic stocks (Central Baltic Herring stock – ICES SD 25-29 ex GOR; Baltic Sprat ICES SD 22-32) in the Baltic Sea and (ii) produce drafts of Ecological (and socio-economic) profiles (ESP) of the pelagic stocks in the Baltic. These profiles should identify quantitative indicators/factors for ecological processes that can be used to scale the species-specific Feco.

We developed potential Feco scaling factor(s) and produced very early drafts of Ecological profiles (ESP) of some of the pelagic stocks in the Baltic. Additionally, the results from the project are part of the work of the ICES Baltic Fisheries Assessment Working Group (WGBFAS) and ICES/HELCOM Working Group on Integrated Assessment of the Baltic Sea (WGIAB). Because of international efforts in data collection, mutual interest, and the need for Ecosystem-Based Advice, results and approach will be discussed and further developed at the ICES WGIAB.

Feco, developed by (Howell et al., 2021) for the Irish Sea, is a promising approach for Baltic stocks. However, over this work, we learned that it couldn't be used directly in the same way as in the Irish Sea due to different ecosystem processes and variables controlling ecological processes.

We used the Spawning Stock Biomass (SSB) and recruitment (R1) time series to reflect overall stock productivity. Based on cross-correlations and regression-based GAM models with environmental and ecosystem variables, we identified a suit of the most influential factors. It seems that there is no single factor to describe the productivity of the stocks, and a combination of factors representing different processes works better than only one variable. However, a single factor can also help explain part of stock productivity. The best candidates were biomass of zooplankton (Acartia and Pseudocalanus in spring or summer), Sea Surface Salinity in Summer and Salinity and Temperature at 60m in summer. The sea surface temperature at 60m in summer agreed with findings by (Casini et al., 2006) for the stock-recruitment relationship, but in our analysis, it was not the most influential factor for sprat.

It is also important to recognise that ecosystem scaling factors, such as Pseudocalanus biomass, may lead to misleading conclusions when used at Feco, despite explaining stock productivity well, i.e. leading to Feco way below Fmsy lower or at the Fmsy upper depending on assuming top-down or bottom-up control in the food-web. The dynamic of some zooplankton species is driven by clupeids' consumption through top-down control in the ecosystem. That is why it's essential to understand the ecology of the food web and further discuss the results in the broader expert group in the WGIAB before they are applied.

Another question raised during the analyses is the shape of the relationship between stock productivity and environmental variable. For example, the Feco approach used linear scaling for Fmsy, while as

given by GAMs, none of the relationships are linear. This suggests the need to modify the Feco approach for the Baltic using non-linear shapes or long-term state and trends (ICES 2017).

In light of using environmental variables as a scaling factor in the long- or midterm, changes in the relationship over time must be considered. STARS results in the five and decade windows show that even if we see influential factors in the long term, about 30 years, for some variables, the time series correlations are different for different periods. An understanding of this is crucial when deciding on the scaling factors. One solution could be to choose the scaling variable with a stable relationship with stock or a subset of data and repeat the analysis for a shorter period i.e. after 1991 (after the regime shift), as suggested by our analysis. That also allows to identify if variables with long-term influence can be used when the ecosystem and stock are in different regimes.

The changes in the interannual relationship it is another issue when choosing the Feco scaling variable. The spatial distributions of the CBH and Sprat stocks have changed significantly over the last years and may differ between years and regions, depending on where the bulk of stock biomass is concentrated. Using spatial distribution modelling (Orio et al., 2017) may help to identify the most influential factors in considering different ecosystem processes in different parts of the Baltic.

While we have in this project identified indicators for stock productivity, and run the Feco type of HCR (as an example) we are not at the application phase yet. To test the Feco HCR/reference points as a fishing opportunity advice use of the Management Strategy Evaluation (MSE) framework is a critical step that needs to be done. Using MSE factors like variability in the Feco HCR and uncertainty in the fisheries assessment (Gårdmark et al., 2011), need to be also taken into account when assessing the risk of applying that into the ICES advisory system. Extending the MSE with the operating model cover food-web as used by (Lucey et al., 2021) will also allow testing zooplankton biomass variables as a scaling factor for Feco in the trophic-control and changing environmental context.

