INTER-BENCHMARK PROCESS ON SARDINE (*SARDINA PILCHARDUS*) IN THE BAY OF BISCAY (IBPSARDINE)

VOLUME 1 | ISSUE 80

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 I © 2019 International Council for the Exploration of the Sea

ICES Scientific Reports

Volume 1 | Issue 80

INTER-BENCHMARK PROCESS ON SARDINE (*SARDINA PILCHARDUS*) IN THE BAY OF BISCAY (IBPSARDINE)

Recommended format for purpose of citation:

ICES. 2019. Inter-benchmark process on sardine (*Sardina pilchardus*) in the Bay of Biscay (IBPSardine). ICES Scientific Reports. 1:80. 50 pp. http://doi.org/10.17895/ices.pub.5552

Editors

John Walter

Authors

Leire Citores • Erwan Duhamel • Leire Ibaibarriaga • Gaël Lavialle • Lionel Pawlowski • Alexandra Silva • Andrés Uriarte • Matthieu Veron • John Walter

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

Contents

i Executive summary

This report documents a Bay of Biscay sardine inter-benchmark tasked with evaluating the stock assessment focusing on retrospective bias, data revisions and updating reference points. The working group (WG) used standard model diagnostics to evaluate a series of interventions designed to evaluate the models and to determine causes of and corrections for the retrospective bias.

The retrospective bias could be corrected by several straightforward interventions. First, fixing selectivity at asymptotic improved model fit and reduced bias. Second, invoking a very weak stock–recruitment relationship (steepness=0.99) and commensurate bias correction ramping on recruitment deviations coupled with not estimating terminal year recruitment, further reduced the bias. Such a treatment of terminal year recruitment and penalizing poorly informed recruitment deviations is common assessment practice. Additional concerns were raised by the estimated catchability coefficients above one for the PELGAS and BIOMAN surveys. There are a number of reasons why these surveys could estimate higher abundance than the assessment model. These include mismatch of timing given the rapid population dynamics, overestimation of acoustic biomass, mismatch of assumed selectivity of the survey as well as many other common issues that support the standard practice of treating most surveys as relative rather than absolute. Once the decision to use these indices as relative inputs, the absolute value of catchability is meaningless as the index could simply be scaled to a mean of one with the same impact in the model.

Given the substantial reduction in retrospective bias achieved through straightforward model interventions and the solid diagnostic performance of the WG-preferred model, the WG recommends the assessment be upgraded from category 2 to category 1. Nonetheless the model cannot estimate MSY-based reference points and this requires proxies. Based on considerations of life history, the WG recommends a proxy of SPR35%. Recommendations for future work include explicitly modelling variability in growth reflecting the declines in mean weight-at-age, incorporating length composition and considering a management procedure approach as the majority of catch comes from ages 1 and 2 which are very poorly informed in catch projection due to the time lag between the assessment and the provision of management advice.

ii Expert group information

Expert group name	**Inter-benchmark process on sardine (*Sardina pilchardus)* in the Bay of Biscay (IBP-Sardine)**
Expert group cycle	NA
Year cycle started	2019
Reporting year in cycle	1/1
Chair(s)	John F. Walter, III, United States
Meeting venue and dates	21–23 October 2019, by correspondence (12 participants)

1 Introduction

An Inter-benchmark Process Review Panel was convened and tasked with the following terms of reference for Bay of Biscay Sardine:

Terms or reference:

a) Evaluate the present analytical assessment method for the stock, particularly;
   1. Investigate the causes of the retrospective patterns in the assessment and potential solutions, including an investigation of fleet's selection patterns and of noise in cohort tracking;
   2. Update the historical French catch time-series.

b) Update the stock annex as appropriate;

c) Re-examine and update MSY and PA reference points according to ICES guidelines (see Technical document on reference points);

d) Develop recommendations for future improving of the assessment methodologies and data collection.

The panel (Annex 1) met by webinar and provides a series of findings and recommendations in this document. Additionally, two external reviewers were present and they provide independent review of the process.

