INTER-BENCHMARK PROTOCOL FOR IBERIAN SARDINE (IBPIS)

VOLUME 3 | ISSUE 108

ICES SCIENTIFIC REPORTS

RAPPORTS
SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA
CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

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

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

ISSN number: 2618-1371

ICES Scientific Reports

Volume 3 | Issue 108

INTER-BENCHMARK PROTOCOL FOR IBERIAN SARDINE (IBPIS)

Recommended format for purpose of citation:

ICES. 2021. Inter-Benchmark Protocol for Iberian sardine (IBPIS).
ICES Scientific Reports. 3:108. 57 pp. https://doi.org/10.17895/ices.pub.8144

Editors

Leire Ibaibarriaga

Authors

Pedro Amorim • Maria Manuel Angelico • Manuela Azevedo • Aaron Berger • Pablo Carrera • Susana Garrido • Leire Ibaibarriaga • Kelli F. Johnson • David Miller • Cristina Nunes • Isabel Riveiro • Andrés Uriarte • Laura Wise

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

Contents

i Executive summary

The Inter-Benchmark Protocol for Iberian sardine (IBPIS) met to revise the available autumn acoustic surveys and to identify a potential recruitment index to be included in the stock assessment. The stock assessment model configuration and the short-term forecast assumptions were updated and the implications for the Precautionary Approach (PA) and Maximum Sustainable Yield (MSY) reference points and for the evaluation of the Harvest Control Rule (HCR) were investigated. The Age 0 estimates of the autumn acoustic surveys in 9aCN from 1997 onwards were considered an informative index of recruitment. The stock assessment model settings were modified for the inclusion of this recruitment index. On the one hand, the catchability of the recruitment index was modelled with a power function, and on the other hand, an extra additive SD parameter was included for all the abundance indices (Acoustic, Daily Egg Production Method and recruitment). Other settings such as the input standard deviation of log number of recruits (sigmaR) and the recruitment deviations were also fine-tuned. The short-term forecast assumptions were updated to account for the interim year recruitment newly estimated in the stock assessment. Recruitment in the interim year was changed to be the value estimated in the assessment and the recruitment in the management year was the geometric mean of the last five years (now, including the interim year estimate). With these new stock assessment and short-term forecast procedures, the prediction skill of the recruitment in the interim year and of SSB in the management year improved by 50% and 46%, respectively. Since the stock assessment results in terms of SSB, recruitment and F were very similar to those from the previous model, the group considered that such small differences would likely have a very minor impact on the reference points and on the evaluation of the HCR. The group recommends continuing monitoring potential recruitment areas other than 9aCN for future inclusion in the recruitment index.

ii Expert group information

Expert group name	Inter-Benchmark Protocol for Iberian sardine (IBPIS)
Expert group cycle	Annual
Year cycle started	2021
Reporting year in cycle	1/1
Chair(s)	Leire Ibaibarriaga, Spain
Meeting venue(s) and dates	25-29 October 2021, Online meeting (13 participants)

1 Introduction

1.1 Terms of reference

The Inter-Benchmark Protocol for Iberian sardine (IBPIS), chaired by Leire Ibaibarriaga, Spain, and reviewed by two invited external experts, Kelli Faye Johnson, USA and Aaron Berger, USA, will be established and work by correspondence as well as meet online from 25 – 29 October 2021 to:

a) Investigate the inclusion of data from autumn acoustic surveys (recruitment index) in the SS3 assessment model currently used for the stock. Where necessary revisit any other aspect of the model configuration that may be impacted by the addition of these data.

b) If new standard analysis methods are proposed:

   i. Update the short-term forecast assumption on recruitment.

   ii. Re-examine and update (if necessary) MSY and PA reference points according to ICES guidelines.

   iii. Re-run MSE analyses to see if the HCR evaluated in 2021 is still precautionary given the new assessment.

c) Agree to, and document the preferred method for evaluating the status of the stock and short-term forecast and update the stock annex as appropriate.

Stocks	Stock coordinator	Stock assessor
Sardine (*Sardina pilchardus*) in divisions 8.c and 9.a (Cantabrian Sea and Atlantic Iberian waters)	Isabel Riveiro	Laura Wise

The Inter-Benchmark Workshop will report by 15 November 2021 for the attention of ACOM and the Benchmark Oversight Group.

1.2 Conduct of the meeting

The list of participants is presented in Annex 1. Before the meeting a working document (Wise et al.) as listed in Annex 3 was made available to all the participants. This working document served to initiate the discussions and provided the basis for all the work conducted during the meeting and presented in this report.

The meeting was organised with presentations and discussions for each of the ToRs. First, autumn acoustic surveys were reviewed, and potential recruitment index time-series were identified and evaluated. Then, the inclusion of the recruitment index in the current stock assessment model was investigated. Other model configurations such as the fishery selectivity pattern were also revisited. Given the new stock assessment model configuration, the short-term forecast assumptions on recruitment were updated. Finally, the MSY and PA reference points were re-examined, and the validity of the HCR evaluated in 2021 was discussed.

1.3 Report structure

The report is structured according to the ToRs. Autumn acoustic surveys to obtain a recruitment index are presented in Section 2 whereas the changes in the current stock assessment to include the recruitment index are discussed in Section 3 (ToR a). The updated short-term forecast assumptions are described in Section 4 (ToR b.i). The improvement in the short-term forecast is quantified in Section 5. MSY and PA reference points according to ICES guidelines are re-examined (ToR b.ii) in Section 6 and the validity of the HCR evaluation conducted in 2021 (ToR b.iii) is discussed in Section 7. Future work identified during the meeting is summarised in Section 8 and have been included in the online tool for future benchmarks' issues. Finally, the main conclusions from the meeting are listed in Section 9 and the external reviewer's report is given in Section 10.

The updated stock annex for Iberian sardine is included as Annex 4.

2 Recruitment index from autumn acoustic surveys

The biomass of the sardine stock in ICES divisions 8c and 9a has shown wide fluctuations in abundance throughout the historical time-series (Figure 2.1a), as is frequent in populations of small pelagic fish (Checkley et al., 2009). The absence of good recruitments (Figure 2.1b) has been the limiting factor in the recent period (2006-2018), in which the historical minimum biomass has been reached in 2015 (ICES, 2021a).

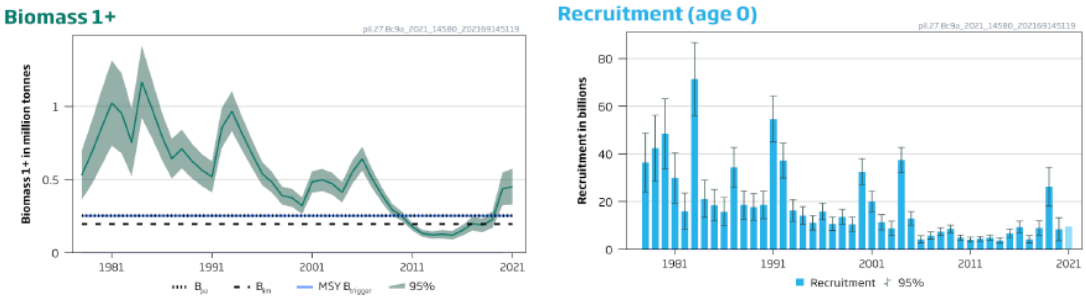


Figure 2.1. Summary of 2021 stock assessment a) Biomass, b) Recruitment (ICES, 2021a).

This sardine population is highly dependent on the strength of the incoming year classes. Since small pelagic species such as sardine individuals mature early in life (individuals are mature at age 1) the impact in the stock biomass is fast and high. Age 1 (recruitment from the previous year) constitutes a very important part of the population in the acoustic spring surveys and even age 0 represents a considerable proportion of the annual catches (Figure 2.2).

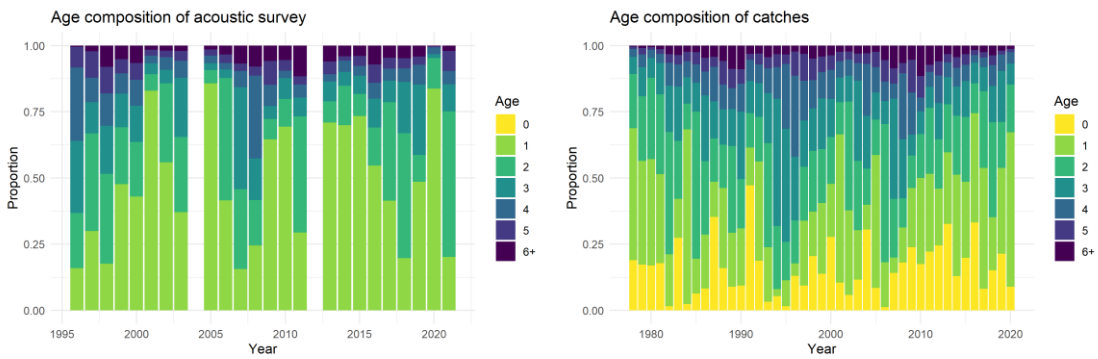


Figure 2.2. Proportion at age in a) acoustic spring surveys and b) catches.

The current short-term forecast assumes that recruitment for the interim year and the subsequent two forecast years is the geometric mean of recruitment of the last five years in the assessment. Knowledge of the incoming recruitment that is subject to high interannual variability could allow a better assessment of the population in the year of the advice.

In the case of the Iberian sardine there is a time-series of autumn acoustic surveys that can provide data on recruitment in the interim year and is not yet included in the assessment model. The inclusion of a recruitment index would provide an estimate of recruitment in the interim year that is not provided by any other source of information entering the model. This could be crucial to improve the short-term catch advice, but it is also expected to improve the estimates of past recruitments in general. This potential inclusion was one of the reasons for changing the advice calendar back in 2019, when the assessment was moved from June to November.

2.1 Overview of autumn acoustic surveys

According to Rodríguez-Climent et al. (2017), the three core areas of juvenile distribution are: Northern Portuguese shelf (centred off Aveiro), the coastal region in the vicinity of the Tagus estuary, and the eastern Gulf of Cádiz. See also Silva et al. (2009). The west coast (9a Central-North subdivision) is considered to be the main recruitment area, with a secondary area of importance in the Gulf of Cádiz, 9a South-Spain subdivision (Figure 2.3). Although the Cantabrian Sea is also a spawning area, 8c division is not considered important for recruitment.

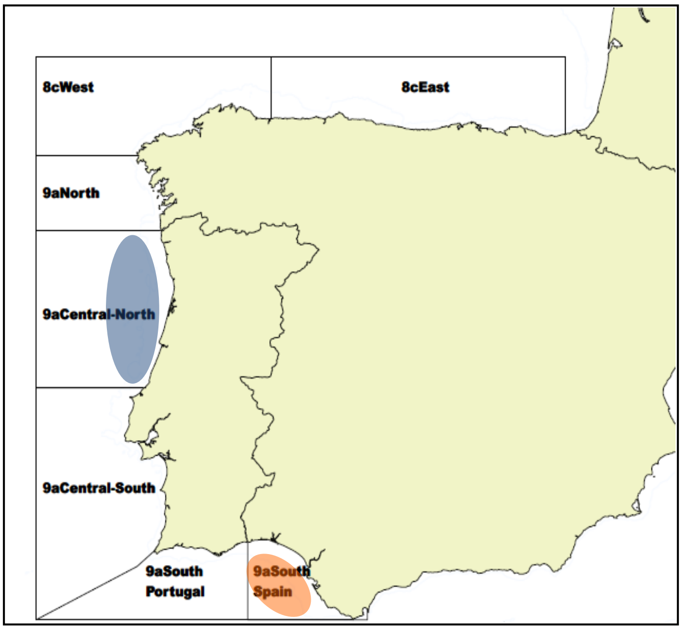


Figure 2.3. Main (blue) and secondary (red) recruitment areas of Iberian sardine stock.

Over the last decades, several autumn acoustic surveys have been carried out in the Iberian sardine stock area with the objective of assessing the incoming recruitment to the fishery.

SAR-PT-AUT

The SAR-PT-AUT survey time-series started in 1984 covering the Portuguese continental shelf. It was interrupted in 1988 and resumed with a systematic periodicity between 1997 and 2008 that extended the coverage to the Gulf of Cádiz. A one-time survey was carried out in 1992 covering also the Spanish Gulf of Cádiz. The acoustic methodology was the echo integration of the acoustic targets and fish schools, surveyed along-transects. The earlier surveys design until the 1992 survey, varied between “zig-zag” and parallel transects 10 nautical miles (nm) apart. In 1997 the survey design changed to acoustic transects perpendicular to the coastline spaced ca. 8 nm. Surveys were conducted both day and night. For acoustic data collection several echosounder models were used through this time-series: Simrad EKS38 between 1984 and 1988, Simrad EK400 in 1992 and Simrad EK500 since 1997. For species identification and biological sampling, trawl hauls were performed with a pelagic or bottom trawlnet. These surveys have had a different spatial coverage and seasonality, but have always covered the main area of juvenile concentration of the stock (subdivision 9a Central North).

JUVESAR

The JUVESAR autumn survey series started in 2013, ended in 2017 and only covered subdivisions 9aCN and part of 9aCS. JUVESAR was always done in the R/V Noruega. In 2017, one purse-seiner also participated and did additional fishing stations. Only biological sampling was collected. The work area ranged from Povoa do Varzim in the subdivision 9aCN and Cape Espichel in subdivision 9aCS, from shoreline (12 m) to the 60-100 m isobath over an adaptive grid with tracks spaced 4 or 8 nm (4nm in the main sardine recruitment areas). In 2014, due to bad weather, only a small area was covered. The methodology was similar to that of the spring acoustic surveys PELAGO. Acoustic equipment consisted of a Simrad EK-500 scientific echosounder, operating at 38 and 120 kHz. The backscattering acoustic energy from marine organisms was measured continuously during daylight. Pelagic or bottom trawls were carried out whenever possible to help identify the species (and size classes) that reflect the acoustic energy. This series provides the size composition (Length Frequency Distribution, LFD) and estimates of numbers and biomass for age 0 sardine and anchovy.

IBERAS

The IBERAS time-series started in 2018 and extended the JUVESAR surveyed area southwards. IBERAS' main objective is to get a recruitment index for both sardine and anchovy in Atlantic waters of the Iberian Peninsula, aiming to improve the estimation of the strength of the recruitment of the Iberian sardine and the western component of the southern anchovy population.

The work area ranged from Finisterra cape (in 2020 from Estaca de Bares cape) until Sao Vicente cape, from shoreline (20 m) to 100 m isobath over an adaptive grid (tracks were enlarged or shortened accordingly) with tracks distanced between 4-8 nm on account the potential recruitment distribution area of sardine. This series provides the size composition and age-structure of the estimated population in numbers and biomass for anchovy and sardine, in particular of age 0 individuals. The methodology is similar to that of the previous surveys and is summarised in Doray et al., 2021. Acoustic equipment consisted of a Simrad EK-80 scientific echosounder, operating at 18, 38, 70, 120 and 200 kHz. The backscattering acoustic energy from marine organisms was measured continuously during daylight except in the northern area where some tracks were steamed at night. Pelagic trawls were carried out whenever possible to help identify the species (and size classes) that reflect the acoustic energy.
In the first year the survey was undertaken in November but due to bad weather conditions, the aggregation and distribution patterns of the fish were anomalous and to improve the precision of the biomass estimates, the timing of the survey was changed. From 2019 onwards the survey was shifted to September which also allows a synoptic coverage of the Iberian Peninsula at the end of summer, beginning of autumn, since this is also the time period when the JUVENA acoustic survey takes place (in the Cantabrian Sea) and still close to the ECOCADIZ-reclutas survey date in October.

ECOCADIZ-RECLUTAS

ECOCADIZ-RECLUTAS is a survey restricted to subdivision 9aS (20 – 200 m depth), conducted over a systematic grid with tracks distanced 8 nm. The survey is conducted during the second fortnight of October, started in 2012 (only Spanish waters sampled) and continued in 2014. A serious breakdown in the RV's propeller system prevented from deriving an acoustic estimate from the 2017 survey. In 2018 ECOCADIZ-RECLUTAS has also experienced methodological problems related with the acoustic sampling coverage (ping rate), meaning that it should be carefully taken into account when dealing with the final acoustic estimates and interpreting their trends. This series provides the size composition (LFD) and age-structure of the estimated population in numbers and biomass for anchovy and sardine.

Figure 2.4 summarizes the timing and the area covered by each of the autumn surveys carried out so far and Figure 2.5 shows the contribution of each subdivision to the recruitment index (age 0). Although 9aCN subdivision is considered to be the main recruitment area, other subdivisions, such as 9aS-Cádiz, represent an important contribution to the index in some years. However, in years of high recruitment, i.e. 2019, the largest proportion corresponds to 9aCN.

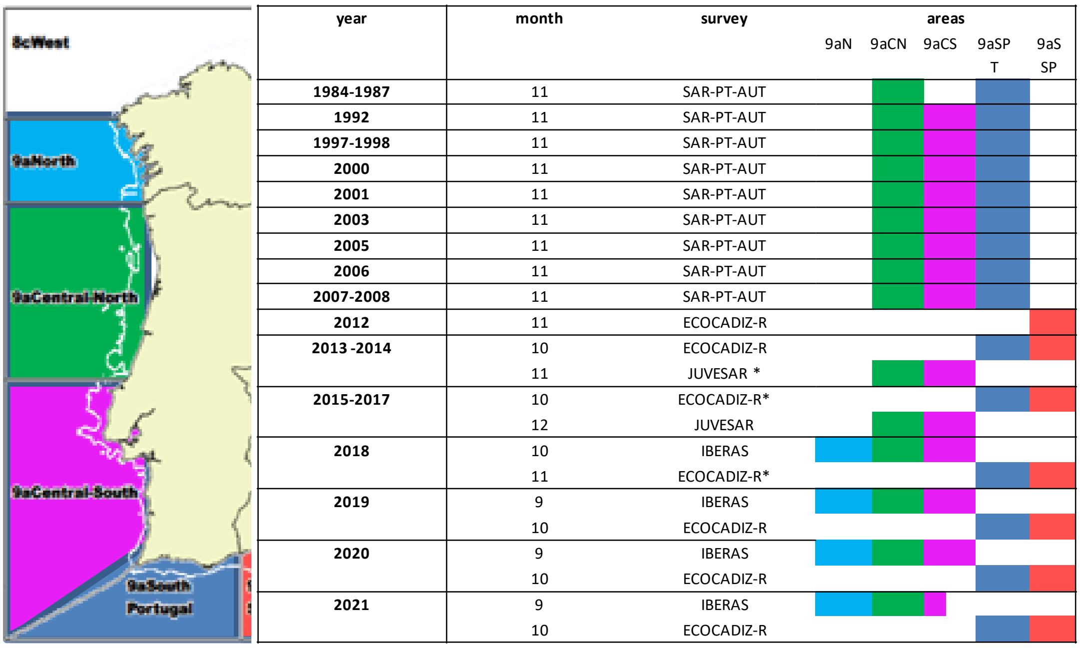


Figure 2.4. Timing and area of the autumn acoustic surveys. *In 2014 there are no JUVESAR acoustic estimates due to methodological problems. *In 2017 there are no ECOCADIZ-R acoustic estimates due to methodological problems. *In 2018 ECOCADIZ-R estimates should be taken with caution due to methodological problems. *In 2021, IBERAS only covers part of the 9aCS subdivision.