Summary of next steps needed:

While we have in this project identified potential indicators for stock productivity, and run the Feco type of HCR (as an example), we are not at the application phase yet. To test the Feco HCR/reference points as a fishing opportunity advice use of the Management Strategy Evaluation (MSE) framework is a critical step that needs to be done. Using MSE factors like variability in the Feco HCR, and uncertainty in the fisheries assessment (Gårdmark et al., 2011), need to be also considered when assessing the risk of applying that into the ICES advisory system. Extending the MSE with the operating model cover food-web as used by (Lucey et al., 2021) will also allow testing zooplankton biomass variables as a scaling factor for Feco in the trophic-control and changing environmental context.

Analysis to perform and tools to apply for indicator selection and testing environmentally based HCR:

- Analyse broader context of stock productivity in the ecosystem context (see ICES/WGIAB ToRs) for a better understanding of relationships between stocks and the ecosystem i.e. zooplankton
- Regression models for testing after a regime-shift time period and potential thresholds
- Test STARS for a post-regime period and stability of correlations
- Create and perform full loop MSE procedure for Baltic stocks evaluate Feco and enviroemnatly/ecosystem-based HCR to support ICES advice on fishing opportunities
- Support development and application of ecosystem operational models for MSE



- Gårdmark, A., Nielsen, A., Floeter, J., Möllmann, C., 2011. Depleted marine fish stocks and ecosystem-based management: on the road to recovery, we need to be precautionary. *ICES J. Mar. Sci.* 68, 212–220. <https://doi.org/10.1093/icesjms/fsq158>
- Lucey, S.M., Aydin, K.Y., Gaichas, S.K., Cadrin, S.X., Fay, G., Fogarty, M.J., Punt, A., 2021. Evaluating fishery management strategies using an ecosystem model as an operating model. *Fish. Res.* 234, 105780. <https://doi.org/10.1016/j.fishres.2020.105780>
- Orio, A., Bergström, U., Casini, M., Erlandsson, M., Eschbaum, R., Hüsey, K., Lehmann, A., Ložys, L., Ustups, D., Florin, A.-B., 2017. Characterizing and predicting the distribution of Baltic Sea flounder (*Platichthys flesus*) during the spawning season. *J. Sea Res.* 126, 46–55. <https://doi.org/10.1016/j.seares.2017.07.002>
- Howell, D., Schueller, A.M., Bentley, J.W., Buchheister, A., Chagaris, D., Cieri, M., Drew, K., Lundy, M.G., Pedreschi, D., Reid, D.G., Townsend, H., 2021. Combining Ecosystem and Single-Species Modeling to Provide Ecosystem-Based Fisheries Management Advice Within Current Management Systems. *Front. Mar. Sci.* 7.
- ICES (2017): Report of the Workshop on DEveloping Integrated AdvICE for Baltic Sea ecosystem-based fisheries management (WKDEICE2). ICES Expert Group reports (until 2018). Report. <https://doi.org/10.17895/ices.pub.8612>

## Working document for Baltic pelagic benchmark workshop, WKBBALPEL, ICES, 14-18 Nov., 2022

### Effects of herring and sprat misreporting on assessment of both stocks – simulations study

Jan Horbowy

National Marine Fisheries Research Institute, Gdynia, Poland

#### Introduction & methods

Herring and sprat in the Baltic are often caught in a mixed fishery. In some cases/fleets, separation of the catches into species may be imprecise. It refers especially to fisheries from which catch is mostly used for fish meal production. For years in the reports of the WGBFAS it was stated that in some fisheries misreporting could be substantial. However, no much data on level of this misreporting are available.

Basic goal of the presented analysis was to test how assessments of both sprat & CBH stocks may be affected when different options of misreporting level and its dynamics are considered.