2 Evaluate the present analytical assessment method

This section describes the current assessment model prior to interventions made at the workshop. The current assessment uses Stock Synthesis 3.24 and is parameterized using a matrix of population annual mean weight-at-age. Catch-at-age is available from commercial catches and the PELGAS survey. The model is set up in a single fleet/single area configuration. It relies on three survey indices (the PELGAS biomass estimate, BIOMAN egg count, both carried out annually, and a triennial Daily egg production (DEPM survey). Fecundity is provided as a matrix combining a maturity ogive and weight-at-age from the PELGAS survey. The ogive is built upon the assumption that half of the individuals are mature at gonad maturity stage 2. Prior that stage, all individuals are considered immature. Natural mortality rates at age are assumed constant between year and derived using the method from Gislason *et al.* (2010) and rescaled so values are consistent with total mortality estimates derived from cohort tracking from the PELGAS data (0.8 times Gislason estimates).

Surveys and commercial fleets selectivities are set as follows. PELGAS has a flat selectivity set at 1 for ages 2 to 5. Selectivities for age 1 and age 6+ were estimated by the model. Catchability is an estimated parameter for each survey. DEPM and egg count surveys have selectivity set at 0 at age 0 and flat at 1 for ages 1 to 6+. There is no prior distribution on selectivity parameters. The commercial fleet has 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.

The primary tools for evaluating the current assessment are series of standard diagnostics (Cass-Calay *et al*., 2014; Carvahlo *et al*., 2017) that included jittering initial starting values to evaluate model stability, profiling key parameters (R_0, 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.

2.1 Investigate the causes of the retrospective patterns in the assessment and potential solutions

The primary task of the Group was to evaluate the causes of and potential solutions for the retrospective patterns. Pawlowski *et al*., (2019) considered a series of runs to address issues with the models (**Table 2**). Most of this first approach focused on exploring alternate input data, parameters and assumptions. The first step was to evaluate the revision of the French data (see details in Section c). Those revisions (as well as the other runs), while not solving the scaling issue, improved substantially the magnitude of the retrospective pattern and were later kept as reference for the runs conducted during the inter-benchmark. Other runs consisted of changing the weight-at-age at Q1, allowing for changing selectivity over time, using two commercial fleets and excluding the French catches from the Douarnenez Bay. An assumption was tested about full maturity at maturity stage 2 (ICES, 2008); the former assumption was 50% fish mature. Overall, none of the initial model explorations substantively ameliorated the retrospective bias. For the purposes of further exploration, the model with the revised French catch data (Section c) was chosen as a reference run.

Table 1. Negative log-likelihood and Mohn's rho estimates from the initial series of runs from Pawlowski *et al*., 2019.

Exploratory	Log-likelihood						AIC	BIC	Mohn's rho (AFSC Hurtado)		
assessments	Total	Catch	Eq_Catch	Survey	Age_comp	Recr			SSB	F	Rec
Benchmark run	77.6	4.62E-09	5.63E-04	-14.15	85.95	5.77	227.17	360.50	0.448	-0.360	0.080
Reference (Revised data)	76.3	4.49E-09	4.95E-04	-14.16	84.77	5.72	224.69	358.03	0.425	-0.349	0.066
Alternate raising (stock area = Biscay & Douarnenez)	67.7	5.13E-09	6.76E-04	-15.28	77.49	5.46	207.34	340.68	0.395	-0.335	-0.014
Alternate raising (stock area = Biscay only)	70.1	8.35E-09	9.73E-07	-16.84	81.09	5.81	212.14	345.47	0.278	-0.278	-0.047
Weight@age from Q1 commercial fleet	79.1	4.49E-09	4.08E-04	-12.30	85.71	5.70	230.26	363.60	0.365	-0.315	0.044
2 blocks of selectivity periods	73.5	3.79E-09	2.56E-05	-13.31	80.94	5.88	227.05	375.20	0.529	-0.374	0.290
All fish mature at age 2	78.1	4.57E-09	1.16E-03	-12.75	84.99	5.87	228.24	361.58	0.513	-0.388	0.091
Alternate raising + w@age Q1	71.8	8.13E-09	7.55E-05	-14.96	80.99	5.74	215.54	348.87	0.358	-0.312	-0.040
Alternate raising (Biscay only)+ w@age Q1	69.0	5.09E-09	6.95E-04	-13.86	77.45	5.42	210.05	343.39	0.236	-0.249	-0.072