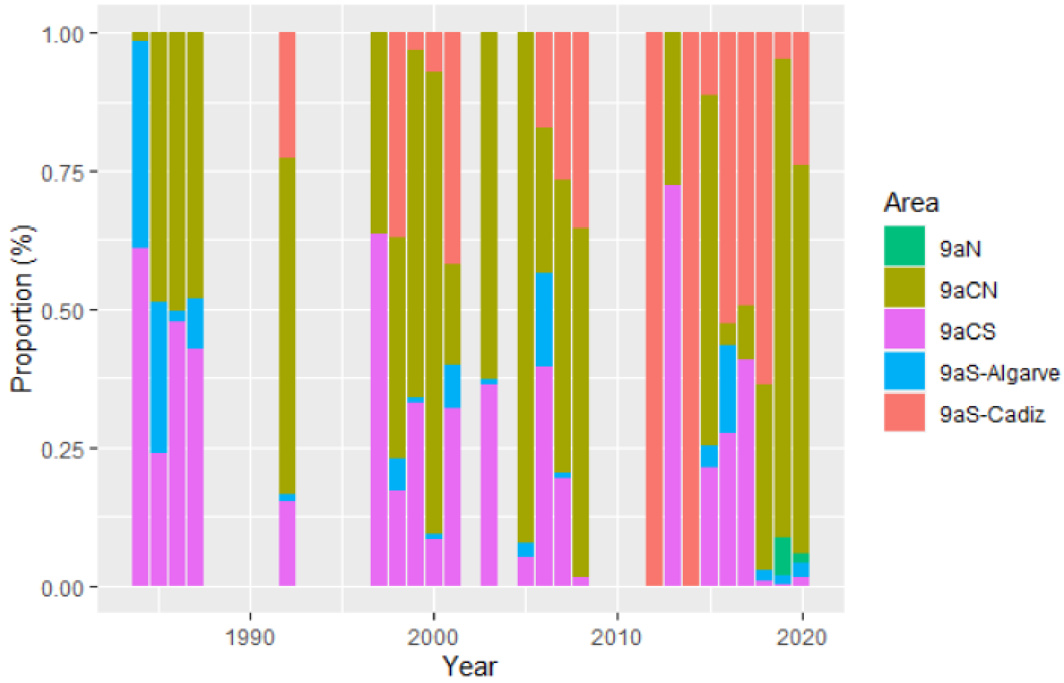


Figure 2.5. Contribution of each area (proportion) to the recruitment index.

During the meeting, the potential recruitment index was analysed by periods, accounting for the methodological changes that occurred throughout the historical series of the autumn acoustic surveys. Three time-series lengths were considered: 1984-2020, 1997-2020 and 2013-2020 (Table 2.1). During the first years (1984-1992) of the SAR-PT-AUT survey, the methodology was still being developed and the design was different. So, the group considered that these years should be considered separately. However, since 1997 (and coinciding with the same criteria applied in the spring acoustic surveys), the methodology was standardised and consolidated. The break in 2013 was initially suggested mainly due to the time gap and the resume of the surveys, but the group considered this should not imply any major changes in catchability. Differences in the echosounder should not have a major effect because the acoustic equipment is calibrated before each survey. Other potential effects as vessel changes could not be evaluated as no intercalibrations have been done. Overall, the group considered that there was no methodological change in the autumn acoustic surveys that prevented the use of the 1997-2020 time-series.

Table 2.1. Methodological changes through the juvenile index time-series.

Years	Surveys	Comments
1984-1987, 1992	SAR-PT-AUT	Zig-zag design, EKS38 and EK400
1997-2008 (gaps in 1999, 2002, 2004)	SAR-PT-AUT	Change survey design to parallel transects 8nm apart, Simrad EK500
2013-2017	JUVESAR	parallel transects 8nm or 4nm apart, Simrad EK500
2018-2020	IBERAS	Since 2019 surveys carried out in September (instead of October-November),

	parallel transects 8 nm or 4 nm apart, Simrad EK80

2.2 Agreement between autumn and spring surveys

The current assessment model uses the spring acoustic surveys PELAGO and PELACUS (covering all the stock area) as a joint index of abundance (from Age 1 to Age 6+). In order to check the validity of the autumn acoustic surveys to provide a recruitment index, the group compared the numbers-at-age 0 in the autumn recruitment surveys (SAR-PT-AUT, JUVESAR, IBERAS and ECOCADIZ-RECLUTAS) and the numbers-at-age 1 estimated in the spring surveys (PELACUS and PELAGO) of the following year and analysed if there was a significant positive correlation (Pearson correlation) between estimates in log scale. The correlation analysis was done for recruitment estimates from:

a) all autumn acoustic surveys (SAR-PT-AUT, JUVESAR, IBERAS and ECOCADIZ-RECLUTAS),
b) all autumn acoustic surveys with the exception of ECOCADIZ-RECLUTAS,
c) common areas in the western continental area of the stock (9aCN and partially 9aCS), and
d) only from the main recruitment area of the stock (9aCN).

The correlation was tested for the three time-series described above (Section 2.1): 1984-2020, 1997-2020 and 2013-2020.

Table 2.2. Pearson correlation between numbers-at-age 0 in the autumn recruitment surveys and numbers-at-age 1 in the spring acoustic surveys of the following year (in log scale).

Autumn surveys	Years	Correlation	df	p-value	Figure
All surveys	1984-2020	0.432	20	0.045	
	1997-2020	0.651	16	0.003	2.2.1
	2013-2020	0.399	6	**0.327**	
Without	1984-2020	0.564	18	0.0096	
ECOCADIZ-R	1997-2020	0.851	14	2.9 e-05	2.2.2.
	2013-2020	0.769	5	0.043	
9aCN & 9aCS	1984-2020	0.559	18	0.010	
	1997-2020	0.843	14	4.05 e-05	2.2.3
	2013-2020	0.758	5	0.048	
9aCN	1984-2020	0.561	18	0.0099	
	1997-2020	0.859	14	1.98 e-05	2.2.4
	2013-2020	0.864	5	0.012	

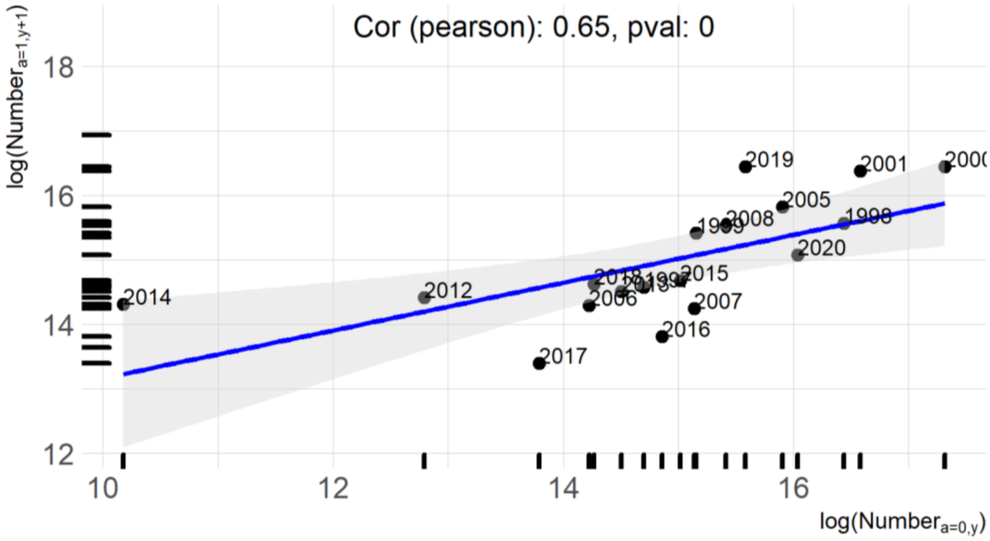


Figure 2.6. ALL SURVEYS. Pearson correlation between recruits estimated in year y with the number of individuals estimated in spring acoustic surveys of the year y+1 (PELACUS + PELAGO). Time-series restricted to 1997-2020. 2014 survey data not included since JUVESAR covered only a small area

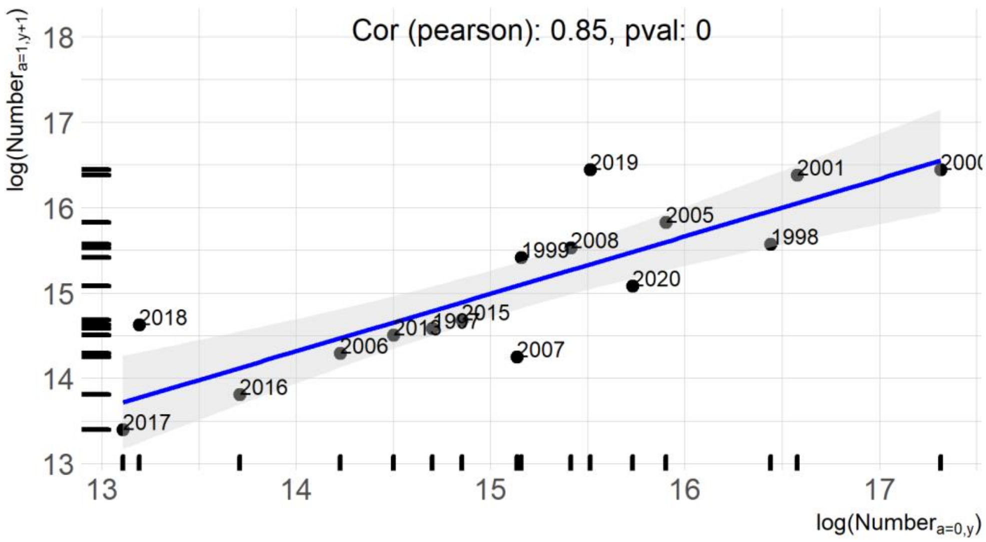


Figure 2.7. WITHOUT ECOCADIZ-RECLUTAS. Pearson correlation between recruits estimated in year y with the number of individuals estimated in spring acoustic surveys of the year y+1(PELACUS + PELAGO). Time-series restricted to 1997-2020. 2014 survey data not included since JUVESAR covered only a small area.

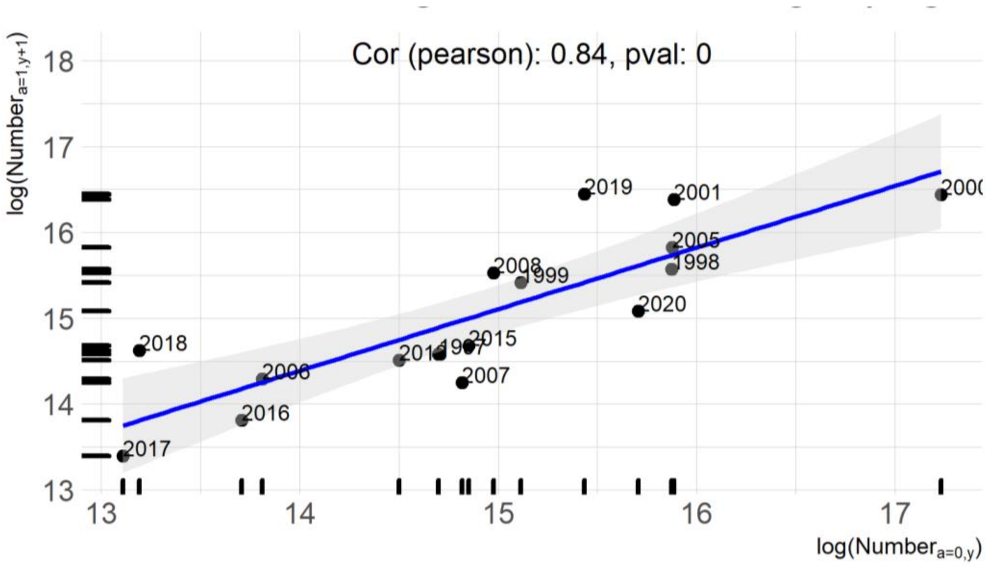


Figure 2.8. 9aCN & 9aCS. Pearson correlation between recruits estimated in year y with the number of individuals estimated in spring acoustic surveys of the year y+1 (PELACUS + PELAGO). Time-series restricted to 1997-2020. 2014 survey data not included since JUVESAR covered only a small area.

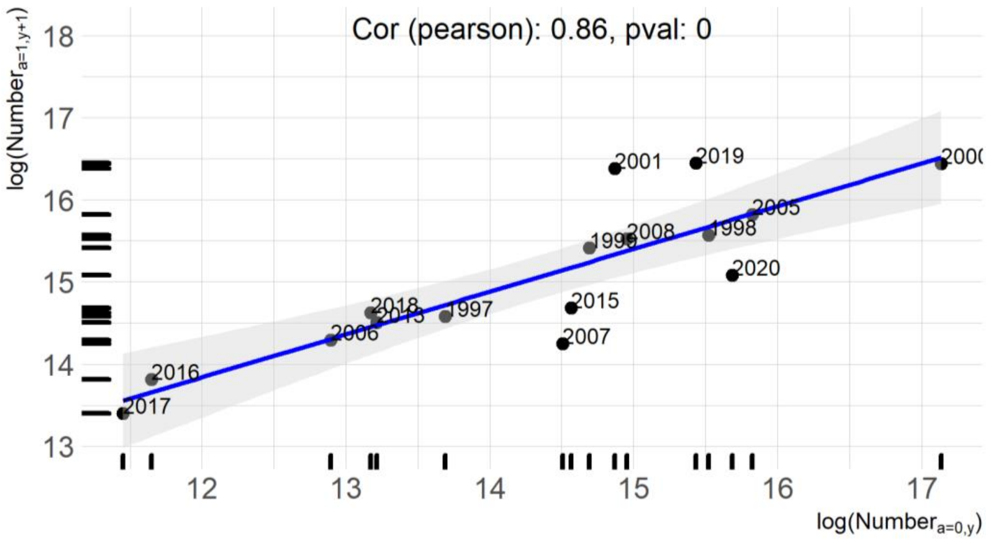


Figure 2.9. 9aCN. Pearson correlation between recruits estimated in year y with the number of individuals estimated in spring acoustic surveys of the year y+1 (PELACUS + PELAGO). Time-series restricted to 1997-2020. 2014 survey data not included since JUVESAR covered only a small area.

All correlations between Age 0 (year y) and Age 1 (year y+1) estimates (in log scale) were significant except for Age 0 estimates from all areas for the shortest time-series 2013-2020 (Table 2.2).

When the recruitment index included all areas, there was a significant positive correlation for both time-series starting in 1984 and in 1997. In this case 9aS was included whenever the SAR-PT-AUT survey was carried out. In some years it covered the Gulf of Cádiz and in others it did not. The difference with the time-series only since 2013 is that the information for 9aS comes from the ECOCADIZ-RECLUTAS survey.

In terms of correlation, the highest difference was found when the ECOCADIZ-RECLUTAS survey covering 9aS subdivision was included. Although the highest correlations were found when only the 9aCN area was considered, the differences between the other three options that excluded the ECOCADIZ-RECLUTAS survey were very small (Table 2.2). Regarding the time-series length, correlation was greater for the time-series starting in 1997 or 2013 than for the whole time-series (1984-2020) but the number of observations was smaller. The Pearson correlation between recruits estimated in year y and the number of individuals estimated in spring acoustic surveys of the year y+1 (PELACUS + PELAGO) for the 1997-2020 time-series are shown in Figures 2.6 to 2.9 for different areas.

Correlation analyses can be hindered when the analysed time-series are strongly autocorrelated (Pyper and Peterman, 1997; Probst et al., 2012). To check the validity of the correlation results presented above, the group analysed the autocorrelation of the time-series for the case of age 0 abundance in subdivision 9aCN in 1997-2020. The standardised time-series of the age 0 abundance estimates in 9aCN from the autumn acoustic surveys and of the age 1 estimates from the spring acoustic surveys showed very similar trends (Figure 2.10). None of time-series depicted any autocorrelation structure (Figure 2.11). This can be in part due to the gaps and the short length of the recruitment time-series, but the conclusions from the correlation analyses were considered valid. Overall, the high significant correlation found supported the inclusion of the recruitment survey series in the Iberian sardine stock assessment model.

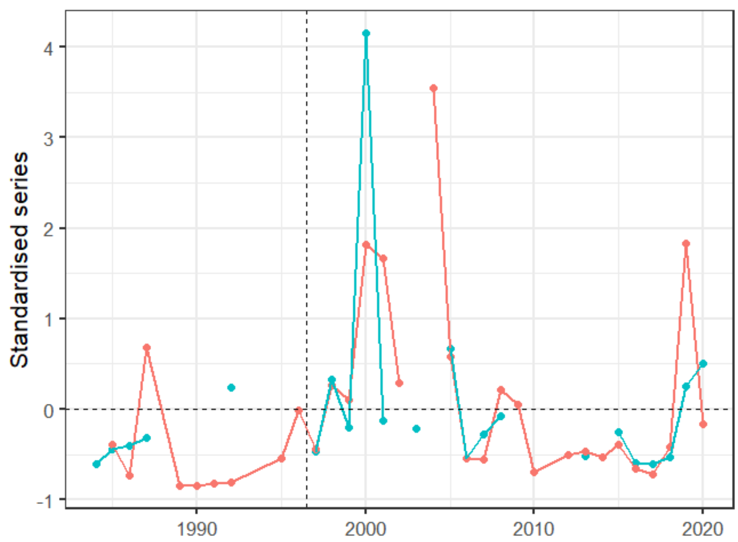


Figure 2.10 Standardised time-series of the age 0 abundance estimates in 9aCN from the autumn acoustic surveys (in blue) and of the age 1 estimates from the spring acoustic surveys (in red).

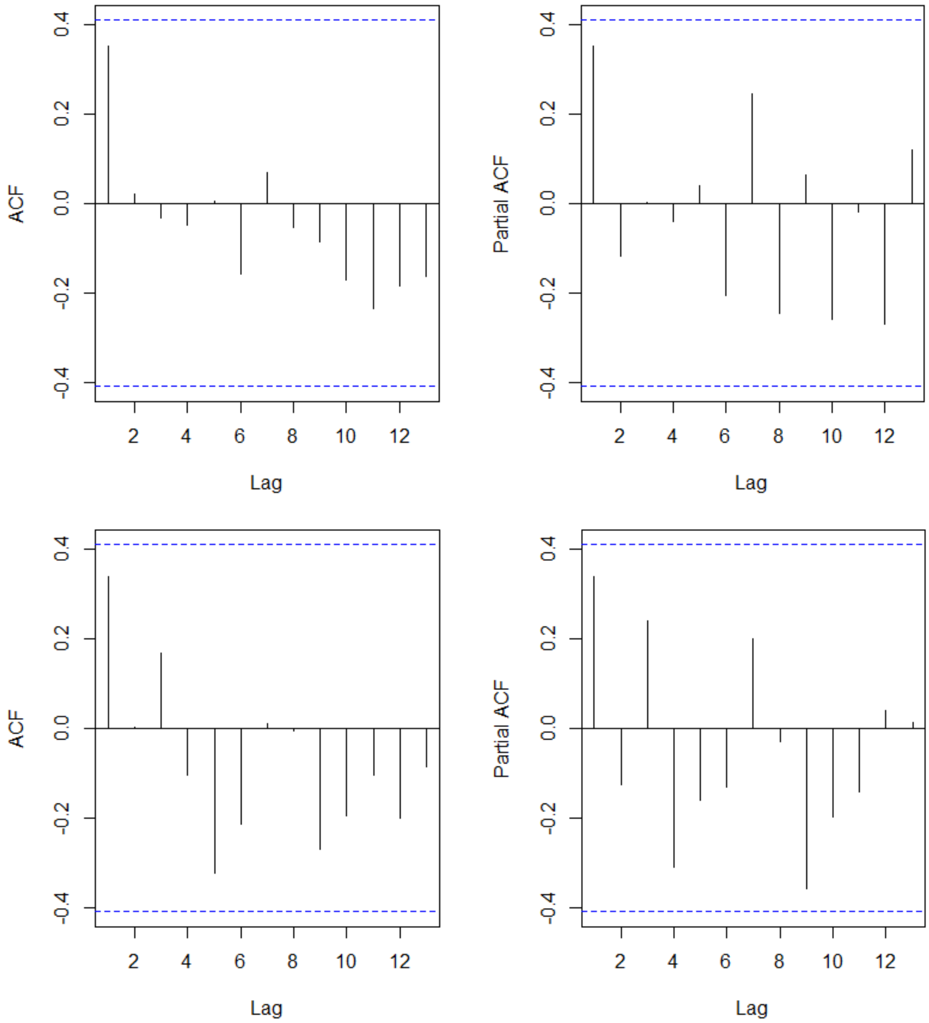


Figure 2.11. Correlation function (on the left) and partial autocorrelation function (on the right) for the logarithm of the age 0 abundance estimates in 9aCN from the autumn acoustic surveys (on the top) and for the logarithm of the age 1 estimates from the spring acoustic surveys (on the bottom).