The analysis was performed using FLR implementation of XSA, which was primary stock assessment model for both stocks till 2022. When SAM is used as a new primary assessment model, some analyses may be repeated. The use of XSA instead of SAM should not have large impact for the results as SAM was used for both stocks as secondary assessment model and XSA & SAM estimates of stock size and fishing mortality were similar.

Two options of misreporting were considered.

**Option 1.** Very simple option in which **x% of sprat catches were assumed to be herring** (thus sprat is overreported, herring is underreported). Then, that amount of catch was taken out from sprat catches and it was added to CBH catches. CANUMs of both stocks were rescaled to comply with new values of total catches of both stocks. Two options for x% were considered: 10% and 20%.

**Option 2.** In second option a simple misreporting „model” was used. It assumed that misreporting is related to ratio of sprat to herring spawning stock biomass: if in a given year, t,

$$\text{ssbSprat}_t / \text{ssbCBH}_t > \text{average}(\text{ssbSprat} / \text{ssbCBH})$$

then in the misreporting “model” it was tendency to report more herring (underreport sprat) and vice versa. The following **misreporting fractions x** in year t in sprat were assumed dependent on the  $\text{ssbSprat}_t / \text{ssbCBH}_t$  ratio:

1. **x = -15%** if  $\text{ssbSprat}_t / \text{ssbCBH}_t < 1^{\text{st}}$  quartile of  $\text{ssbSprat}_t / \text{ssbCBH}_t$  distribution
2. **x = -5%** if  $\text{ssbSprat}_t / \text{ssbCBH}_t$  is within 1<sup>st</sup> quartile and median of  $\text{ssbSprat}_t / \text{ssbCBH}_t$  distribution
3. **x = 5%** if  $\text{ssbSprat}_t / \text{ssbCBH}_t$  is within median and 3<sup>rd</sup> quartile of  $\text{ssbSprat}_t / \text{ssbCBH}_t$  distribution
4. **x = 15%** if  $\text{ssbSprat}_t / \text{ssbCBH}_t > 3^{\text{rd}}$  quartile of  $\text{ssbSprat}_t / \text{ssbCBH}_t$  distribution

Following above misreporting in sprat, misreporting in CBH was determined. It is not claimed that above misreporting “model” is a valid model of sprat & herring misreporting; it is used only as an example to have some variability in misreporting level opposite to Option 1. where misreporting is assumed to be constant fraction of catches. The misreporting according to Option 2 is shown in Fig. 1.