A series of model explorations were considered at the meeting (**Table 2, Figure 4–6**) several of which evaluated whether the retrospective pattern was a result of the substantial (~27%) decline in mean weight-at-age that appears to be a defining characteristic of the biological state of the stock (**Figure 1**). Additional runs (**Table 2**) considered fixing selectivity at asymptotic for the fishery and the PELGAS survey, removing a single index one at a time, starting the model in 1989 (see (e), below), several growth parameterization and incorporating a Beverton–Holt stock–recruitment relationship estimating sigmaR and using the bias correction of Methot and Taylor (2014). The assumption of full maturity at stage 2 was tested. Given the decrease in weight-at-age, four treatments of growth were considered. First a run with a fixed mean weight-at-age matrix using the first year of data to simply test whether the changes in mean weight-at-age caused the retrospective pattern. Next, the model was parameterized to use the von Bertalanffy growth model, length–weight parameters from Duhamel *et al.*, 2018, and placeholder maturity parameters. These allowed Stock Synthesis to estimate population and fishery weight-at-age internally rather than relying on the input matrix and allowed for growth to be sequentially fixed. Then deviations in K and deviations in growth by cohort were estimated.

Table 2. Negative log-likelihood and Mohn's rho estimates from the exploratory runs carried out during the Inter-benchmark.