2.3 Recruitment index for the stock assessment

Recruitment index areas

During the meeting, the group decided not to use age-0 estimates from ECOCADIZ-RECLUTAS in the recruitment index at this time. This survey covers the second recruitment area of the stock, but the correlation is much noisier that the combination of IBERAS-JUVESAR survey, because most of the recruitment is occurring in the northern part.

ECOCADIZ-RECLUTAS covers 9aSouth subdivision since 2012, but this survey had some methodological problems in 2017 and 2018. High recruitment estimates from ECOCADIZ-RECLUTAS do not seem to be reflected into next year abundances and there is no correlation between age 0 from ECOCADIZ-RECLUTAS and age 1 from the joint spring acoustic survey PELAGO-PELACUS. For that reason, the correlation between age 0 from all the autumn acoustic surveys and age 1 from the joint spring acoustic survey PELAGO-PELACUS gets worse when including the ECOCADIZ-RECLUTAS in the age 0 index. In addition, the recruitment hot spot in the eastern

Gulf of Cadiz seem to be not connected with the rest (Silva et al., 2019) and oceanographic conditions driving the recruitment distribution in this area are different from the Northern Portuguese shelf (Rodríguez-Climent et al., 2017).

The other core areas for juveniles have been covered by a series of surveys (SAR-PT, JUVESAR, IBERAS). The group agreed that the age 0 index should include at least 9aCN subdivision, but the inclusion of a larger area covering 9aCN+9aCS subdivision was also discussed. The group considered that ecologically the best option would be to include at least the northern part of 9aCS, where the bulk of the juvenile distribution within 9aCS occurs. The coverage of this northern part of 9aCS is warranted as the Spanish and Portuguese governments have agreed that the IBERAS survey (2018-onwards) should cover 9aN (Spanish waters), 9aCN (Portuguese waters) and northern part of 9aCS (Portuguese waters). However, due to vessel availability, right now it cannot be warranted that the whole 9aCS zone will be covered in the future. The irregular coverage of 9aCS area of the autumn acoustic surveys along the time-series supposes an additional challenge that couldn't be solved during WKIBPIS due to the lack of time and lack of data at the beginning of the time-series.

Based on these arguments, the group decided to use an index of the age 0 abundance from the autumn acoustic surveys in 9aCN only. However, the group recommends continuing monitoring the other potential areas (particularly 9aN and 9aCS), in order to detect spatial changes in the recruitment area that could trigger future changes of the autumn survey design.

Time-series length

Regarding the time-series length of the recruitment index, based on the methodological changes of the autumn acoustic surveys (Section 2.1) and the correlation analysis with the age 1 abundance estimates from the spring acoustic surveys (Section 2.2), the initial preference of the group was to use the 1997-2020 time-series. To study the impact of the recruitment time-series length on the stock assessment, the group carried out some preliminary runs including the three recruitment time-series (1984-2020, 1997-2020 and 2013-2020) and compared them to the 2020 model (ICES, 2020) and the 2020 model including the 2020 DEPM which was already included in the runs with the recruitment index. These runs are presented and discussed in Section 3.

3 Stock assessment

The current assessment model of the Iberian sardine stock (prior to interventions made at the inter-benchmark) uses Stock Synthesis (Methot and Wetzel, 2013) since 2012. From 2019 onwards the version used to run the assessment model is version 3.30. The model is age-based and is set up in a single fleet/single area configuration with a yearly season and genders combined. Numbers-at-age are available from commercial catches and a joint acoustic survey (PELAGO and PELACUS spring acoustic surveys) which provide a full coverage of the stock area. The model includes two survey indices (the annually joint spring acoustic survey and a joint triennial Daily Egg Production Method - DEPM - survey). Both surveys are used in the assessment as relative indices of abundance. Fecundity is provided as a matrix combining a maturity ogive and weight-at-age derived from the DEPM survey. Natural mortality rates at age are assumed constant between years and were derived using the method from Gislason et al. (2010) (0.7 times Gislason estimates). Surveys and commercial fleets selectivities are set as follows. Both the joint spring acoustic survey and the joint DEPM survey have selectivity set at 0 at age 0 and a flat selectivity set at 1 for ages 1 to 6+. Catchability is an estimated parameter for each survey. There is no prior distribution on selectivity parameters. The commercial fleet has a time-variant selectivity where three periods were defined (1978–1987, 1988–2005, 2006–final year of the assessment). The three periods may have different selectivity patterns, in all cases selectivity assumed to be fixed within each time period. The commercial fleet age selectivity is a flat selectivity at 1 for ages 3 to 5. Selectivities for ages 1, 2 and 6+ are estimated and catchability is specified for the commercial fleet as a scaling factor such that the estimate is median unbiased, similar to allowing catchability to float.

3.1 Inclusion of recruitment index

Up until 2019, the assessment of the state of the stock in the interim year (year ***y+1)*** and advice for the following year (year ***y+2***) would take place in June during WGHANSA-1. The current assessment model (prior to interventions made at the inter-benchmark), that includes fishery data up to year ***y*** (final year of the assessment) and acoustic data up to year ***y+1*** (interim year), has information on recruitment (at Age 0) up to the final year of the assessment (Figure 3.1). This recruitment estimate is supported by the catch-at-age 0 data in the final year of the assessment and by the acoustic survey that takes places in the first quarter of the interim year. The short-term forecast assumes that recruitment for the interim year and the subsequent two forecast years is the geometric mean of recruitment of the last five years in the assessment (***y-4,…,y***).

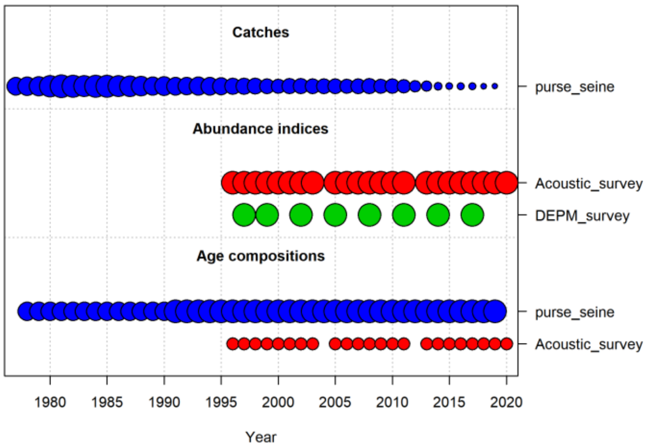


Figure 3.1 Data presence by year for each fleet and data type.

The inclusion of another source of information on recruitment is thought to improve the advice that is provided since small pelagic species such as sardine have highly interannual recruitment variability that have major impacts in the stock biomass. Sardine individuals mature early in life (individuals are mature at Age 1) so the impact in the stock biomass is fast and high. In the case of the Iberian sardine there is a time-series of autumn acoustic surveys (see Section 2) that can provide data on recruitment in the interim year and is not yet included in the assessment model. The inclusion of a recruitment index would provide an estimate of recruitment in the interim year that is not provided by any other source of information entering the model. This could be crucial to improve the short-term catch advice but it is also expected to improve the estimates of past recruitment in general. This potential inclusion was one of the reasons for changing the advice calendar back in 2019, when the assessment was moved from June to November (WGHANSA-2).

The most recent assessment of the Iberian sardine stock took place in June 2021 (ICES, 2021b; WGHANSA 2021) but models tested during the IBPIS2021 were compared with the previous assessment of the stock, fitted to data 1978-2020 (ICES, 2020; WGHANSA 2020). This option was preferred by the group since it mimics the future assessment process where in the interim year we have information from both the spring and the autumn acoustic surveys. The assessment takes place in November, during WGHANSA-2 and information from catches is provisional (only total catches, age composition in the interim year ***y+1*** is not known). At the time of the IBPIS2021 estimates from the 2021 autumn acoustic survey were not available since the survey ended on the 19th of October.

The primary objective of the inter-benchmark was to include the autumn acoustic survey data in the Iberian sardine model. This new index was included as an index of abundance with a selectivity tailored to young fish, where age selectivity options were used to choose a single age, age 0. The primary tools for evaluating a new assessment model that includes a recruitment index are series of standard diagnostics that included profiling key parameters (steepness), retrospective analyses, and evaluation of parameter standard errors, correlations, model fit and residual patterns. Diagnostics such as these can determine whether the assessment model has converged and estimated all parameters to a necessary degree of precision and is appropriately specified for and conditioned on the available data.

3.2 Testing alternative model settings

Evaluate recruitment index time-series length

A series of runs were brought forward during the inter-benchmark to examine the performance of the model with the new recruitment index. Settings for recruitment deviation were modified to accommodate for the new index series (last year of main recruitment deviations is now the interim year as opposed to the last year of catch data). All other settings were the same as the current assessment model.

Considering the agreement between the autumn and the spring acoustic surveys and the group decision regarding the area to estimate number of Age 0 from the autumn acoustic surveys (see Sections 2.2 and 2.3), a first set of runs was presented to evaluate the performance of the assessment when different lengths of the time-series are included. The 9aCN area has been covered by different surveys since 1984. The time-series considered for the recruitment index were: (i) long - the whole time-series, from 1984 up to 2020; (ii) medium - from 1997, when surveys start using the EK500, up to 2020; and (iii) short - from 2013, when autumn acoustic surveys are resumed, up to 2020.

Evaluate other selectivity settings

After the WG's selection of a preferred model formulation regarding the recruitment index data, additional explorations were considered regarding different selectivity settings for both the commercial fleet and the acoustic survey.

Table 3.1 summarises the current model assumptions for fishery selectivity and catchability (setup a) and the other two selectivity settings tested (setup b and setup c). Figure 3.2 shows the time and at age selectivity settings in the 3 set ups tested.

Table 3. 1. Model assumptions for fishery selectivity and catchability for the different setups.

Model assumptions	Description
Fishery selectivity-at-age	
Setup a and c	S-at-age are parameters, each estimated as a random walk from the previous age; S-at-age 0 used as the reference; S-at-ages 4 and 5 assumed to be equal to S-at-age 3.
Setup b	S-at-age are parameters, each estimated as a random walk from the previous age; S-at-age 0 used as the reference; S-at-ages 4, 5 and 6 assumed to be equal to S-at-age 3.
Time-varying selectivity	
Setup a	Three periods: 1978–1987, 1988–2005 and 2006–onwards. Selectivity-at-age is estimated for each period and within each period assumed to be fixed over time.
Setup b and c	Selectivity-at-age is estimated every year (1978-2020) for ages up to 2.
Fishery catchability	
All models	Scaling factor, median unbiased
Acoustic selectivity-at-age	
Setup a and c	S-at-age are parameters, each estimated as a random walk from the previous age; S-at-age 0 is set to be zero. S-at-age 1 used as reference.
Setup b	S-at-age are parameters, each estimated as a random walk from the previous age; S-at-age 0 is set to be zero. S-at-age 1 used as reference. Min and Max values differ from Setups a and b.
Acoustic and DEPM survey catchability	

Model assumptions	Description
Setup a	Parameter, mean unbiased
Setup b and c	An extra parameter is estimated that will contain an additive constant to be added to the input standard deviation of the survey variability.

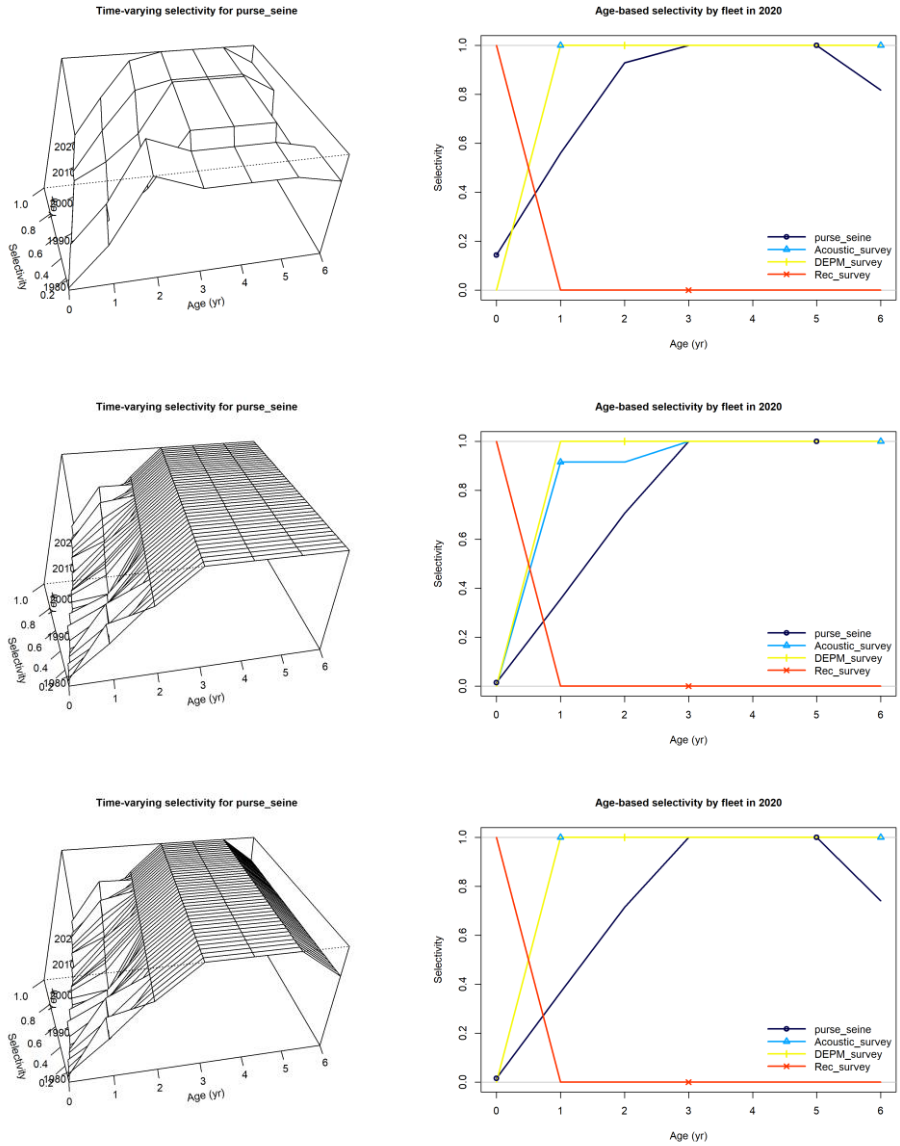


Figure 3.2 Time-varying selectivity surface for the purse-seine fleet (left panels) and selectivity at age for the multiple fleets in the final year (2020) (right panels). Top panels reflect the selectivity settings for setup a, middle panels reflect the selectivity settings for setup b and the bottom panels reflect the selectivity settings for setup c.

Evaluate other catchability settings

Once a preferred assessment model was selected by the WG, it was tested to evaluate the improvement of the assessment model when the catchability recruitment index was modeled as a power type instead of a linear relationship. Adjustments were also tested for all surveys included in the assessment model such as the inclusion of an extra standard deviation parameter for the catchability of the surveys.

With a small set of candidates assessment models a profiling of the steepness parameter was run and adjustments in both σ_R and the recruitment deviations were made. In the latter case the least squares estimate of alternative bias adjustment relationship for recruitment deviations settings followed the recommendations automatically estimated by SS (Methot and Taylor, 2011). Finally, retrospective patterns were estimated based on ICES guidelines using the Mohn's rho estimate (Mohn, 1999) and some diagnostics were assessed with the R package ‘ss3diags’ (Winker et al., 2020).

3.3 Results

All the runs brought forward by the group were compared with the official 2020 assessment (ICES, 2020), the ***current*** assessment model, but also with the official 2020 assessment ***updated*** to include the now available 2020 DEPM survey point estimate and the now known total catches of 2020. All the runs including the recruitment index also included the 2020 DEPM survey point estimate and the total catches of 2020.

Evaluate recruitment index time-series length

Figure 3.3. shows the spawning biomass, fishing mortality, Age-0 recruits and recruitment deviations outputs from all the runs. The likelihoods and AIC’s from the models fitted to different datasets cannot be directly compared (Table 3.1), however the longest time-series (***All***) was not considered a suitable candidate because during the first years (1984-1992) of the SAR-PT-AUT survey, the methodology was still being developed and the design was different (zig-zag design as opposed to the parallel design used since 1997). Since 1997, the methodology seems to be standardised and quite consistent, hence the group considered that it could be regarded as a unique index. The shorter time-series (***2013***-2020) encompasses a period with only low recruitments, which can affect the overall fit to the recruitment index. The group preferred to have information from the longer time-series (***1997***-2020) encompassing a larger range of recruitments. In general, the model doesn’t fit well high values of acoustic surveys and high values of the recruitment index (Figure 3.4.). The discrepancy observed between the high recruitment suggested by the autumn acoustic survey in 2020 was not confirmed by the 2021 spring acoustic survey (Age 1). In general, high recruitment values not being well fitted by the model and then not being confirmed by the spring surveys have been observed before in the time-series. It is not clear if this is due to catchability issues or high natural mortality events or something else. But, again, they are arguments favouring a longer time-series. Therefore, **the group decided to use the 1997-2020 time-series.**

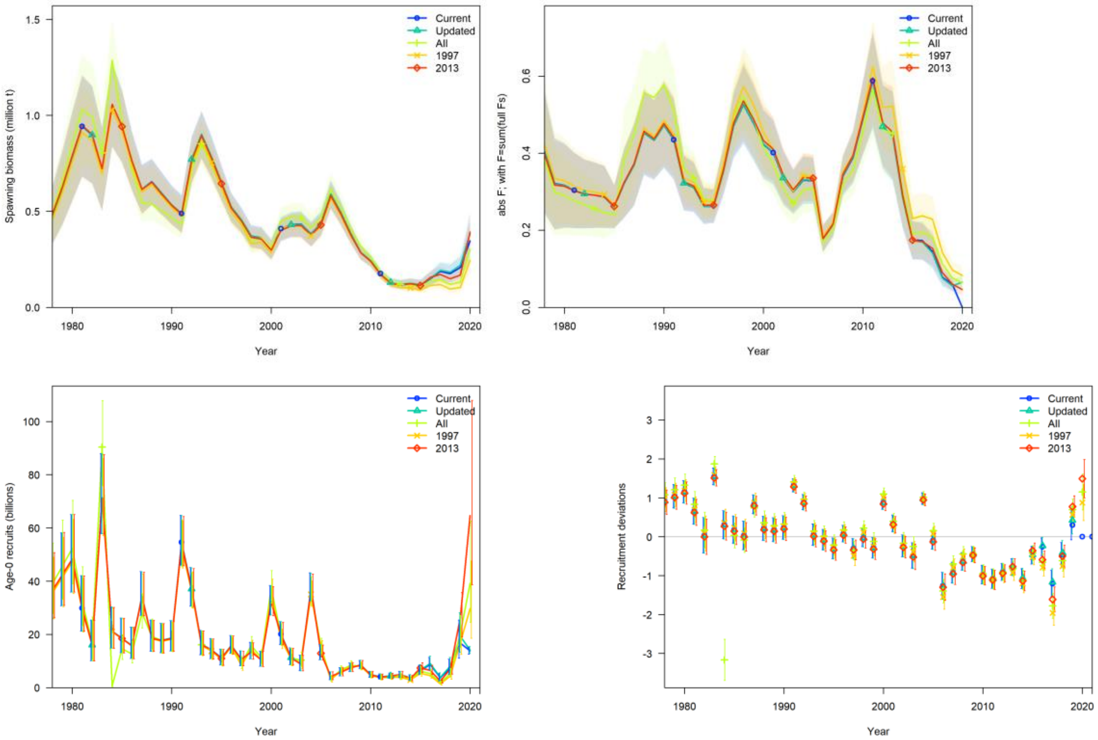


Figure 3.3. Spawning biomass (top left), fishing mortality (apical, top right), Age-0 recruits (bottom left) and recruitment deviation (bottom right) outputs from all the runs to evaluate the recruitment index time-series length.