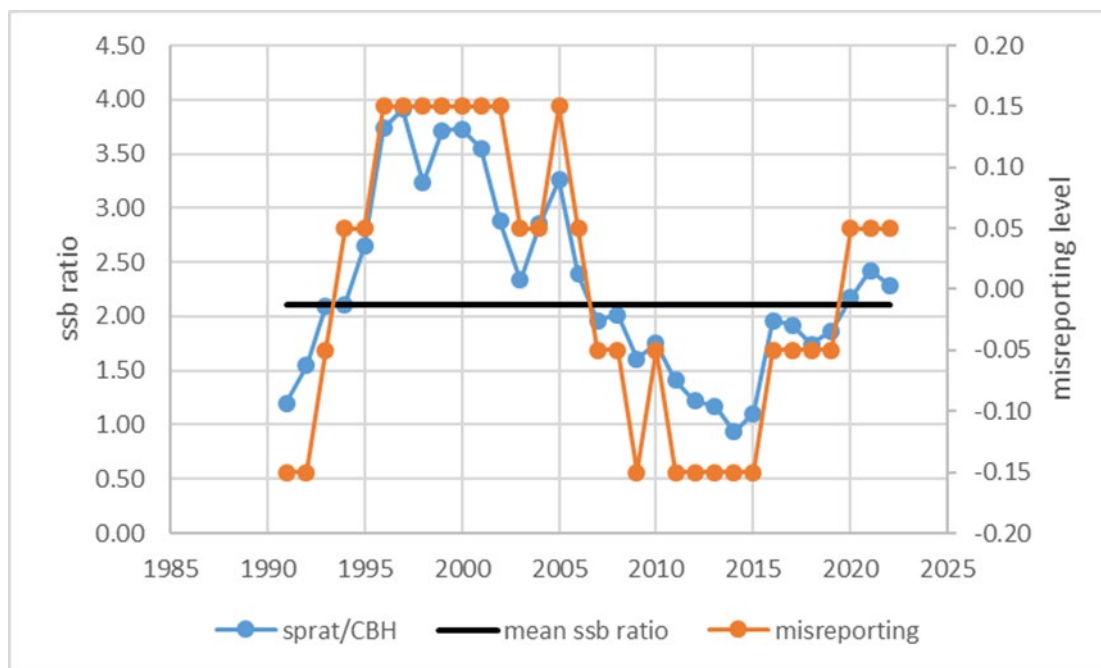


Fig. 1. The ratio of sprat spawning biomass to CBH biomass (sprat/CBH) and misreporting level of sprat assumed as dependent on that ratio.

In both options misreporting level was disturbed by random noise with assumed standard deviation, SD. For each option and assumed SD of misreporting level 200 repetitions were performed to obtain distribution of estimated yields, biomasses, fishing mortalities, and recruitments.

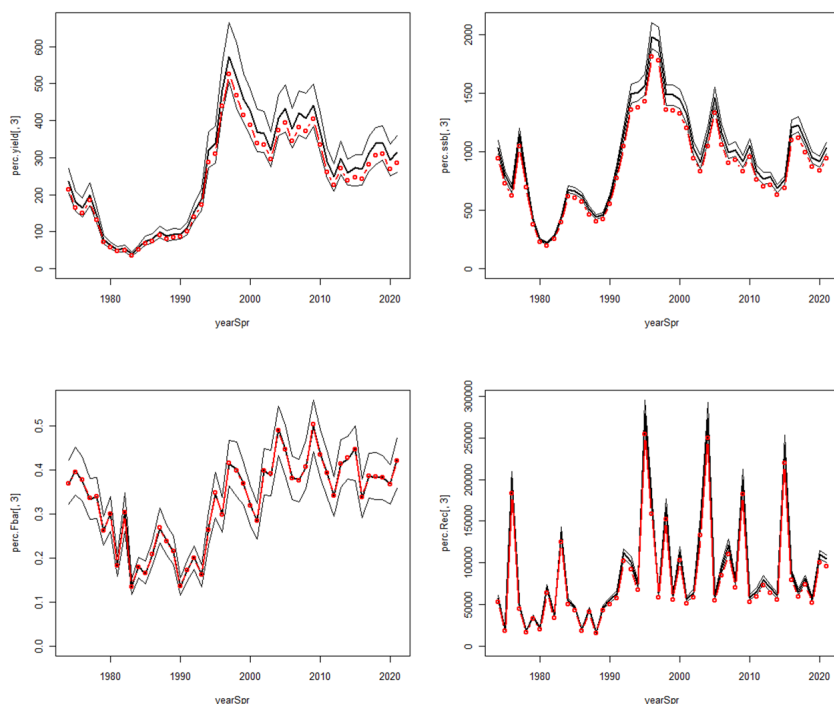
## Results

The Figures 2 and 3 show estimates of SSB, fishing mortality, and recruitment of sprat and CBH stocks under options 1 and 2 of sprat misreporting. Base case (assessment without considering misreporting) is shown in red. Thick black line is median of the estimates, thin lines represent 90% confidence intervals. The SD of  $x$  was assumed at three levels: 0.1, 0.2, 0.3.

Main outcome of the analysis was as follows:

1. Overreporting of one stock by  $x\%$  leads to overestimation of its biomass by approximately the same percentage while the estimates of average fishing mortality are only slightly affected.
2. If misreporting „fluctuates” (catches in some years are underreported and in some overreported as in Option 2) then
  - a) changes in biomass fluctuate similarly as misreporting (in some years SSB is underestimated and in some it is overestimated comparing to basic run),
  - b) no big effect of misreporting was observed on fishing mortality.
3. With increase of SD of misreporting level confidence intervals of estimated quantities increase but medians are not markedly affected.