Exploratory assessments	Log-likelihood						AIC	BIC	Mohn's rho (AFSC Hurtado)		
	Total	Catch	Eq_Catch	Survey	Age_comp	Recr			SSB	F	Rec
Reference (Revised data)	76.3	4.49E-09	4.95E-04	-14.16	84.77	5.72	224.69	358.03	0.425	-0.349	0.066
Additional runs											
Reference + historical catch back to 1989	82.0	6.44E-09	9.48E-07	-8.835	85.08	5.77	236.05	369.39	0.528	-0.395	0.087
increased equilibrium catch by 2kt	76.3	4.53E-09	4.22E-05	-14.24	84.85	5.68	224.63	357.96	0.450	-0.361	0.092
Standard error of log catch set to 0.1	76.3	1.12E-09	1.85E-03	-14.16	84.77	5.72	224.69	358.03	0.426	-0.349	0.067
BIOMAN indices removed	80.9	4.14E-09	8.90E-05	-8.44	84.10	5.17	233.70	367.04	0.261	-0.243	0.019
DEPM indices removed	76.5	4.46E-09	5.16E-04	-13.89	84.65	5.73	225.03	358.36	0.393	-0.331	0.042
PELGAS indices removed	85.0	4.34E-09	5.87E-04	-4.66	84.03	5.66	242.08	375.42	0.357	-0.313	0.032
Empical constant weight at age	74.2	0.00	0.0	-15.45	84.60	5.04	220.42	353.76	0.387	-0.331	0.130
Constant growth rate	74.9	5.00E-09	7.35E-04	-14.87	84.49	5.25	221.78	355.12	0.374	-0.321	0.087
Time varying growth rate	74.8	5.03E-09	8.68E-04	-15.15	84.50	5.25	259.50	463.21	0.369	-0.319	0.087
Cohort Growth Deviations	73.6	5.25E-09	2.26E-03	-16.81	84.61	5.10	251.16	443.76	0.354	-0.304	0.130
Logistic selectivity at age	77.0	4.68E-09	5.31E-03	-14.38	85.63	5.74	219.99	342.21	0.248	-0.234	-0.051
Logistic selex + Beverton/Holt S/R relationship	63.5	4.26E-09	1.83E-02	-13.53	86.52	-9.49	197.04	326.67	0.107	-0.117	-0.139
same as above + historical series	68.9	6.36E-09	1.36E-05	-8.03	86.38	-9.49	207.72	337.35	0.174	-0.179	-0.119
Candidate runs											
Logistic selex + no zero weight at age values	76.99	4.68E-09	5.31E-03	-14.38	85.63	5.74	219.99	342.21	0.248	-0.234	-0.051
Logistic selex + acoustic selex settings at 0	13520.40	4.48E-09	5.93E-03	13429.80	84.97	5.63	27106.80	27229.02	0.171	-0.180	-0.077
Logistic selex + full maturity at stage 2	78.70	4.77E-09	7.00E-03	-13.00	85.81	5.88	223.41	345.63	0.309	-0.273	-0.038
Logistic selex + alternate raising (Biscay + Douarnenez)	67.84	5.24E-09	3.13E-03	-15.41	77.76	5.48	201.68	323.90	0.227	-0.228	-0.122
Logistic selex + alternate raising (Biscay only)	70.33	8.90E-09	2.03E-03	-17.08	81.45	5.95	206.66	328.88	0.104	-0.141	-0.153
Logistic selex + Beverton/Holt SR (Biscay + Douarnenez)	54.01	4.86E-09	1.60E-02	-14.53	78.54	-10.01	178.03	307.66	0.072	-0.092	-0.176
Logistic selex + Beverton/Holt SR (Biscay only)	57.04	1.07E-08	2.06E-02	-16.34	82.04	-8.69	184.07	313.71	-0.026	0.012	-0.205
Logistic selex + Beverton/Holt SR + Full maturity stage 2	65.52	4.34E-09	2.20E-02	-11.96	86.58	-9.12	201.05	330.68	0.151	-0.156	-0.135
Logistic selex + Beverton/Holt SR + Qpelgas = 1	86.61	9.14E-10	1.22E-02	-11.50	104.25	-6.15	241.22	367.15	0.308	-0.231	-0.259
Logistic selex + Beverton/Holt SR H=0.99	63.17	4.29E-09	1.66E-02	-13.55	86.46	-9.77	196.33	325.96	0.110	-0.121	-0.127
Logistic selex + Beverton/Holt H=0.99 (Biscay + Dz)	53.66	4.88E-09	1.40E-02	-14.55	78.50	-10.31	177.32	306.95	0.076	-0.096	-0.164
Logistic selex + Beverton/Holt H=0.99 (Biscay only)	56.66	1.03E-08	1.69E-02	-16.36	82.00	-9.00	183.32	312.95	-0.022	0.005	-0.190
Logistic selex + Beverton/Holt H=0.99 + Qpelgas = 1	86.39	9.17E-10	1.20E-02	-11.47	104.14	-6.30	240.77	366.70	0.309	-0.232	-0.256
Logistic selex + Beverton/Holt H=0.99 + Full maturity s2	65.08	4.37E-09	1.96E-02	-12.01	86.52	-9.45	200.15	329.79	0.156	-0.161	-0.121
Logistic + BH H=0.99 + Full mat. S2 + no age borrowed	55.30	4.25E-09	1.75E-02	-11.62	75.54	-8.65	180.60	310.24	0.188	-0.173	-0.082
Log. + BH H=0.99 + Fmat.S2 + no age + acoustic selex=0	13498.4	4.04E-09	1.27E-02	13433.2	74.65	-9.53	27066.8	27196.43	0.086	-0.080	-0.061
Preferred run											
same as above with added PELGAS weight@age matrix	56.34	4.41E-09	1.98E-02	-11.31	75.77	-8.14	182.69	312.32	0.147	-0.133	-0.138

The dome shaped selectivity of the PELGAS survey was questioned and, in particular, the lower age 6 selectivity. This shape was estimated during the ICES, 2017 benchmark process. The rationale at the time was a better fit to the observations. However, as there was no evidence to support a lower catchability for fish of ages 6+, a run explored logistic selectivity with full selectivity at age 3. This run substantially reduced retrospective patterns and had a significantly lower AIC. This setting was kept as a first candidate run.