Table 3.1 Log-likelihood, number of parameters, SSB, Recruitment, F, AIC and final gradient for the different runs (current assessment model, updated assessment model and runs with recruitment index estimated in 9aCN for the all time-series available, from 1997 and from 2013 onwards).

Label	Current	Updated	1984-2020	1997-2020	2013-2020
Total	135.14	134.89	301.31	228.17	180.16
Catch	1.21e-09	1.26e-09	1.53e-09	2.79e-09	1.44e-09
Equil_catch	0.407	0.356	1.521	0.856	0.232
Survey	-20.37	-20.83	62.08	44.11	7.84
Age_comp	131.36	131.93	195.26	152.45	144.55
Recruitment	23.7	23.4	42.4	30.7	27.5
Parm_softbounds	5.11e-04	5.11e-04	4.73e-04	5.03e-04	5.11e-04
Parm_devs	0	0	0	0	0
N parm	62	62	64	64	64
SSB_2020	346.83	384.88	302.57	245.87	393.38
Recr_2020	13.8	14.14	39.2	29.72	64.64
F_2019	0.06	0.06	0.08	0.1	0.06
AIC	394.28	393.79	730.62	584.34	488.32
Max Grad	6.29e-05	4.63e-05	1.92e-05	2.26e-05	4.26e-05

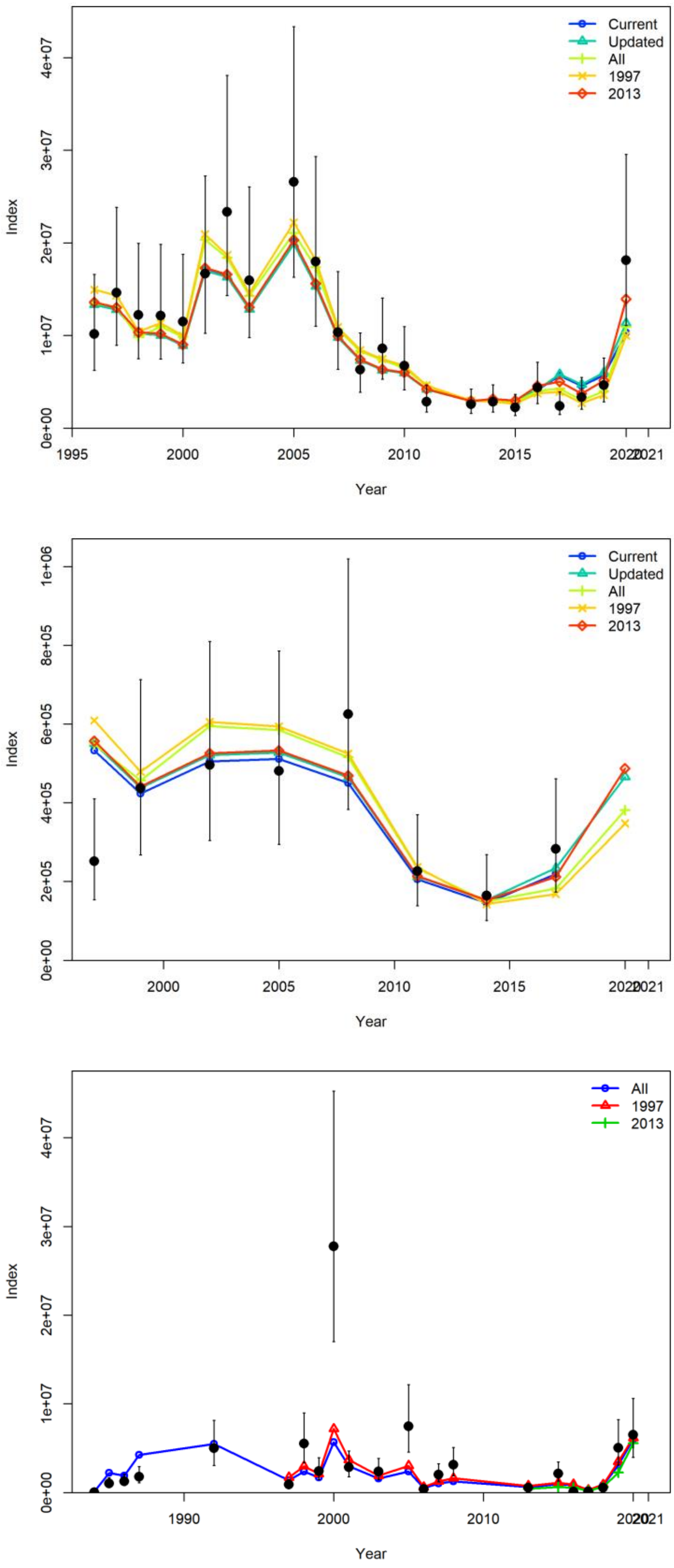


Figure 3.4. Fit for the different runs to the spring acoustic survey (top), to the DEPM survey (middle) and to the autumn acoustic survey (bottom).

Testing other selectivity settings

For this analysis we used the recruitment index series starting in 1997 and only for area 9aCN according to the group decisions so far. Three alternative model settings (***Setup a, b and c***) were evaluated depending on the selectivity of the fleet, the selectivity of the age-structured index of the acoustic survey and the inclusion of additional standard deviation parameter for catchability. Outputs from the different runs and model diagnostics are shown in Table 3.2. and Figures 3.5. to 3.7.

Table 3.2. Log-likelihood, number of parameters, SSB, Recruitment, F, AIC and final gradient for the runs with different selectivity settings.

Label	Current	Updated	SetupA	SetupB	SetupC
Total	135.14	134.89	228.17	139.5	138.66
Catch	1.21e-09	1.26e-09	2.79e-09	6.40e-15	5.88e-15
Equil_catch	0.407	0.356	0.856	0.025	0.017
Survey	-20.37	-20.83	44.11	22.68	22.5
Age_comp	131.36	131.93	152.45	93.87	92.79
Recruitment	23.7	23.4	30.7	14.9	15.3
Parm_softbounds	5.11e-04	5.11e-04	5.03e-04	7.32e-04	7.36e-04
Parm_devs	0	0	0	8.01	8.03
N parm	62	62	64	144	142
SSB_2020	346.83	384.88	245.87	258.04	267.26
Recr_2020	13.8	14.14	29.72	24.72	25.38
F_2019	0.06	0.06	0.1	0.28	0.26
AIC	394.28	393.79	584.34	566.99	561.31
Max Grad	6.29e-05	4.63e-05	2.26e-05	3e-05	2.17e-05

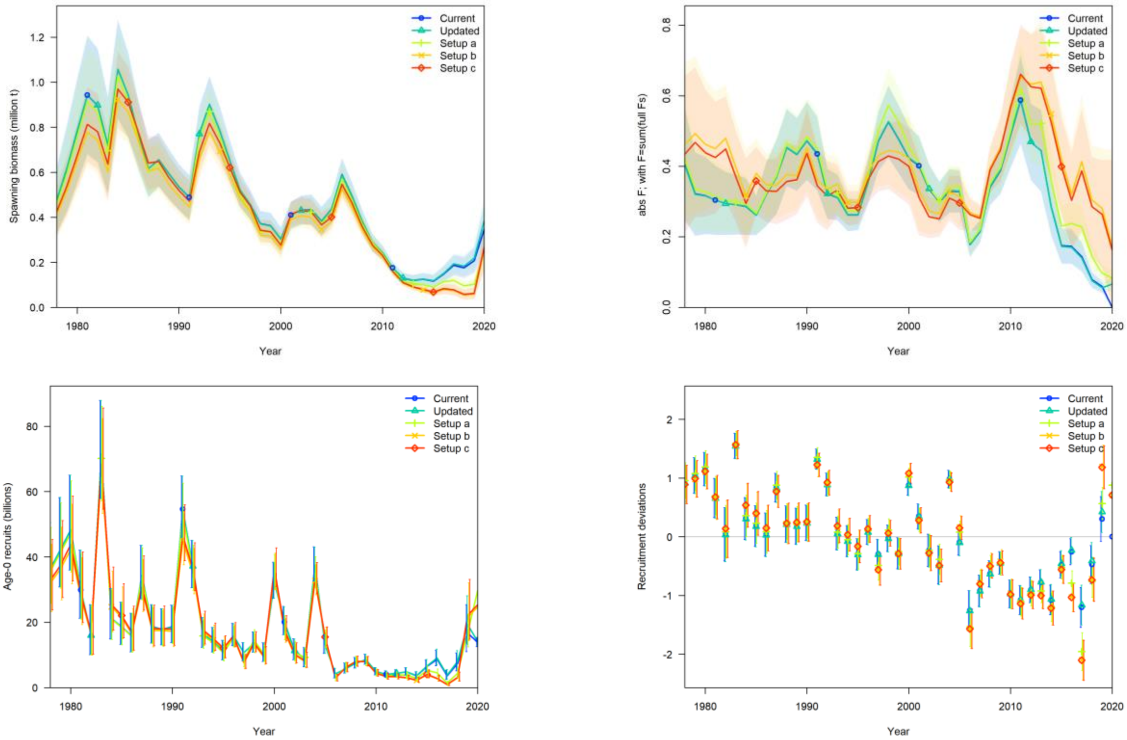


Figure 3.5. Spawning biomass (top left), fishing mortality (apical, top right), Age-0 recruits (bottom left) and recruitment deviation (bottom right) outputs from the runs to evaluate the selectivity settings.

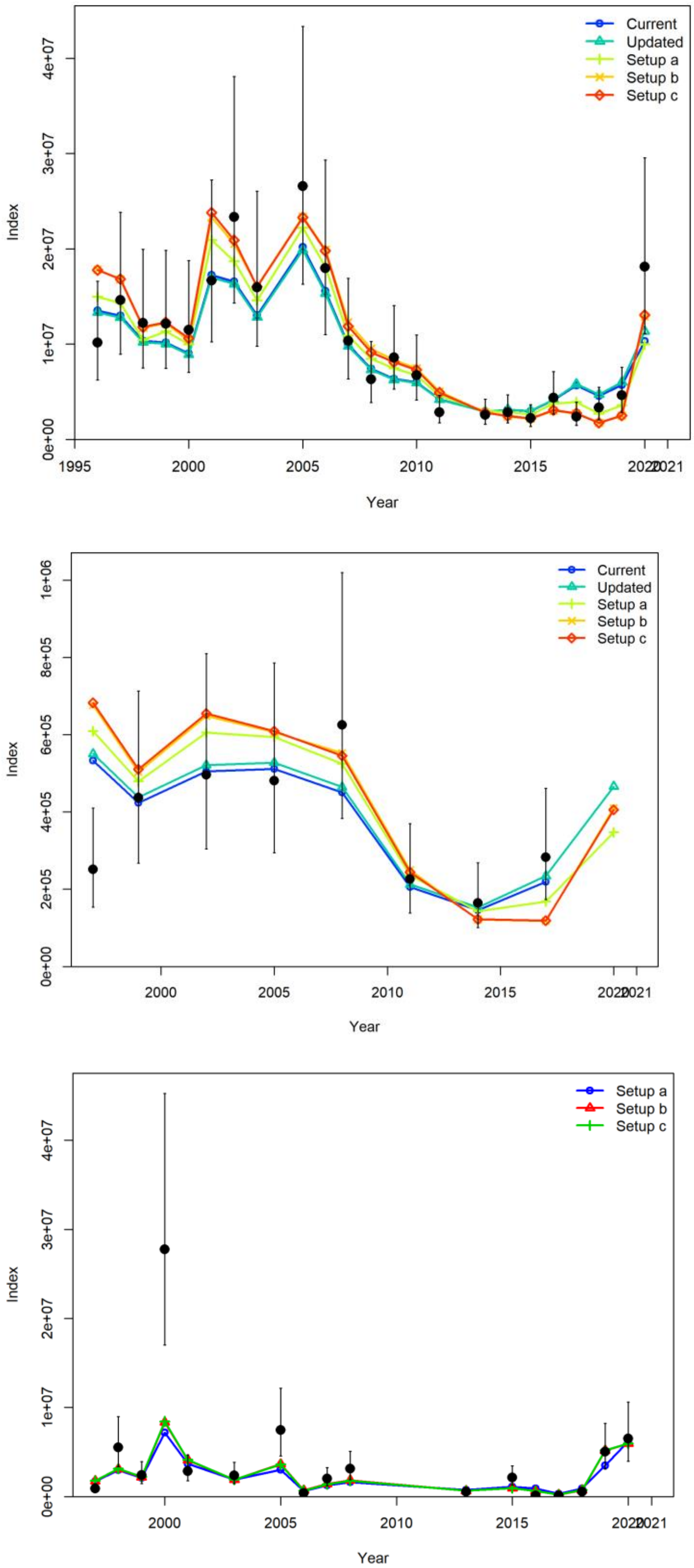


Figure 3.6. Fit to the spring acoustic survey (top), to the DEPM survey (middle) and to the autumn acoustic survey (bottom)

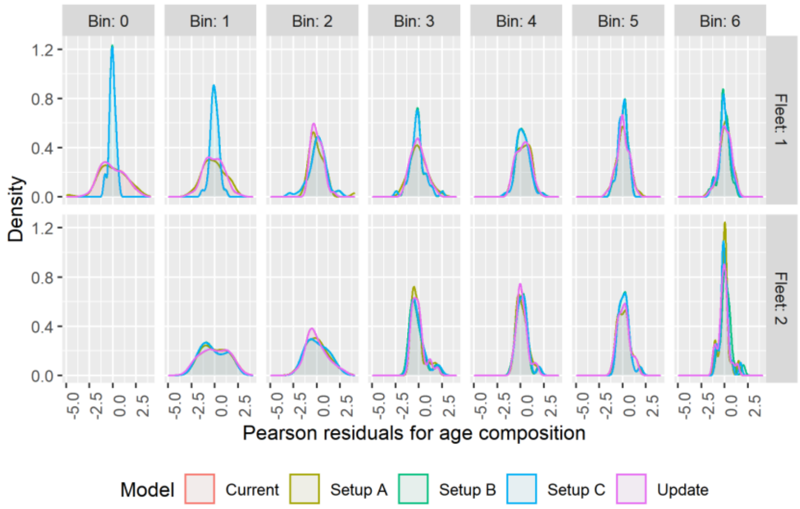


Figure 3.7. Density plot of Pearson residuals for age composition, comparing across fleets (the commercial fleet is Fleet 1, top panel, and the acoustic survey is Fleet 2, bottom panel) and runs (colours).

In general, there is no reason to believe the selectivity pattern of the purse-seine fleet is changing every year. The interannual changes showed by the time-varying selectivity could be reflecting biological or environmental variations. In addition, even if the model fit improves a bit with the time-varying selectivity, the changes in the model fit are not very big, and the diagnostics for the age composition residuals of the acoustic survey get worse in Setup b and c. There is also the suspicion that the big change in SSB from 2019 to 2020 could be an artefact of the annual time-varying selectivity. Therefore, **the group decided to keep the three selectivity blocks for the purse-seine fleet as in the current model** (***Setup a***).

The three selectivity blocks were defined in the previous benchmark (ICES, 2017; WKPELA) based on the percentage of the catches occurring in the recruitment areas (Figure 3.8.). This is more related to catchability than to selectivity. So, it reinforces the idea of keeping the three blocks instead of the annual time-varying selectivity.

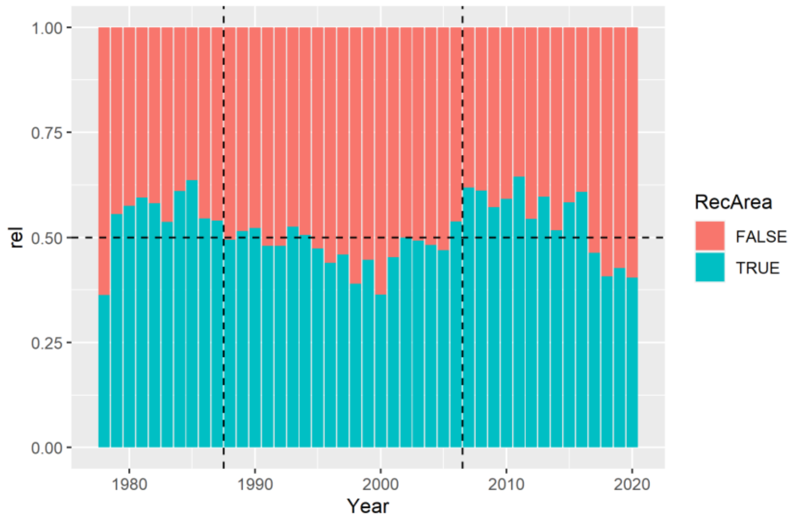


Figure 3.8. Percentage of sardine catches (in biomass) from recruitment areas (9aN + 9aCN + 9aS-Cadis in blue) and non-recruitment areas (8c+9aCS+9aS-Algarve in red).

The selection pattern of the acoustic survey was kept as before (i.e. 0 for Age 0 and 1 for all the other Ages). When allowing the model to estimate the selection for ages 1 and 2 (setup b), the estimated values were high (above 0.9), so the impact expected is to be low.

In summary, the group decided to keep the selectivity settings of the commercial fleet and the acoustic survey as in the current model (setup a).

Evaluate other catchability settings

In the previous set of runs an extra SD parameter was included for both the DEPM and the spring acoustic surveys in ***Setup b*** and ***Setup c***. During the inter-benchmark it was suggested to add this extra standard deviation parameter for the autumn acoustic survey (***Setup aSD***). Additionally, it was also questioned how the autumn acoustic survey should be included in the assessment. In the previous set of runs, the autumn acoustic survey was included in the same way as the other surveys, with a constant catchability that is estimated by the model, mean unbiased with lognormal error (option 1 of SS on how to model catchability). It was suggested that a power function would be more appropriate to model the catchability of the autumn acoustic surveys (option 3, Q with power, on SS). This option would be more in accordance with the correlation analyses done previously (see Section 2.2.) and maybe could accommodate better the high recruitments observed with the model estimates (***Setup aSDQ***). The additional parameter needed would reflect the change from the expected abundance to the vulnerable abundance. The group decided to investigate this option further, i.e. model catchability as $\log I = \log q + c * \log R = \log q + \log R^c$ instead of $\log I = \log q + \log R$.

The model fit statistics improved from ***Setup a*** to ***Setup aSD*** and, to a lesser extent, from ***Setup aSD*** to ***Setup aSDQ*** (Table 3.3). Overall, ***Setup aSDQ*** provided the best model fit. ***Setup aSDQ*** improved the fit to the recruitment index in comparison to ***Setup aSD***, especially for the higher recruitment values (Figure 3.9., bottom panel). When including the extra SD the confidence intervals around the recruitment index observations included the fitted values. So, this addition seemed necessary. The fit to acoustic and DEPM indices was almost the same for ***Setup aSD*** and ***Setup aSDQ*** (Figure 3.9, top and middle panels). Exploration of the runs indicated that the extra SD for the autumn acoustic survey is quite high (between 0.62 and 0.83), whereas the extra SD for the spring acoustic (between 0.051 and 0.064) and DEPM surveys (between 0.077 and 0.096) are smaller. Furthermore, it was reassuring that the recruitment index did not change the past, the trends were smooth (Figure 3.10), but it appeared useful to point the direction of the recruitment signal.