## sprat



**CBH**

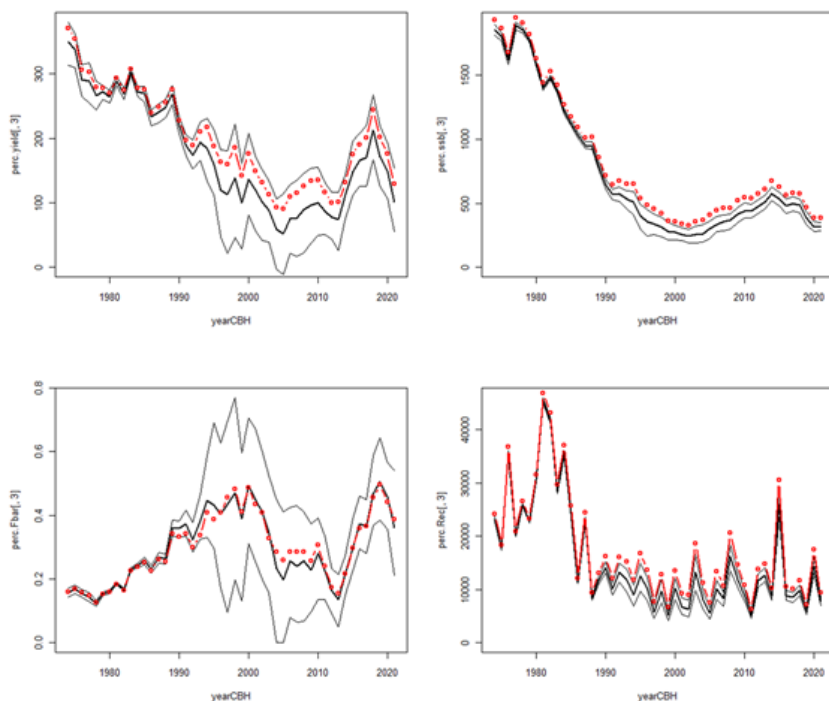
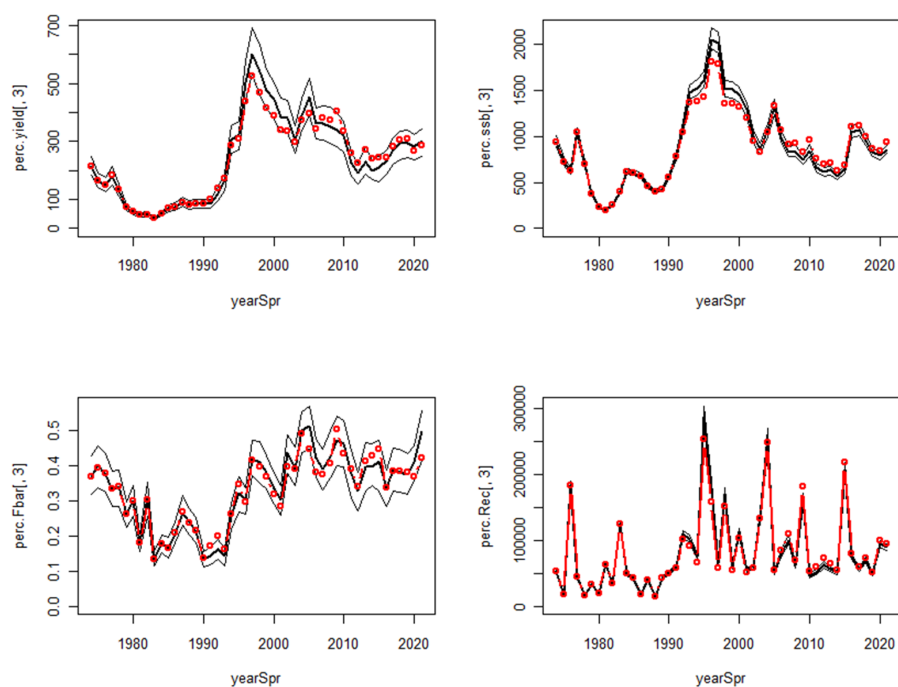


Fig. 2. Sprat (upper panel) and Central Baltic herring (bottom panel) assessments (yield, SSB, F, recruitment) assuming option 1 and sprat underreporting by 10% ( $x=-10\%$ ). Thick black lines are medians of the estimates considering misreporting, thin lines represent confidence intervals. Base case (assessment without considering misreporting) is shown in red.