A further test was carried out with catchability equal to 1 for the PELGAS survey. This assumption increased the retrospective bias, reduced the fit substantially and was discarded. The effect of removing one survey at a time was also tested. Retrospective bias was reduced in each case with the highest decrease in Mohn's rho estimate, when BIOMAN was removed and lowest when DEPM was excluded. Nonetheless none of the single index removal treatments ameliorated the retrospective bias to within acceptable (<0.2) levels and these options were not pursued further.

The different treatment of growth only slightly improved the fit of the model and the retrospective bias. Therefore, the growth parameterization was discarded as a major source of the retrospective patterns and these variations of handling growth were not kept for the candidate runs. One reason that estimating time-varying growth and cohort deviations did not substantively change the log-likelihood is that the mean weight-at-age matrix is not included in the likelihood terms as it would be if it were used as input data. Future model considerations may want to fit to this as data. Another run tested whether the mean population weight-at-age for age 0 which was input with a weight of zero created problems for the modelling. This was done by replacing the empirical weights-at-age for age 0 in the beginning of the year with the mean mid-year weight. These did not substantially change or improve overall fit or retrospective bias and therefore the mean weights of zero for age zero at the beginning of the year were left as is.

The lack of set stock–recruitment relationship was considered as a potential source of retrospective bias. Although no stock–recruitment relationship was immediately evident, a Beverton–Holt relationship was specified in SS3 with a steepness (H) parameter set to 0.9 and with sigmaR estimated with the Methot and Taylor (2016) bias correction ramping. This run lowered substantially the magnitude of retrospective patterns while lowering AIC of the model. This run was considered as an intervention. Profiling of steepness indicated that it was not estimable (**Figure 7**). A subsequent run set H to 0.99 to be consistent with the previous model that did not invoke a stock–recruitment relationship. This run slightly improved the fit of the model while slightly reduced retrospective patterns. While the stock–recruitment assumption is weak, the added structure imposed on the recruitment deviations being penalized from a stock–recruitment relationship and the bias correction function substantially improved the retrospective patterns.

The major interventions (logistic selectivity, Beverton–Holt stock–recruitment with both steepness H=0.9 and H=0.99) were then combined with other settings. These settings were tested with the alternate time-series including or excluding the Douarnenez Bay catch (see point c). Including data from DZ Bay simply increased the scale of the population (commensurate with adding ~30% greater catch) but did not substantively change the population trajectory or reduce the retrospective bias. Setting catchability at Q=1 for PELGAS increased the average level of biomass but at the cost of one of the worst log-likelihood values and increased retrospective bias. On this basis, as well as in keeping with general modelling practices to treat most surveys as relative indices, this option was discarded for the candidate run.

The initial maturity assumption was of half of the individuals mature at stage 2 and the final runs considered the new assumption of having all fish mature at stage 2 (ICES, 2008). This assumption increases slightly the magnitude of retrospective patterns but is more in line with the observations during the pelagic surveys in spring, therefore this new assumption was kept for the final runs. Size selectivity pattern for PELGAS was set in the control file to 0 (Selex=1.0 for all sizes) instead of 30 (linked to spawning biomass) as the acoustics detect biomass rather than just mature fish. In the practice of updating the models the indices will be available one year prior to the age composition. In previous model settings the age compositions of the fishery for the terminal year was carried over from the previous year. This is a problem as it will diminish true cohort signals and is unnecessary in Stock Synthesis. Hence, age composition in the terminal year was removed from the model input data in the final run and retrospective analyses were conducted manually to mimic this situation that would occur in practical updating of the models. Furthermore, as the model has no information to estimate terminal year recruitment, e.g. fish born in the terminal year of the model have yet to be seen in the indices or the age composition, the WG recommended to not estimate terminal year recruitment. Working group-preferred model settings are documented below.