Table 3.3. Log-likelihood, number of parameters, SSB, Recruitment, F, AIC and final gradient for runs with extra SD parameter and catchability power function for the recruitment index.

Label	SetupA	SetupA_SD	SetupA_SDQ
Total	228.17	143.28	140.49
Catch	2.79e-09	1.06e-09	1.18e-09
Equil_catch	0.856	0.289	0.392
Survey	44.11	-10.96	-15.08
Age_comp	152.45	130.39	131.22
Recruitment	30.7	23.6	24.0
Parm_softbounds	0	0	0
Parm_devs	5.03e-04	5.16e-04	5.18e-04
N parm	64	67	68
SSB_2020	245.87	336.78	304.77
Recr_2020	29.72	23.04	18.19
F_2019	0.1	0.06	0.06
AIC	584.34	420.566	416.97
Max Grad	2.43e-05	3.31e-05	2.26e-05

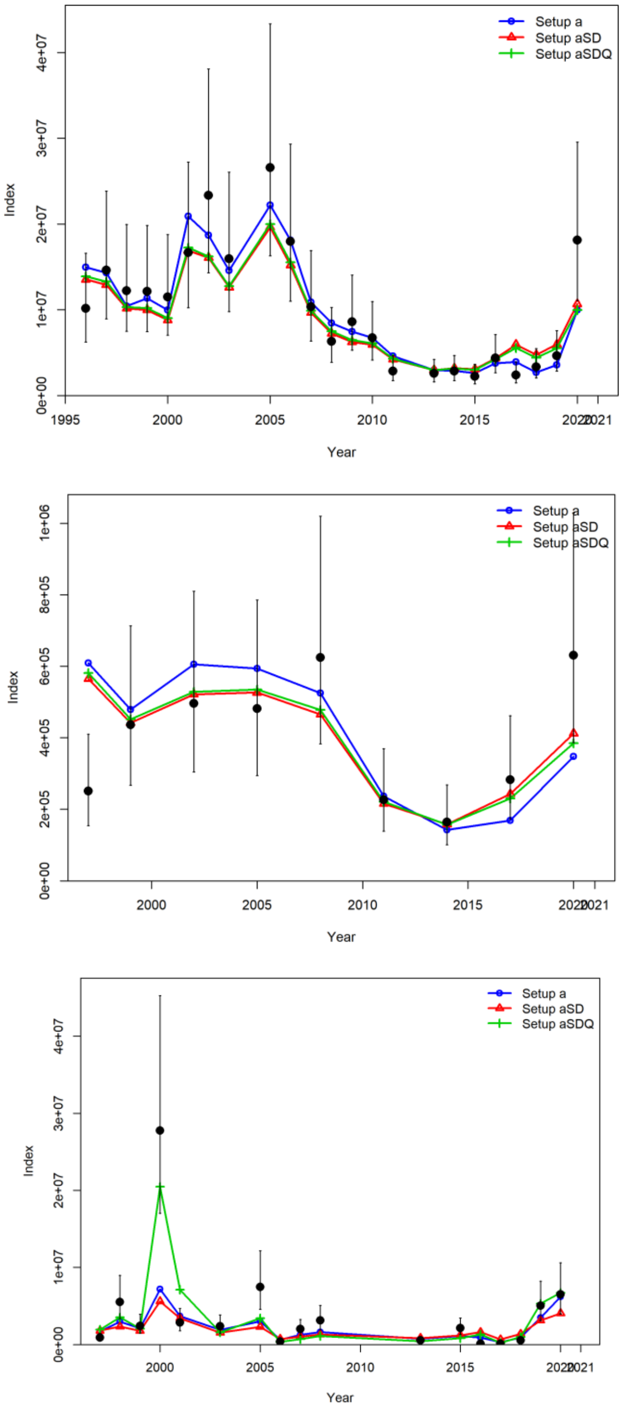


Figure 3.9. Fit for runs with extra SD parameter and catchability power function for the recruitment index to the spring acoustic survey (top), to the DEPM survey (middle) and to the autumn acoustic survey (bottom).

F

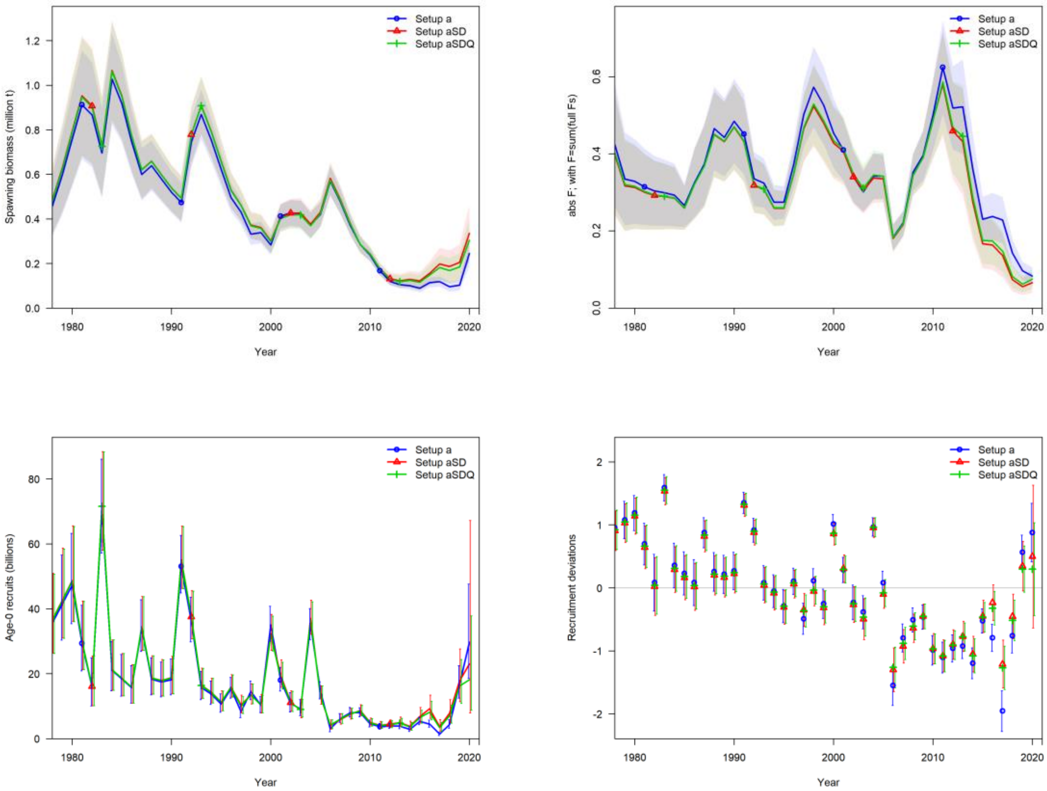


Figure 3.10. Spawning biomass (top left), fishing mortality (apical, top right), Age-0 recruits (bottom left) and recruitment deviation (bottom right) outputs from the runs to evaluate the catchability settings.

The additional parameter estimated in the power catchability model for the recruitment index (***Setup aSDQ***) was very close to 1, indicating a minor deviation from the linearity assumption (***Setup aSD***). In an additional run ***Setup aQ***, where the recruitment index was modelled with a power function but without the extra SD parameter for all the indices, the power parameter was around 1.4. In contrast to ***Setup a*** (catchability modelled with a linear function and no extra SD for the indices), ***Setup aQ*** was able to fit the high recruitment values and improve the model fit of the recruitment index. So, it seemed that to better accommodate the high values of the recruitment index either the power catchability, or the extra SDs or both were needed. A more detailed analysis of the contribution of each index to the total likelihood of the indices (Figure 3.11) showed that the likelihood of the indices attained its lowest value for ***Setup aSDQ***, followed by ***Setup aSD*** and ***Setup a***.

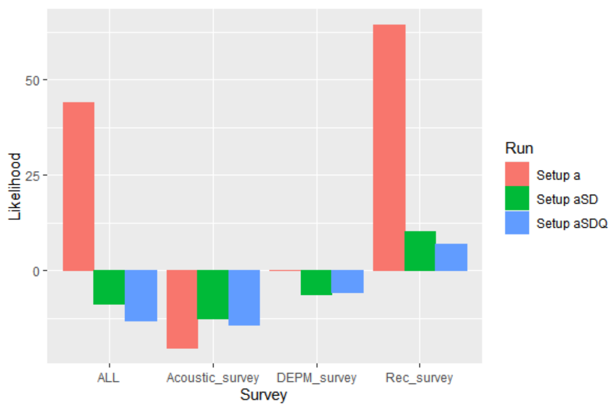


Figure 3.11. Total likehood (ALL) and likehood by fleet for the runs to evaluate the catchability setting.

The group decided to keep the extra SD parameters for all the surveys (even if not significant for the acoustic and DEPM surveys) **and that they should be estimated** (float = 0) as it's done with all the other parameters. In the future, this will allow to evaluate the evolution of these parameters (extra SD for the surveys) and use them as quality checks. In addition, **the group decided to model the catchability of the recruitment index with a power function.** Hyperstability (catchability model as a power function) is plausible for a pelagic stock with aggregative behaviour and hot spots in their spatial distribution and it accommodates better to high values of the recruitment index. Having the extra parameter for catchability updated every year as the time-series lengthens can be useful to evaluate the evolution of the power relationship, especially as larger recruitment indices inform the power relationship.

The current assessment model uses a Beverton–Holt stock-recruitment model with steepness set at 0.71, the median steepness for family *Clupeidae* from Myers et al. (1999) meta-analysis. Stock assessment trials to estimate steepness conducted during the Inter-Benchmark showed a preference for low steepness values and therefore a strong dependence of recruitment on stock biomass up to high biomass levels (Figure 3.12). The value of steepness pointed out by the profiling is considered too low for a small pelagic fish for which recruitment is primarily affected by environmental factors (Katara, 2014). In the case of sardine, the literature exploring drivers of recruitment variability ofdicate that recruitment is affected by a variety of environmental factors and only weakly related to spawning biomass (e.g. Santos et al. 2012). The group considered that we were in the same position as in the past but there was no other indication, additional information or proposed steepness value. So, given that this was not one of the issues list of the IBP, the group decided to keep the steepness assumption.

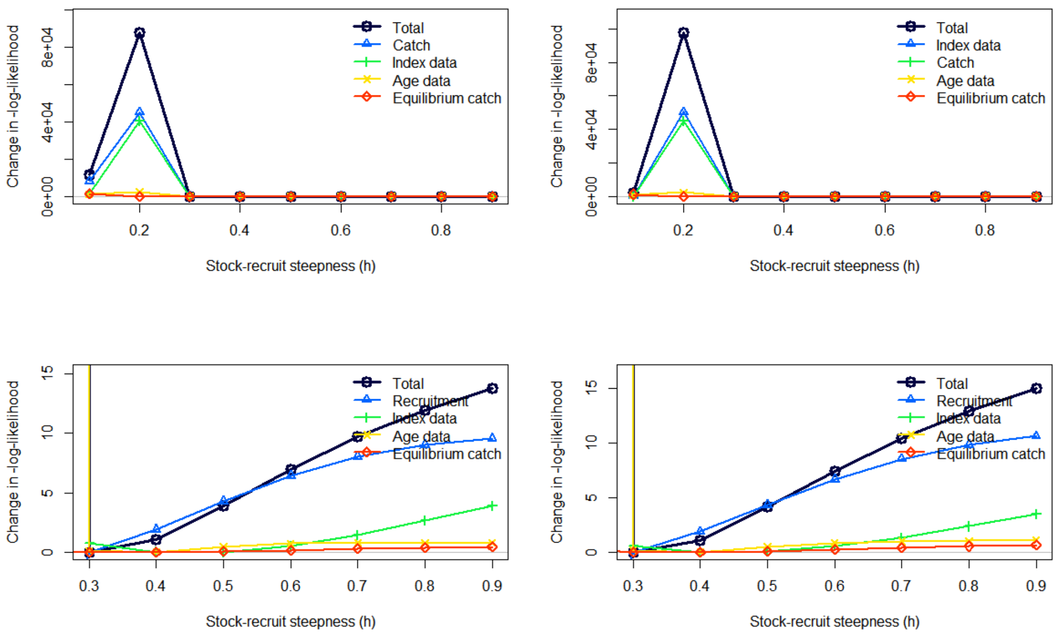


Figure 3.12. Steepness (h) likelihood profile for *Setup aSD* (left panels) and *Setup aSDQ* (right panels) runs. Bottom panels show a close up of the steepness (h) likelihood profile for values of h between 0.3 and 0.9. Please note that changes in log-likehood within this range of steepness values are very small.

The sigma R (input standard deviation of log number of recruits) and the recruitment deviations settings of the two preferred assessment models (***Setup aSD and Setup aSDQ***) were fine tuned based on the outputs/suggestions of SS. Two additional cycles (take outputs, use as inputs, run again and check outputs) were performed to get the final settings (runs ***Setup aSDTune***, ***Setup aSDTune2*** and ***Setup aSDQTune***, ***Setup aSDQTune2***). There were almost no differences in the estimates of sigma R without and with the fine tuning, but the results with the fine tuning were slightly better in terms of model fit. Sigma R parameter changed from 0.7 to 0.74 to be consistent with the residual mean standard error of the recruitments estimated by the model.

The retrospective patterns are quite good for both ***Setup aSDTune2*** and ***Setup aSDQTune2***. However, in the specific case of retro -2 of ***Setup aSDQTune2*** the SSB changes at the beginning of the time-series. After the inter-benchmark online meeting, a new run of the specific retro -2 for ***Setup aSDQTune2*** where the number of iterations in the F hybrid method was increased from 4 to 7 showed that it was in fact a convergence issue. Here we present the retrospective plots with this updated retro -2 (Figure 3.13). Changes in model led to smaller values of Mohn's rho for Recruitment, in the case of the ***Setup aSDQTune2*** run, compared to the 2020 assessment run (Table 3.4. shows the Mohn's rho values with the updated retro -2).

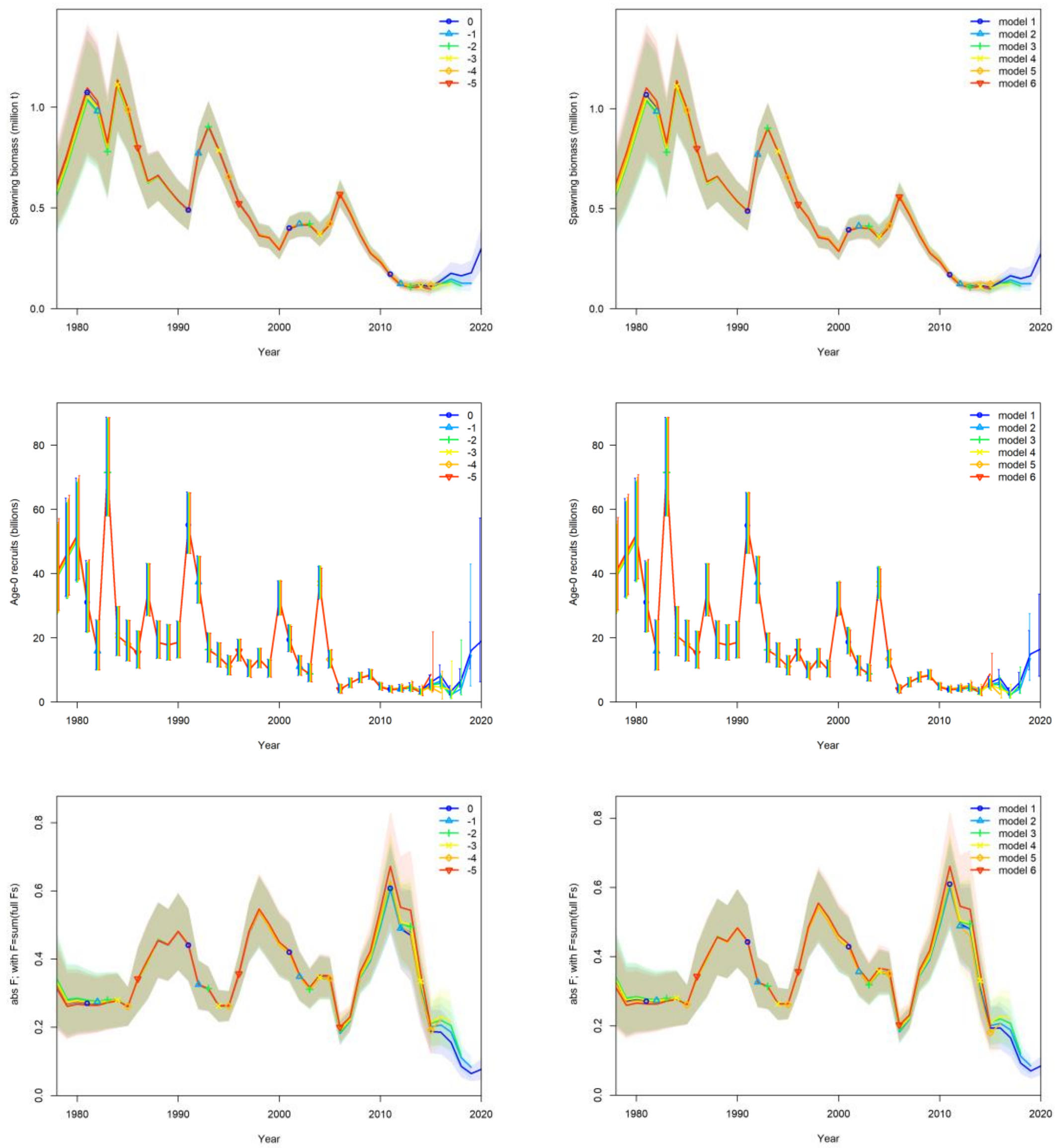


Figure 3.13. Retrospective patterns for the Setup aSDTune2 (left panels) and the Setup aSDQTune2 (right panels) runs. From top to bottom spawning output, age 0 recruits and apical F with 95% confidence intervals.

Table 3.4. Mohn's rho values for the runs 2020 assessment (current), Setup aSDTune2 and Setup aSDQTune2 .

Run	SSB	Recruits	Fvalue
Current	-0.112	-0.240	0.059
Setup aSDTune2	-0.045	0.262	-0.220
Setup aSDQTune2	-0.116	0.189	-0.155

The DEPM index didn't pass the runs test residuals for mean composition data of ss3diags (Figure 3.14.) for runs ***Setup aSDTune2*** and ***Setup aSDQTune2***. This is not considered a big concern as these tests calculate the 2-sided p-value of the Wald-Wolfowitz runs test (see Carvalho et al., 2017; Carvalho et al., 2021), which is a non-parametric statistical test that checks a randomness hypothesis for a data sequence.