**sprat**



**CBH**

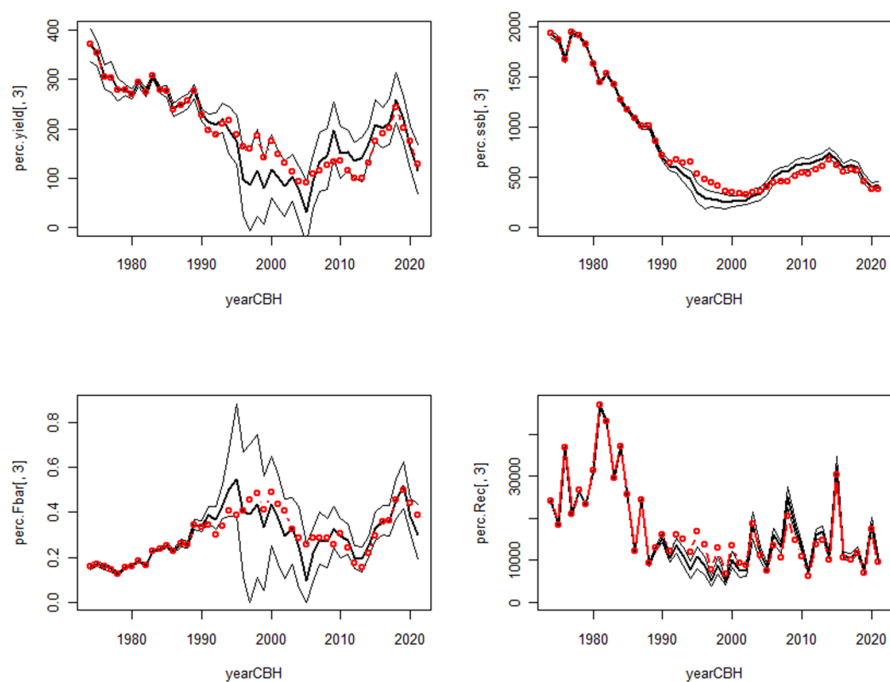


Fig. 3. Sprat (upper panel) and Central Baltic herring (bottom panel) assessments (yield, SSB, F, recruitment) assuming misreporting as in option 2. Thick black lines are medians of the estimates considering misreporting, thin lines represent confidence intervals. Base case (assessment without considering misreporting) is shown in red.

## Annex 6: Updated stock annexes

The table below provides an overview of the WKBBALTPEL Stock Annexes. Stock Annexes for other stocks are available on the ICES website Library under the Publication Type “[Stock Annexes](#)”. Use the search facility to find a particular Stock Annex, refining your search in the left-hand column to include the year, ecoregion, species, and acronym of the relevant ICES expert group.

Name	Title
<a href="#">ple.27.24-32</a>	Stock Annex: Plaice ( <i>Pleuronectes platessa</i> ) in subdivisions 24–32 (Baltic Sea, excluding the Sound and Belt Seas)
<a href="#">her.27.25-2932</a>	Stock Annex: Herring ( <i>Clupea harengus</i> ) in subdivisions 25–29 and 32, excluding the Gulf of Riga (central Baltic Sea)
<a href="#">her.27.28</a>	Stock Annex: Herring ( <i>Clupea harengus</i> ) in Subdivision 28.1 (Gulf of Riga)
<a href="#">spr.27.22-32</a>	Stock Annex: Sprat ( <i>Sprattus sprattus</i> ) in subdivisions 22–32 (Baltic Sea)