2.2 Investigation of fleet's selection patterns and of noise in cohort tracking

The working paper presented to the group evaluated models that included time blocks in selectivity (Pawlowski *et al*., 2019) and found that time blocks did not improve the model fit based on AIC. Similarly evaluation of the Pearson residuals to the age composition for the reference model and the eventual Group-preferred model (**Figure 2**) did not indicate a systematic lack of fit indicative of a change in selectivity or strong diagonal patterns indicative of failing to track cohorts, at least in the aggregated age composition data.

2.3 Evaluate updates to the historical French catch time-series

When the ICES WGHANSA meeting was held in 2018, some substantial downward revision of the French catches had been done for 2016. Some investigations were carried out by Ifremer during the summer 2018, based on production data provided by the French fishing organisations. Some inconsistencies in catches were pointed in some harbour on some quarter. It is unknown why the downward revision occurred in the official databases as data in WGHANSA 2017 had better matches with production data from the fishing organisations. Production data in 2016 were consistent with the official data used at WGHANSA 2017. Therefore it was assumed that the production data reflected the actual catches and were included in this update assessment with a revision from 2013 to 2016 (**Table 3**).

In parallel, some work in the biological datasets (weight@age, number@age) showed some mismatches in French data between catches as used in the stock assessment and the sum of products (SOP) for the respective biological data on some years. This is most likely related to some various revisions to the total catches over time that were followed by rescaling of the number-at-age matrices to match those revisions. These input data were considered by the inter-benchmark as the new reference datasets used for the stock assessment within ICES.

Table 3. Differences between total catch estimates from ICES, WGHANSA 2018 and revised catches. Total catches represent both Spanish and French catches in the Bay of Biscay. Revised catches are the sum of Spanish catches and production data provided by the French fishing industry. For 2018, preliminary catch information from the Spanish and French industry provided in December 2018 (not shown in table).

Year	Total catches (t)	Revised total	Difference
	ICES WG	catches (t)	(t)
	June 2018	December 2018	
2010	20 217	20 217	0
2011	23 208	23 208	0
2012	30 900	30 900	0
2013	32 489	32 938	449
2014	33 943	35 704	1761
2015	27 284	28 756	1472
2016	25 498	29 754	4256
2017	30 318	30 435	117

Another investigation considered inclusion of the two rectangles from the 7e area (25E5&24E5 statistical rectangles called "Douarnenez bay"(DZ bay)). This area is exclusively exploited by a coastal fleet of French purse-seiners (vessels >17 meters) representing around 19% of the total catch (average 2014–2017). The DZ bay was included in the stock, as there is little evidence, at present, to indicate that it should be separated (ICES, 2017a). Considering the high level of catch coming from this area, its poor biological sampling and observed differences in growth between stocks in different regions (e.g. the ICES Subarea 7 stock; Huret *et al*., 2019; ICES, 2017a), the WG raised the need to investigate the differences in biological parameters or stock delineation between the two areas.

Considering the short time window to conduct the study (Lavielle *et al*., 2019) before the Interbenchmark, only two sources of data were used to compare growth and length-at-age: European Union Data Collection Framework (DCF) data from France and *ad hoc* data collected during May 2019. Despite the sampling bias of the first source, it seems there is a higher growth of the DZ bay sardine especially at the first ages. But the erratic sampling of the DCF data from the DZ bay during the time-series and the partial nature of the *ad hoc* sampling do not allow drawing a definitive conclusion. In terms of stock structure based on growth parameters, it is not possible to say if the sardine from the DZ bay belongs to a different stock than those from the north of the 8a area. A complementary study including monthly sampling from the two areas and genetic analysis may be forthcoming. In the absence of clear information to separate the stocks, the group recommends keeping the DZ bay in the stock area definition. Several model runs, described above, including or excluding the DZ catch data, found little improvement in retrospective bias in either case.

2.4 Biological data sampling

Since the current modelling uses catch-at-age data, some concerns about the raising procedure of the catch-at-age matrix were highlighted and the conversion of catch-at-size to catch-at-age through an age–length key (ALK):

- Regarding the French data, no ALK is available during the first quarter of each year, except over 2006–2008. The first semester ALK is borrowed (almost) entirely from the PELGAS survey.
- Difficulties remain in obtaining biological samples from the catch for France and Spain (especially in recent years for France). There is an increasing dependence of the ALK from the survey data (PELGAS/EVOHE) which do not cover some parts of the fishery area (7e part of the stock but also some coastal parts of the Bay of Biscay).