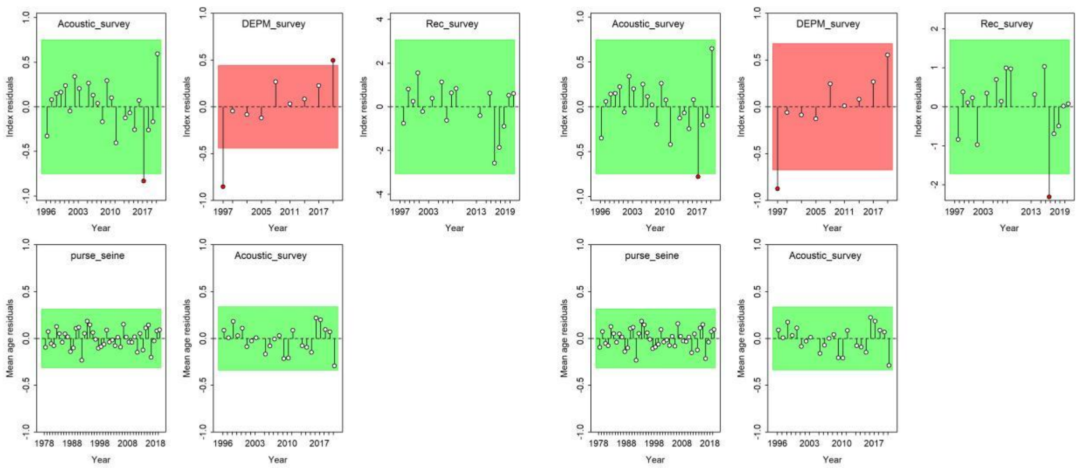


Figure 3.14. Runs Test residuals for mean composition data. From top to bottom: Index residuals (top panels) and mean age residuals (bottom panels). From left to right: *Setup aSDTune2* and *Setup aSDQTune2*.

Residual mean standard error of joint index residuals and mean age residuals are similar between runs, slightly smaller for **Setup aSDQTune2** (Figure 3.15).

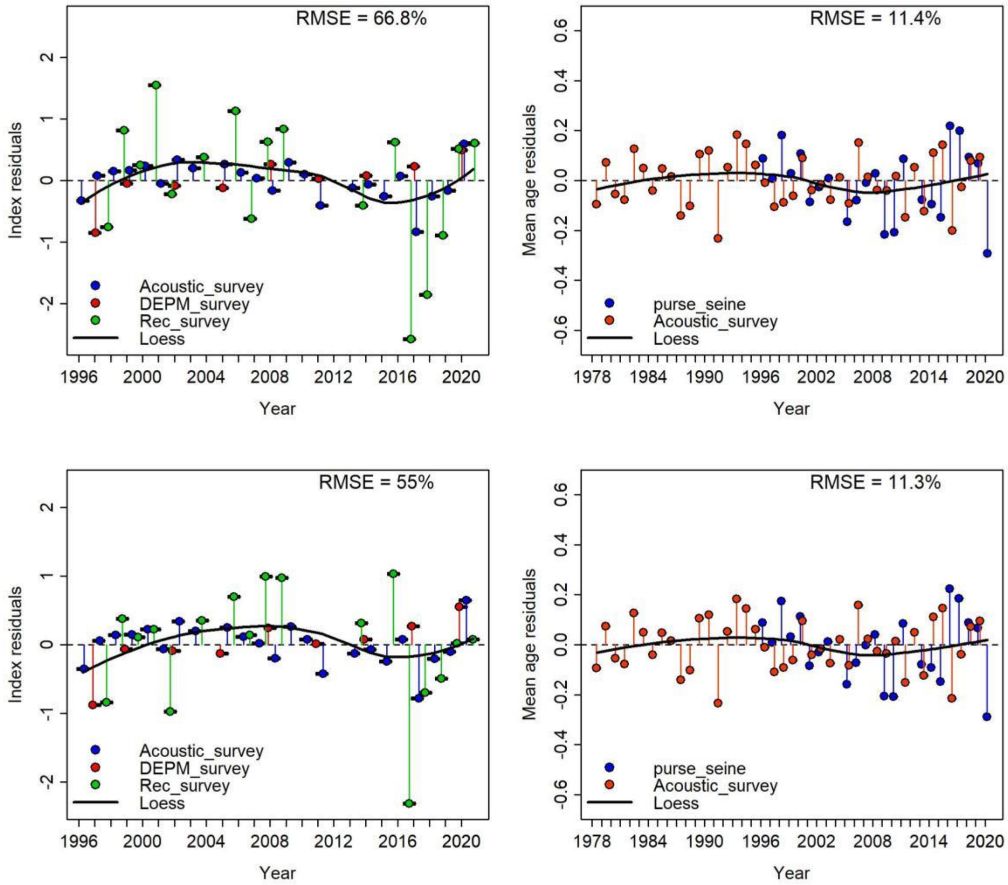


Figure 3.15. Joint residuals to check for conflicts in index residuals (left panels) and mean age residuals (right panels). From top to bottom: *Setup aSDTune2* and *Setup aSDQTune2*.

Based on all the above, **the group decided to use the model with the fine tuning** (***Setup aSDQTune2***) as the final candidate for the stock assessment model of the Iberian sardine stock.

3.4 Final stock assessment

The group preferred model shows an acceptable retrospective performance, with Mohn's rho values less than 0.2 (Table 3.4., Figure 3.13). Fits to the aggregated age composition is very good (Figure 3.16) and fits to indices (Figure 3.17) are also good. Overall, the group preferred model is not substantively different in scale, trend or variability of SSB, recruitment or fishing mortality from the initial inter-benchmark model or the updated model with the 2020 DEPM survey point estimate and the total catch of 2020 (Figure 3.18). The group preferred model simply adds a recruitment index to have a more reliable recruitment estimate in the final year of the assessment and other fine tuning settings that lead to a better fit of the high values of the recruitment index and fine tuning of the sigma R and recruitment deviations.

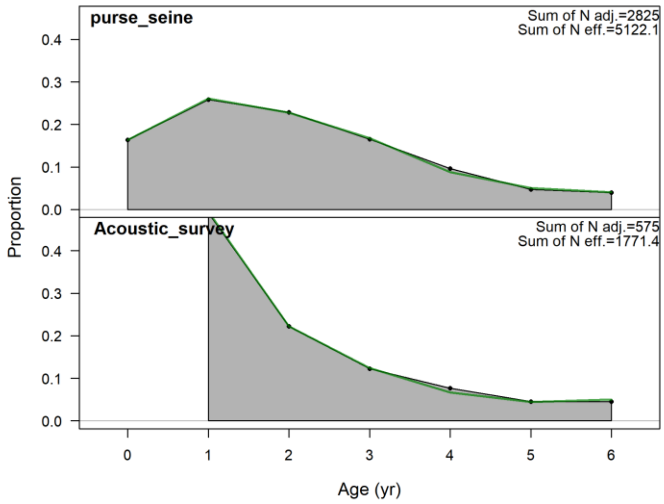


Figure 3.16 Age composition fits, aggregated across time by fleet.

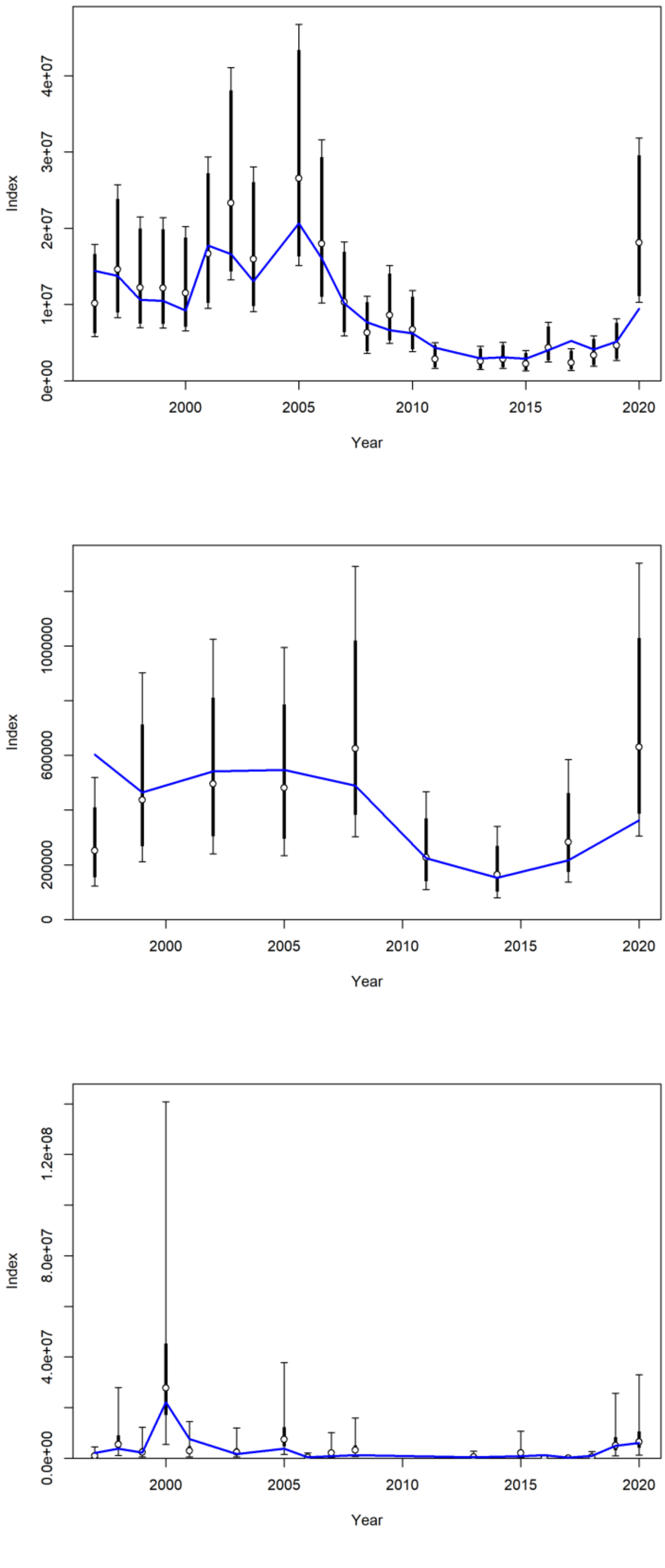


Figure 3.17. Fit to the index data for spring acoustic survey (top), to the DEPM survey (middle) and to the autumn acoustic survey (bottom). Lines indicate 95% uncertainty interval around index values based on the model assumption of lognormal error. Thicker lines indicate input uncertainty before addition of estimated additional uncertainty parameter.

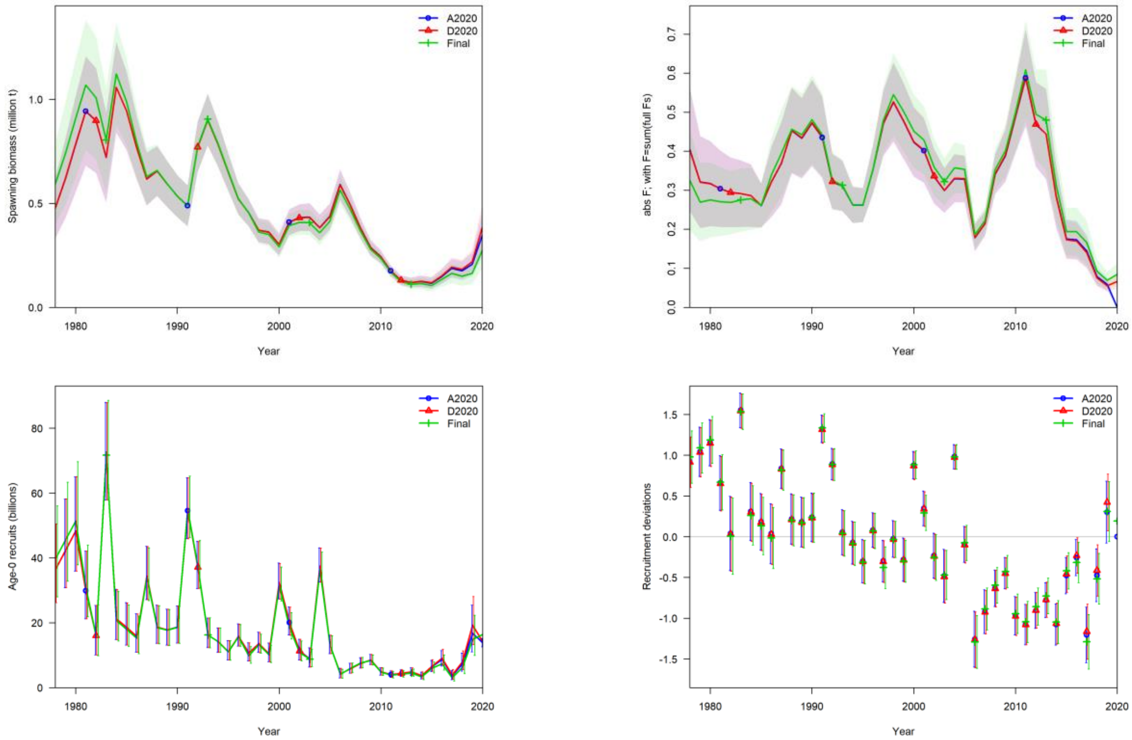


Figure 3.18. Spawning biomass (top left), fishing mortality (apical, top right), Age-0 recruits (bottom left) and recruitment deviation (bottom right) outputs comparing the current (A2020), the updated (D2020) and the group preferred model (Final). Note that the estimates of fishing mortality for the current model is different because catches for 2020 were assumed to be zero, whereas in the updated and the group preferred model catches for 2020 were equal to the quota for that year.

4 Short-term forecast

The main changes in the short-term forecast procedure from the last sardine benchmark (WKPELA2017; ICES 2017) relates to the assumption of recruitment in the interim year and the stock number-at-age 1 in the following year. Previously, the recruitment in the interim year was assumed to be the geometric mean of the recruitment estimates in the last five years of the assessment and the stock number-at-age 1 in the following year the survivors of the assumed geometric mean. Now, the stock assessment model includes a new recruitment index (autumn acoustic survey) allowing estimating recruitment in the interim year. Therefore, recruitment in the interim year is the value estimated in the assessment. The recruitment in the management year will be based on a geometric mean of the last five years (now, including the interim year estimate).

The deterministic short-term forecasts were carried out using the R-FLR (Kell et al., 2007) package FLash (version 2.5.11) with the following settings:

Input	**Settings**	**Comment**
Initial stock size	Assessment estimates for ages 1–6+ at the final year of the assessment	
Maturity	Arithmetic mean of the last six years of the assessment	Age dependent
Natural mortality (M)	As used in the assessment	Age dependent
Proportion of F and M before spawning	Zero	Corresponds to the beginning of the year when the SSB is estimated by the model
Weight-at-age in the stock	Arithmetic mean value of the last six years of the assessment	No significant trends are found in historical weight-at-age in the stock.
Weight-at-age in the catch	Arithmetic mean value of the last three years of the assessment	No significant trends are found in historical weight-at-age in the stock.
Exploitation pattern	Equal to the last year of the assessment	
Fishing mortality in the intermediate year	Assuming an F_{sq} equal to the average estimates of the last three years in the assessment	$F_{multiplier}$ ranging from 0 to 2
Recruitment in the interim year and forecast year	Recruitment in the interim year is the value estimated in the assessment; Recruitment in the forecast year is assumed to be the geometric mean of the R estimates in the last five years (which includes the R estimate in the interim year)	The new stock assessment model includes a recruitment index, allowing to estimate R in the interim year

During the inter-benchmark the group discussed the possibility of a stochastic forecast within SS. The potential advantages of this would be to have everything integrated under the same framework and therefore, a less error-prone method. However, implementing a stochastic forecast within SS requires additional work to find the most suitable settings for this particular stock which was not feasible to carry out during the meeting. **The group decided to explore this option in the future**. The group also considered that if the requirement of a stochastic STF is a general request for all the ICES stocks, then this discussion could be better addressed from a more general perspective and following some general/common guidelines.

5 Improvements in recruitment estimations

We assessed the forecast skill of the group preferred assessment model and the changes in the short-term forecast (STF) procedure by comparing the root mean-square error (RMSE) of the recruitment and spawning-stock biomass estimates of a retrospective forecast (Kiaer et al., 2021).

For each of the six retrospectives runs of the group preferred model two STF procedures were followed: one according to the reference forecast and another according to the new proposed procedure (see Section 4). The reference forecast followed the current stock assessment practices: the recruitment in the interim year (y) and in the management year was assumed to be the geometric mean of the recruitment estimates in the previous five years of the assessment (y-5 to y-1). In the new proposed procedure (minimal changes) the recruitment in the interim year (y) was assumed to be the recruitment estimate of the assessment (R_y) and in the management year was assumed to be the geometric mean of the recruitment estimates in the last five years of the assessment(y-4 to y), i.e., including the last recruitment estimate of the assessment. The RMSE of the recruitment and spawning biomass was estimated for the two groups of STF and compared.

Results show that in the interim year there is a improvement of 50% in the RMSE of recruitment estimates, while in the management year there is an improvement of 7% and 46% for recruitment and spawning biomass estimates (Table 5.1.).

Table 5.1. Root mean square error for recruitment and spawning-stock biomass for the reference and proposed forecast based on the outputs of the retrospective forecast. Improvement is the percentage of reduction of the root mean square error when comparing the proposed with the reference forecast.

		Interim Year	Management year	
		Recruitment	Spawning stock biomass	Recruitment
Root mean square error	Reference forecast	5278750	69037	7733998
	Proposed forecast	2664861	37381	7215234
Improvement (%)		50	46	7

The significant improvement in the prediction skill of recruitment and spawning biomass when incorporating the recruitment index in the stock assessment, reinforces the proposal of WGHANSA to move the stock assessment of Iberian sardine to November. The stock assessment and short-term forecast proposed by IBPIS could incorporate the results of the autumn acoustic surveys as soon as they are ready to provide management advice for the next management year (Jan-Dec). This will also avoid the recurrent re-evaluation of the advice in the spring of the management year.

6 Reference points

6.1 Current reference points

The stock's reference points were re-evaluated in April 2021 during the Workshop for the Evaluation of the Iberian sardine HCR (WKSARHCR; ICES, 2021c). Only Maximum Sustainable Yield (MSY) and Precautionary Approach (PA) fishing mortality reference points were updated (Table 6.1) using a management strategy evaluation (MSE) methodology developed to evaluate a generic HCR proposed by Portugal and Spain EU members within a management plan for 2021-2026.

Table 6.1. Previous and updated reference points. The previous biological reference points were estimated during WKSARMP (ICES, 2019) with EqSim and the current were estimated during WKSARHCR (ICES, 2021c) within a MSE framework. Both are based on the state of low productivity. Weights are in tonnes.

BRP	2006–2017	2006-2019
B_{lim}	196 334	196 334
B_{pa}	252 523	252 523
F_{lim}	0.156	0.26
MSY $B_{trigger}$	252 523	252 523
F_{pa}	0.032	0.092
F_{MSY}	0.224	0.22
Adopted F_{MSY}	0.032	0.092

The basis for the reference points, as established during WKSARHCR (ICES, 2021c), are presented in Table 6.2.

Table 6.2. Summary of Iberian sardine stock reference points and their technical basis for the low productivity regime.

Reference	Value	Technical basis
B_{lim}	196 334 t	B_{lim} = Hockey-stick change point
B_{pa}	252 523 t	$B_{pa} = B_{lim} * \exp(1.645 * \sigma)$
F_{lim}	0.26	Stochastic long term simulations within an MSE framework (50% probability B1+ < B_{lim})
MSY $B_{trigger}$	252 523 t	MSY $B_{trigger} = B_{pa}$
F_{pa}	0.092	Stochastic long-term simulations within an MSE framework with ICES MSY AR (≤ 5% probability B1+ < B_{lim}); Constraint F_{msy} to F_{pa} if $F_{pa} < F_{msy}$
F_{MSY}	0.22	Median F_{target} which maximizes yield without $B_{trigger}$

Adopted F_{MSY}*	0.092	If F_{pa} < F_{MSY} then F_{MSY} = F_{pa}

*** The F that maximizes long-term yield under the constraint that the long-term probability of B1+ < B_{lim} is ≤ 5% when applying the ICES MSY advice rule (ICES, 2021c).**

B_{lim} is the breakpoint of a segmented stock–recruitment (S–R) model and was kept the same as in WKSARMP (ICES, 2019) since the updated estimate of B_{lim} was within the 95% confidence interval of the B_{lim} estimated in WKSARMP (ICES, 2019). As B_{lim} was considered the same, for consistency, B_{pa} and MSY $B_{trigger}$ were also kept the same as in WKSARMP (ICES, 2019).

6.2 TOR b ii) - Re-examine and update (if necessary) MSY and PA reference points according to ICES guidelines.

Based on the preferred stock assessment model from IBPIS (Section 3.4), the updated estimate of B_{lim} (B_{lim} = 182 836 t) is within the 95% confidence interval of the B_{lim} estimated in WKSARMP (ICES, 2019) for the low productivity regime. Therefore, **the group decided to keep the current adopted B_{lim} = 196 334 tonnes (CI = 98 389 - 296557 t) for the low productivity regime.** As B_{lim} was considered the same, for consistency, **B_{pa} and MSY $B_{trigger}$ were also kept the same** as in WKSARMP (ICES, 2019). This was the same reasoning followed during WKSARHCR (ICES, 2021c).

The new assessment provides a new S-R data pair (we now have an estimate of recruitment in 2020) when compared to the previous assessment used for the calculation of the current reference points. However this new S-R data point does not deviate from the S-R model used for the current biological reference points B_{lim} and B_{pa} (Figure 6.1).

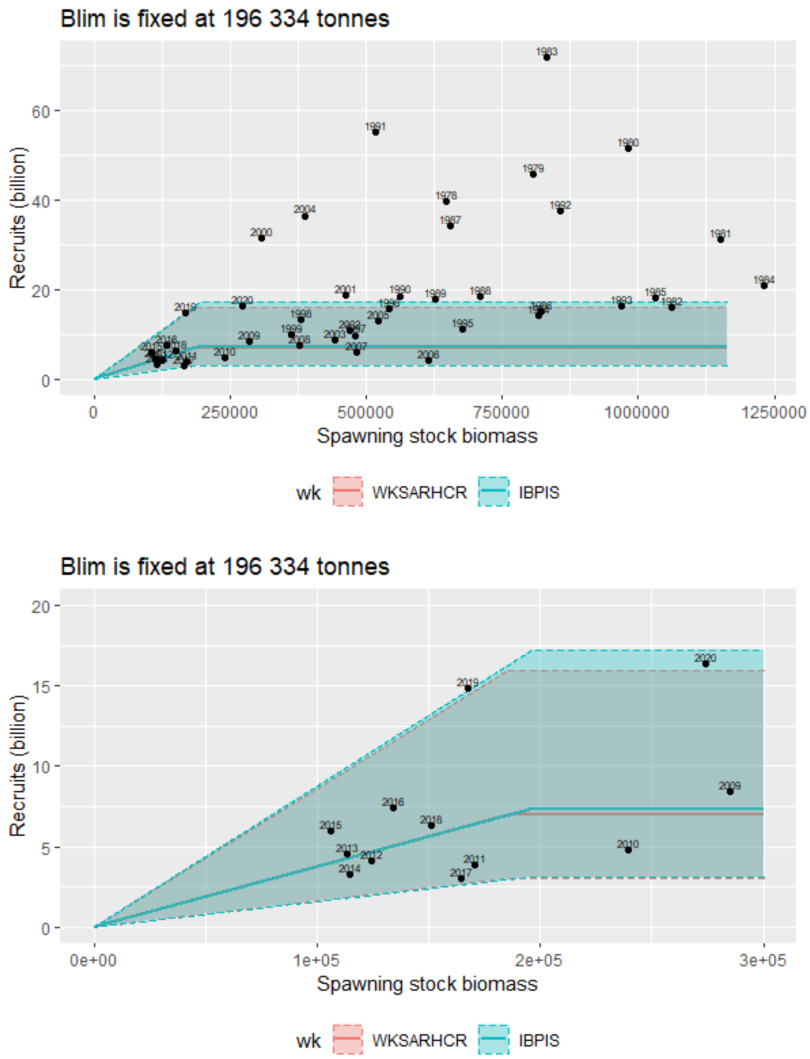
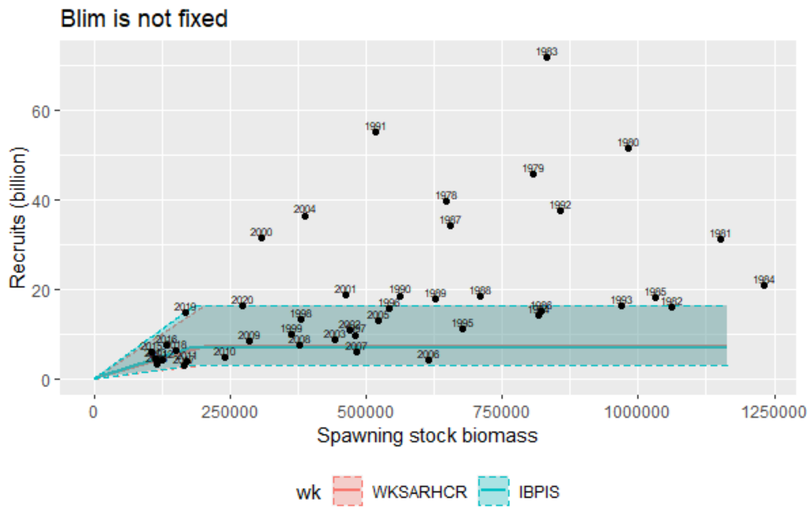


Figure 6.1. - Fitted (solid line) and 95% confidence intervals (dashed lines) for the Hockey-stick S–R relationship for the period 2006–2019 (red lines) during WKSARHCR (ICES, 2021c) and the period 2006–2020 (blue lines) during IBPIS (ICES, 2021c) while fixing the inflexion breakpoint of the model as B_{lim} = 196 334 tonnes. Bottom panel limits x-values up to 300 000 tonnes and y-axis up to 20 billion recruits. The S-R points are from the preferred group assessment model.



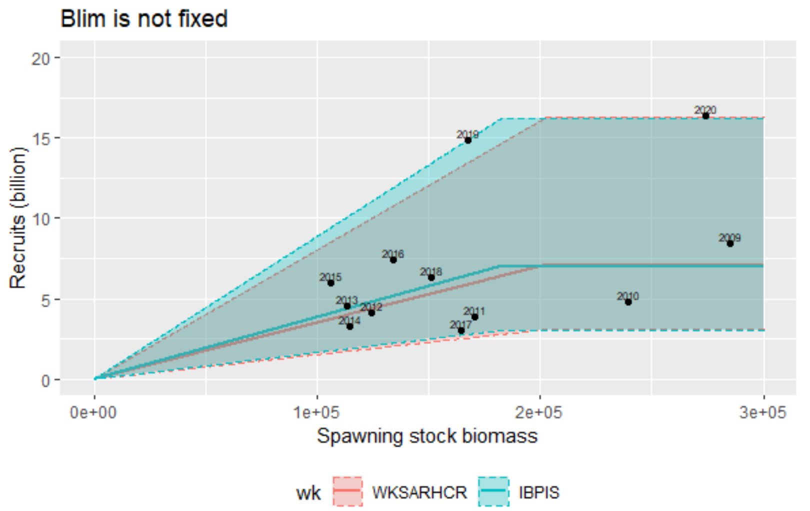


Figure 6.2. Fitted (solid line) and 95% confidence intervals (dashed lines) for the Hockey-stick S–R relationship for the period 2006–2019 (red lines) during WKSARHCR (ICES, 2021c) and the period 2006–2020 (blue lines) during IBPIS (ICES, 2021c) without fixing the inflexion breakpoint of the model as B_{lim} = 196 334 tonnes. Bottom panel limits x-values up to 300 000 tonnes and y-axis up to 20 billion recruits. The S-R points are from the preferred group assessment model.

Another factor that might affect the reference points calculation is the selectivity curve. The selectivity settings for the preferred assessment model are the same as the current assessment model. As a result, the ending year selectivity of the commercial fleet for the preferred stock assessment model is similar to the current model one (Figure 6.2). Therefore, it’s not expected that this has a big impact on the reference points.

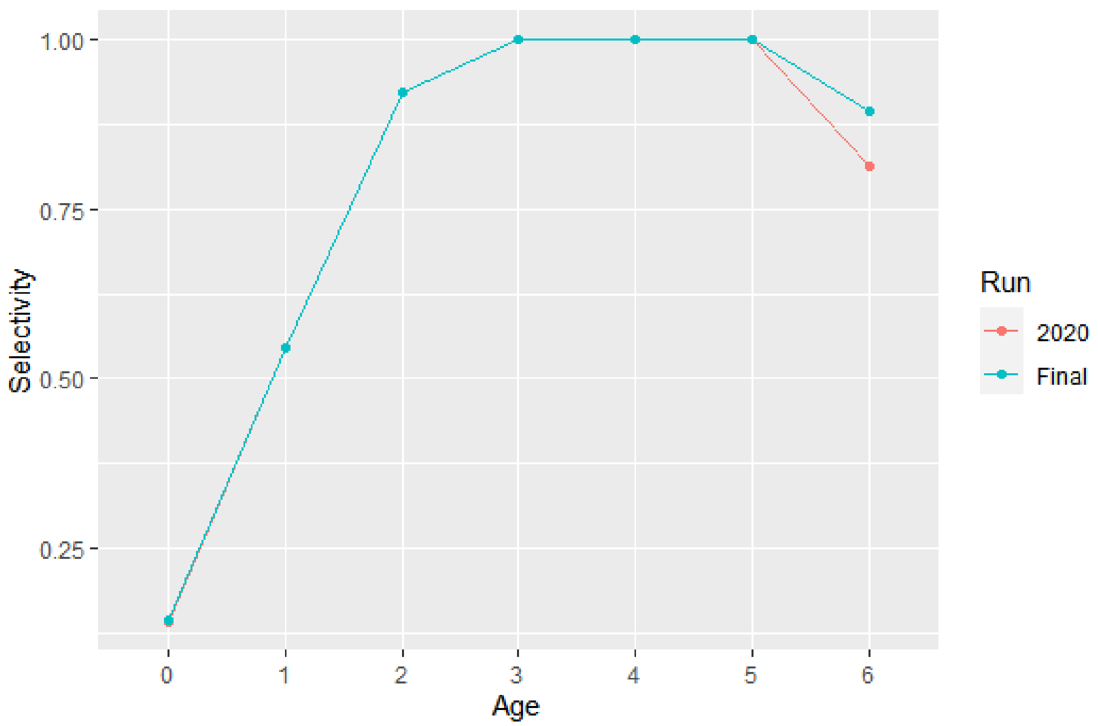


Figure 6.2. Ending year selectivity for the commercial fleet (purse-seine). In red we have the selectivity for the current assessment model and in blue the selectivity for the preferred assessment model.

In WKSARHCR2021 (ICES, 2021c) the fishing mortality reference points were calculated using the same MSE framework as for the evaluation of the management plan. This depended on the initial conditions (population in 2019 and associated uncertainty), biology (natural mortality, maturity, weight at age in stock and in catch), S-R model (segmented regression with fixed breakpoint at B_{lim}), fleet dynamics represented by Cobb-Douglas production model (catch/biomass), observation error and assessment error (note that the assessment in the MSE framework shows positively biased estimates of recruitment and spawning stock). Given the small differences between the current and preferred new assessment (see subsection 3.4), the impact on these settings is expected to be minor. However, given the high burden of the MSE framework, the group did not have time during the Inter-Benchmark Protocol to redo the calculations.

As a sensitivity analysis, it was decided to run EqSim and compare the results to the Eqsim results of WKSARHCR2021 (ICES, 2021c) to test the impact of the new preferred stock assessment (Table 6.3.). EqSim was run twice, the first run where a time-series from 2006 up to 2020 was used and another run where a time-series from 2006 up to 2019 was used. The latter run is to mimic the S-R pairs used to run the EqSim during WKSARHCR2021 (ICES, 2021c). The EqSim approach gives different results to the MSE, but it was considered that it could provide a reliable indication of the sensitivity of reference points to the stock assessment change. When using a segmented regression model with the breakpoint fixed at B_{lim}, F_{pa} changed from 0.13 to 0.1 or 0.12 and F_{MSY} changed from 0.26 to 0.23 or 0.25 depending on the time-series used in the EqSim runs (Table 6.3.). These changes were evaluated as small changes (specially when the time-series period was the same) and the group considered that the reference points are still valid.

Table 6.3. Summary of Iberian sardine stock reference points estimated in WKSARMP 2019, WKSARHCR 2021 and IBPIS 2021.

	EqSim					MSE framework
	WKSARMP 2019	WKSARHCR 2021		IBPIS 2021		WKSARHCR 2021
	w/o fixing B_{lim}		Fixing B_{lim}			
	2006-2017	2006-2019	2006-2019	2006-2019	2006-2020	**2006-2019**
B_{lim}	**196334**	202815	196334	196334	196334	**196334**
B_{pa}	**252523**	254501	252523	252523	252523	**252523**
F_{lim}	**0.156** **(27 709 t)**	0.23 (42 660 t)	0.26 (45 922 t)	0.25 (45 363 t)	0.27 (48 363 t)	**0.26**
F_{MSY}	**0.22** **(28 312 t)**	0.27 (44 592 t)	0.26 (45 356 t)	0.25 (45 040 t)	0.23 (47 962 t)	**0.23**
F_{pa}	**0.032** **(8 044 t)**	0.073 (20 097 t)	0.13 (32 848 t)	0.12 (32 115 t)	0.10 (30 140 t)	**0.092**
Adopted F_{MSY}*	**0.032**	0.073	0.13	0.12	0.10	**0.092**

For all these reasons above, **the group decided to keep the current MSY and PA reference points.** Nevertheless, the group acknowledges that it was not possible during IBPIS to do any formal verification of the impact of the new assessment on the reference points calculated using the MSE framework. The high computing facilities that were used in WKSARHCR2021 were not available during IBPIS (re-installation and maintenance) and will probably only be available in December 2021. The available time frame for additional analyses before the Advice is due is very short and any additional testing won't be possible until beginning of 2022.

The reference points for this stock were updated in spring 2021. The group considered that it is not recommendable to update the reference points very often. Moreover, ICES is reconsidering redefining the procedure to estimate reference points (WKREF1 and WKREF2). At this point it might be useful to consider the conclusions from WKREF1 and WKREF2 on reference point calculation before updating the reference points for this stock.

7 Evaluation of the management plan

A Harvest Control Rule (HCR) proposed by Portugal and Spain for the management of the Iberian sardine was evaluated by ICES in April 2021 (WKSARHCR; ICES, 2021c). The HCR specified different levels of fishing mortality associated with reference biomass levels (Figure 7.1), and three candidate values of maximum allowed catches. The HCR was evaluated under an MSE framework (ICES, 2021c) and ICES found the HCR to be precautionary with maximum allowed catches between 30 000 and 50 000 tonnes (ICES, 2021d).

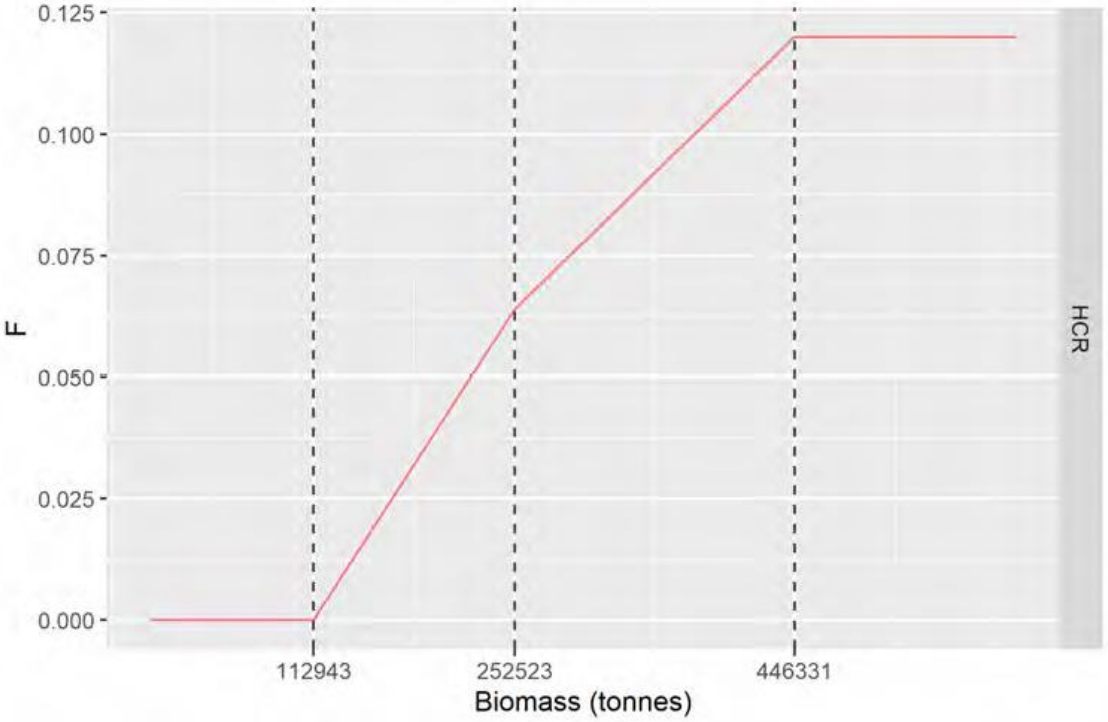


Figure 7.1. HCR: The biomass reference levels (B1+) are B_{loss}(2018) = 112 943 t, MSY $B_{trigger\_low}$= B_{pa_low}= 252 523 t and MSY $B_{trigger\_med}$ = B_{pa_med} = 446 331 t.

The group discussed the rationale for re-evaluating the HCR. Since the new stock assessment model, adopted by the group during the Inter-Benchmark meeting, showed minor deviations of the stock key-parameter estimates (F, SSB, R) in relation to the previous assessment (Section 3) and that both fishing and biomass reference points are still valid (Section 6), the group considered that such small differences would likely have a very minor impact on the evaluation of the HCR.

It is also noted that in case such evaluation is deemed necessary it will not be possible to carry out the analysis before early 2022 given the high computational burden required to redo the MSE analysis for the several HCR options.

8 Future work

During the Inter-Benchmark the group identified the following issues for future research:

- Continue research to improve the estimates of steepness of the Stock-Recruitment model and natural mortality.
- Continue monitoring other potential areas of recruitment (specially the northern area of 9aCS) for the autumn acoustic surveys.
- Monitor the power function for the catchability of the recruitment index, especially when large recruitments occur.
- Study further causes of changes in selectivity over time.
- Investigate methods to detect changes in the productivity regime and to identify the current productivity regime.
- Evaluate the possibility to conduct the short-term forecast and calculation of reference points within SS.
- Evaluate the possibility to conduct stochastic short-term forecast to evaluate the probability of being below B_{lim}. This might require additional guidelines from ICES about the sources of uncertainty to be considered.
- Study the magnitude of the changes in the stock assessment that would trigger significant changes in the reference points.

9 Conclusions

The main conclusions from the Inter-Benchmark Protocol for Iberian Sardine (IBPIS) were the following:

- The Age 0 estimates of the autumn acoustic surveys in 9aCN from 1997 onwards were considered a reliable index of recruitment to be included in the stock assessment. Given that this index only covers a fraction of the main recruitment areas for this stock, the group recommended continuing monitoring other potential areas (particularly 9aN and 9aCS) in order to detect spatial changes in the recruitment area that could trigger future changes of the IBERAS autumn survey design. In addition, the monitoring of 9aS with the ECOCADIZ-RECLUTAS survey should be continued.
- The settings of the Stock Synthesis (SS) stock assessment model were modified for the inclusion of the recruitment index. The catchability of the recruitment index was modelled with a power function and an extra additive SD parameter was included for all the abundance indices (Acoustic, DEPM and recruitment). In addition, the input standard deviation of log number of recruits (sigmaR) and the recruitment deviations were fine-tuned based on the suggestions of the SS output.
- The stock assessment results in terms of SSB, recruitment and F were very similar to those from the previous model.
- The inclusion of the recruitment index in the stock assessment allowed the estimation of recruitment in the interim year. The short-term forecast assumptions were updated. Previously, the recruitments in the interim year and in the management year were assumed to be the geometric mean of the recruitment estimates in the last five years of the assessment. Based on the new stock assessment settings, recruitment in the interim year was changed to be the value estimated in the assessment and the recruitment in the management year was the geometric mean of the last five years (now, including the interim year estimate).
- The prediction skill of the new assessment and short-term forecast procedures improved notably. The root-mean square error of recruitment in the interim year and of SSB in the management year improved by 50% and 46% respectively.
- Since the new stock assessment model showed minor deviations of the stock key-parameter estimates (F, SSB, R) in relation to the previous assessment, the group considered that such small differences would likely have a very minor impact on the reference points and on the evaluation of the HCR.
- The updated estimate of B_{lim} was within the 95% confidence interval of the previous B_{lim}. Therefore, the group decided to keep B_{lim}, and for consistency, B_{pa} and MSY $B_{trigger}$ were also kept the same. A sensitivity analysis using Eqsim indicated that the impact of the new assessment on F_{pa} and F_{MSY} was small. However, the group acknowledged that it was not possible during IBPIS to do any formal verification of the impact of the new assessment on the F reference points and on the evaluation of the HCR using the MSE framework. In case such check is deemed necessary, the group emphasized that it will not be possible to carry out the analysis before early 2022.

10 External reviewers report

Aaron M. Berger (National Ocean and Atmospheric Administration, United States Department of Commerce) and Kelli F. Johnson (National Ocean and Atmospheric Administration, United States Department of Commerce) acted as the external experts for the ICES Inter-Benchmark Protocol for Iberian sardine in 2021, which took place over October 25-29, 2021. The reviewers would like to thank the organizers and the participants of the Inter-Benchmark Protocol for the well-run meeting and all of the hard work that was performed on such a short timeline to answer questions that came up throughout the review. The ability to make informed decisions was facilitated by the high level of organization, clear presentation of results, and the intimate understanding of the biology of the stock at hand by those present.

The primary task of the review was to decide which survey or combination of surveys should be used as an index of recruitment for Iberian sardine. All other tasks depended on this decision. Much of the discussion focused on the timing of the survey, the length of the time-series, and the spatial coverage of the survey relative to recruitment and spawning events. The timing of the survey is relevant in two major ways, with respect to recruitment events such that the survey adequately captures trends in recruitment and with respect to the stock assessment deadlines such that information from the survey can be post-processed and included in the stock assessment while still meeting management deadlines. Additional discussions were had regarding the survey protocols to ensure that survey data could be combined and not violate assumptions regarding survey selectivity and catchability.

In short, the autumn acoustic surveys in 9aCN were chosen as the best representation of recruitment because surveys in this area covered the bulk of the core areas of juvenile habitat, i.e., Northern Portuguese shelf, Tagus estuary, and eastern Gulf of Cadiz. Coverage of the eastern Gulf of Cadiz was not seen as having high importance because the trends in recruitment were not correlated with the age-one fish in the spring acoustic survey. It was suggested that the estimates of correlations between time-series may be spurious if autocorrelation is present in the data (Pyper and Peterman, 1997; Probst et al., 2012). Further analyses were performed during the review to assess time-series relationships after removing autocorrelation (‘prewhitening’). This procedure led to similar results because autocorrelation was low. There was a large interest in including information collected in 9aCS but the spatial coverage and timing of the surveys in this area have not been consistent and the data, currently, cannot be partitioned between north and south.

The time-series of data from 9aCN were truncated to start in 1997 although the survey extends back in time to 1984 because of differences in survey protocols that could lead to differences in selectivity or catchability. Largely, the change in the survey design from a zig-zag pattern to parallel transects precludes any immediate and straightforward combination of these two time-series into a single cohesive time-series.

Recommendations for future research regarding survey data used as a recruitment index include:

- Determine the oceanographic conditions and/or movement patterns that lead to recruitment hot spots in the eastern Gulf of Cadiz that are not seen in the age-one fish in the subsequent year.
- Conduct analyses to partition the data collected in 9aCS such that information can be summarized for the northern portion of 9aCS. These partitions will allow for more detailed spatio-temporal analyses, including irregular survey coverage and variable recruitment distributions.
- Conduct a spatio-temporal analysis of data inclusive of 9aCN and 9aCS that accounts for differences in timing and spatial coverage of these areas given that the current analysis cannot account for changes in the coverage of 9aCS. This will potentially allow the information content in 9aCS to be utilized in the index in a statistically rigorous way.
- Document procedures used to combine survey data collected using different acoustic frequency levels, as this was also highlighted as a difference between surveys, to show assumptions, advantages, and limitations of combining the available data.

During the review, several additional modelling explorations were evaluated to examine the inclusion of the autumn acoustic age-zero survey index in the stock assessment. The addition of a power function for age-zero survey catchability was examined to improve the lack of fit to extreme recruitment events when a proportional relationship was assumed between the index and the recruitment of age-zero fish. Although the degree of non-linearity (hyperstability) in catchability was estimated to be small, it did improve model diagnostics and was, in principle, appropriate to a dynamic aggregating stock that is surveyed with incomplete spatial coverage. Hyperstability in catchability should be evaluated carefully when developing subsequent stock assessments with further data and supporting analyses. Along with the inclusion of the age-zero index, parameters to accommodate additional variance in all surveys, above and beyond the input variances (CVs), were investigated to account for unmeasured sources of error in surveys and to evaluate how constricting both measured and assumed input variances were relative to other data sources in the integrated stock assessment model. The extra standard deviation term estimated for the autumn age-zero acoustic survey led to better fits to the spring acoustic survey and the DEPM survey because the assumed CV for the age-zero acoustic survey index was too constrictive. The assessment model was further ‘tuned’ after these changes. Specifically, the use of the bias adjustment ramp to accommodate the transfer between a normal and a lognormal distribution was used for the estimation of recruitment deviations and that the input value for the variability of recruitment deviations, σ_r, matched the variability of the estimates of the deviations themselves.

Other minor model discussions and analytical explorations undertaken during the review week did not result in changes to the stock assessment model but were noted as future research endeavors. Of particular note for determining stock status relative to reference points was the examination of steepness in the assessment model. Likelihood profiles were suggested and conducted for the steepness of the Beverton–Holt stock-recruitment relationship, which was fixed at 0.71 in the assessment as suggested by a meta analysis by Meyers et al. (1999). The likelihood profile provided a comparison of alternative values with respect to what data are available to inform the steepness parameter. The profiles suggested that there was very little information in the data to estimate steepness, similar to the Pacific sardine assessment (Kuriyama et al., 2020); although Kuriyama et al. (2020) chose to fix steepness at a much lower value ($h = 0.3$) than used in this assessment. As in the previous benchmark of Iberian sardine, the data suggested a much lower value of steepness than the meta analysis. Future iterations of this assessment should continue to evaluate how best to parameterize steepness and, if input as a fixed parameter, provide sensitivity analyses of alternative plausible fixed values to explore management outcomes such as stock status and expected management performance relative to reference points.

Recommendations for future work regarding the stock assessment for Iberian sardine include:

- Investigate the use of the Dirichlet-multinomial distribution / parameterization for composition data such that weighting of input sample sizes can be done internally.
- Provide rationale for the use of constant input sample sizes in the assessment model. If these are not based on the actual number of samples taken, then work should be initiated to determine year-specific sample sizes that reflect sampling protocols.
- Devote time to understanding the runs-test diagnostics such that a rationale for the continued lack of fit to the DEPM survey data can be summarized.
- Investigate a model structure that simultaneously estimates natural mortality and steepness to determine if the low estimates of steepness are a by-product of an overly-strict model structure because these parameters are correlated and directly related to many reference points.
- Update the Stock Synthesis executable to the newest version available. This newer version has improvements that pertain to the calculation of initial fishing mortality that could improve the estimation capabilities of the retrospective runs.
- Implement the Stock Synthesis model using Markov Chain Monte Carlo estimation techniques to generate stochastic forecasts to preserve the benefits of using an integrated modelling approach into the forecast period (i.e., not conduct forecasts outside the assessment model) while also improving the propagation of uncertainty into the forecast period (i.e., attempt Markov Chain Monte Carlo).

The changes included in the stock assessment as a result of this Inter-Benchmark Protocol, i.e., the inclusion of the autumn acoustic survey, the tuning of recruitment deviations, the estimation of additional error for the surveys, were not thought to have considerable impact on the calculation of reference points. In particular, there was no change in the approach to model selectivity (potential to have large impact fishing mortality based reference points) or steepness (potential to have large impact biomass based reference points) with the updated stock assessment model. Additionally, there was not an immediate concern to re-run the Management Strategy Evaluation at this time because there was only a minor difference in reference points after model changes. This was advantageous as the additional analytical time needed to run the MSE would not be easily amenable to existing management cycle deadlines. Nonetheless, future work should confirm these notions by conducting the following:

- Rerun the MSE to see if the HCR is still precautionary
- Determine if existing reference points remain applicable to management objectives.

If policies are not already in place, ICES should consider appropriate timelines to periodically reconsider reference points (especially limit reference points) and harvest rules. Re-evaluation cycles should ideally be based on species life history characteristics, significant model changes, or the considerable addition of new information.

The use of the Stock Synthesis stock assessment model that was developed for Iberian sardine with the addition of fitting to the autumn acoustic index provides improved forecasts of recruitment relative to the Stock Synthesis model used for management prior to this review. More can be done, as always, but having at least one additional year of informed recruitment deviations was deemed more useful than two years based on the geometric mean of the previous five years. Indices of recruitment are inherently variable and the Ibearian sardine age-zero survey is no exception. Hence, it will be critical to accurately portray this uncertainty and allow it to propagate through the forecast period within the integrated assessment model itself to adequately capture the context within which to characterize future recruitment and interpret the range of plausible future biomass levels. As such, the reviewers recommend that future work consider integrating the existing 'outside the model' forecast procedure directly into the integrated assessment. This new approach should then be tested within the Management Strategy Evaluation framework to examine the performance and robustness of existing, or proposed new, management procedures.

11 References

Caeiro, F. and Mateus, A. (2014). randtests: Testing randomness in R. R package version 1.0. https://CRAN.R-project.org/package=randtests

Carvalho, F., Punt, A. E., Chang, Y.-J., Maunder, M. N., and Piner, K. R. 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? Fisheries Research, 192: 28-40.

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., et al. 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research, 240: 105959.

Checkley, D., Alheit, J., Oozeki, Y., & Roy, C. (Eds.). 2009. Climate Change and Small Pelagic Fish. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511596681.

Doray, M., Boyra, G., and van der Kooij, J. (Eds.). 2021. ICES Survey Protocols – Manual for acoustic surveys coordinated under ICES Working Group on Acoustic and Egg Surveys for Small Pelagic Fish (WGACEGG). 1st Edition. ICES Techniques in Marine Environmental Sciences Vol. 64. 100 pp. https://doi. org/10.17895/ices.pub.7462.

Gislason H., Daan N., Rice J.C., Pope J.G. 2010. Size, growth, temperature and the natural mortality of marine fish. Fish and Fisheries. https://doi.org/10.1111/j.1467-2979.2009.00350.x.

ICES. 2017. Report of the Benchmark Workshop on Pelagic Stocks, 6–10 February 2017, Lisbon, Portugal. ICES CM 2017/ACOM:35. 278 pp.

ICES. 2019. Workshop on the Iberian Sardine Management and Recovery Plan (WKSARMP). ICES Scientific Reports. 1:18. 125 pp. http://doi.org/ 10.17895/ices.pub.5251.

ICES.2020. Working Group on Southern Horse Mackerel, Anchovy and Sardine (WGHANSA). ICES Scientific Reports. 2:41. 655 pp. http://doi.org/10.17895/ices.pub.5977.

ICES. 2021a. Advice on fishing opportunities, catch, and effort Bay of Biscay and the Iberian Coast ecoregion. ICES Advice 2021 – pil.27.8c9a – https://doi.org/10.17895/ices.advice.8203.

ICES. 2021b. Working Group on Southern Horse Mackerel, Anchovy and Sardine (WGHANSA). ICES Scientific Reports. 3:55. 455 pp. https://doi.org/10.17895/ices.pub.8138.

ICES. 2021c. The Workshop for the evaluation of the Iberian sardine HCR (WKSARHCR). ICES Scientific Reports. 3:49. 115 pp. https://doi.org/10.17895/ices.pub.7926

ICES. 2021d. Request from Portugal and Spain to evaluate a new Harvest Control Rule for the management of the Iberian sardine stock (divisions 8.c and 9.a). In Report of the ICES Advisory Committee, 2021. ICES Advice 2021, sr.2021.05. https://doi.org/10.17895/ices.advice.8163

Katara, I., 2014. Recruitment variability. Biol. Ecol. Sardines Anchovies, 242.

Kell, L. T. (2017). FLash: Forecasting and Projection of Fish Stocks. R package version 2.5.11.

Kiaer, C., Neuenfeldt, S., and Payne, M. R. 2021. A framework for assessing the skill and value of operational recruitment forecasts. ICES Journal of Marine Science. In press. https://doi.org/10.1093/icesjms/fsab202.

Magnusson, M., Millar, C. and Anne Cooper. icesAdvice: FunctionsRelated to ICES Advice, 2018. R package version 2.0-0.

Methot, R. and Taylor, I. G. (2011). Adjusting for bias due to variability ofestimated recruitments in fishery assessment models.Can. J. Fish. Aquat.Sci., 68:1744–1760.

Methot, R. and Wetzel, C. R. (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management.Fisheries Research, 142:86–99.

Mohn, R. (1999). The retrospective problem in sequential population analysis; An investigation using cod fishery and simulated data. ICES Journal of Marine Science, 56: 473–488.

Myers, R. A., Bowen, K. G., and Barrowman, N. J. (1999). The maximum reproductive rate of fish at low population sizes. Canadian Journal of Fisheries and Aquatic Sciences, 56: 2404–2419.

Probst, W. N., Stelzenmüller, V., and Fock, H. O. 2012. Using cross-correlations to assess the relationship between time-lagged pressure and state indicators: an exemplary analysis of North Sea fish population indicators. ICES Journal of Marine Science, 69: 670-681.

Pyper, B. J., and Peterman, R. M. 1997. Comparison of methods to account for autocorrelation in correlation analyses of fish data. Canadian Journal of Fisheries and Aquatic Sciences, 55: 2127-2140.

Rodríguez-Climent, S., Angélico, M. M., Marques, V., Oliveira, P., Wise, L., and Silva, A. 2017. Essential habitat for sardine juveniles in Iberian waters. Scientia Marina, 81: 351.

Santos, M.B., González-Quirós, R., Riveiro, I., Cabanas, J.M., Porteiro, C., Pierce, G.J., 2012. Cycles, trends, and residual variation in the Iberian sardine (*Sardina pilchardus*) recruitment series and their relationship with the environment. ICES J. Mar. Sci. 69 (5), 739–750, http://dx.doi.org/10.1093/icesjms/fsr186.

Silva, A., Skagen, D. W., Uriarte, A., Massé, J., Santos, M. B., Marques, V., Carrera, P., et al. 2009. Geographic variability of sardine dynamics in the Iberian Biscay region. ICES Journal of Marine Science, 66: 495-508.

Silva, A., Garrido, S., Ibaibarriaga, L., Pawlowski, L., Riveiro, I., Marques, V., Ramos, F., et al. 2019. Adult-mediated connectivity and spatial population structure of sardine in the Bay of Biscay and Iberian coast. Deep Sea Research Part II: Topical Studies in Oceanography, 159: 62-74.

Taylor, I. G., Stewart, I. J., Hicks, A. C., Garrison, T. M., Punt, A. E., John R. Wallace, J. R., Wetzel, C. R. et al. (2020). r4ss: R Code for Stock Synthesis. R package version 1.40.1. https://github.com/r4ss/r4ss

Winker, H., Carvalho, F., Cardinale, M., and Kell, L. (2020). ss3diags: What the Package Does (One Line, Title Case). R package version 1.0.4.

Annex 1: List of participants

Name	Institute	Country (of institute)	E-mail
Pedro Amorim	IPMA	Portugal	pedro.amorim@ipma.pt
Maria Manuel Angelico	Portuguese Institute for the Sea and the Atmosphere	Portugal	mmangelico@ipma.pt
Manuela Azevedo	IPMA	Portugal	mazevedo@ipma.pt
Aaron Berger (External Reviewer)	NOAA	United States of America	aaron.berger@noaa.gov
Pablo Carrera	The Spanish Institute of Oceanography	Spain	pablo.carrera@ieo.es
Susana Garrido	Portuguese Institute for the Sea and the Atmosphere	Portugal	susana.garrido@ipma.pt
Leire Ibaibarriaga (chair)	AZTI	Spain	libaibarriaga@azti.es
Kelli F. Johnson (External Reviewer)	NOAA	United States of America	Kelli.johnson@noaa.gov
David Miller	ICES	Denmark	david.miller@ices.dk
Cristina Nunes	Portuguese Institute for the Sea and the Atmosphere	Portugal	cnunes@ipma.pt
Isabel Riveiro	The Spanish Institute of Oceanography	Spain	isabel.riveiro@ieo.es
Andrés Uriarte	AZTI	Spain	auriarte@azti.es
Laura Wise	IPMA	Portugal	lwise@ipma.pt

Annex 2: Resolutions

IBPIS – Inter–Benchmark Protocol for Iberian sardine

2021/2/FRSG74 An **Inter-Benchmark Protocol for Iberian sardine (IBPIS)**, chaired by Leire Ibaibarriaga, Spain, and reviewed by two invited external experts, Kelli Faye Johnson (USA) and Aaron Berger (USA) will be established and work by correspondence as well as meet online from 25–29 October 2021 to:

d) Investigate the inclusion of data from autumn acoustic surveys (recruitment index) in the SS3 assessment model currently used for the stock. Where necessary revisit any other aspect of the model configuration that may be impacted by the addition of these data.

e) If new standard analysis methods are proposed:

  i. Update the short-term forecast assumption on recruitment.

  ii. Re-examine and update (if necessary) MSY and PA reference points according to ICES guidelines.

  iii. Re-run MSE analyses to see if the HCR evaluated in 2021 is still precautionary given the new assessment.

f) Agree to, and document the preferred method for evaluating the status of the stock and short-term forecast and update the stock annex as appropriate.

Stocks	Stock coordinator	Stock assessor
Sardine (*Sardina pilchardus*) in divisions 8.c and 9.a (Cantabrian Sea and Atlantic Iberian waters)	Isabel Riveiro	Laura Wise

The Inter-Benchmark Workshop will report by 15 November 2021 for the attention of ACOM and the Benchmark Oversight Group.

Annex 3: List of working documents

Wise, L., Riveiro, I., Silva, A., Carrera, P., Moreno, A., Ibaibarriaga, A., Garrido, S., Azevedo, M., Uriarte, A., and Miller, D. 2021. Preliminary analysis to include data from autumn acoustic surveys in the assessment of the Iberian sardine stock. WD presented to IBPIS 2021.

Annex 4: Stock annex

The text below provides a link to the stock annex for sardine in divisions 8.c and 9.a (Cantabrian Sea and Atlantic Iberian waters). 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.

Stock ID	Stock name	Last up-dated	Link
pil.27.8c9a	Sardine (*Sardina pilchardus*) in divisions 8.c and 9.a (Cantabrian Sea and Atlantic Iberian waters)	October 2021	Sardine 8c and 9a