In consideration of the ongoing work dealing with the identification of different growth patterns between the areas of the stock, it would be recommended to keep at least the same level of data coverage and quality. However, moving to an SS configuration that uses the length data input directly and estimated the age–length key internally could provide time and cost savings as the total number of age samples that would need to be collected and aged annually would likely be much lower. Further, the length sampling coverage of the catch is very satisfactory, even on a quarterly basis for the French data as well as for the Spanish data.

2.5 Consideration of the historical (pre-2000) Spanish purse-seine data

The Group considered the historical Spanish (Basque) purse-seine catch data from 1971–1986 (**Figure 3**). The early data in Figure 3 were considered to be unreliable, both due to the potential that the catches came from areas other than the Bay of Biscay, and due to unreliable reporting. Secondly, much of the catch went to local markets at that time, rather than canneries and hence it would be unlikely that the local Basque markets could have consumed such volumes. Hence the Group agreed with the decisions of previous working groups to exclude this catch time-series. The Group did consider that there was likely a baseline catch of around 2000 t in the period prior to 2000 and that the most appropriate way to account for this was to add this to the initial equilibrium catch and to increase the CV on this initial input in the assessment from 0.05 to 0.10 to account for the fact that these catches were not known precisely. The purpose of the initial equilibrium catch is to allow the model to estimate initial fishing mortality rates, a critical aspect of starting the model later than the start of fishing. The implementation of these settings did not change substantially the overall fit of the model. Mohn's rho values were similar to the reference run for the CV of 0.1 while adding 2000 t to the initial equilibrium catch slightly increased the magnitude of retrospective patterns. Those settings were not kept for the candidate runs.

2.6 Exploration of an earlier starting point

Best practices in integrated modelling generally attempt to account for the entire history of removals, and start the model at the earliest possible time period. This obviates the need to estimate initial fishing mortalities and may improve the estimation of R_0 as the model has a high negative correlation between R_0 and initial F. Hence starting the model earlier could improve the model.

Given the perceived quality of the catch data, the Group considered it possible to push the starting time earlier to 1989 and models were constructed to do this. These models (**Table 2**) did not substantively improve the precision of the initial F estimates over the initial reference model, nor did it reduce the correlation between R_0 and initial F. Furthermore, it did not ameliorate the main issue of the retrospective pattern. Hence, while the Group considers that efforts to reconstruct the historical catch time-series could improve the model, it would require further review of the available data and was not necessary to pursue at this point to address the retrospective pattern.

2.6.1 Recommendations for Group-preferred model formulation

The group identified a series of additional recommended changes to the model outlined below:

- asymptotic selectivity (fix parameter 6 selectivity at 1) for both PELGAS and fishery;
- Beverton–Holt stock–recruitment relationship, fixed steepness at 0.99, estimated;
- sigmaR with Methot and Taylor bias correction;
- remove the relict selectivity deviation parameterization;
- Retain start in 2000;
- Add 2000 to the initial equilibrium catch and to increase the CV on this initial input in the assessment from 0.05 to 0.10;
- Do not estimate terminal year recruitments as these are little informed by any data (this is commensurate with general practice of other modelling approaches to replace terminal year recruitments in VPAs or to employ some type of 'shrinkage' to the mean penalty);
- do not carry over terminal year age composition;
- full maturity at stage 2.

2.6.2 Diagnostic evaluation of the Group-preferred model

The group performed and report here upon a series of standard diagnostics on the preferred model settings indicating that the model. Profiles of steepness (**Figure 7**) indicate that there is limited ability to estimate steepness, which is not surprising given the short duration of the modelled time-series and the lack of contrast in indices. The model has a log-likelihood minima for R_0 which indicates that it can estimate population scale, which is a key function of the modelling (**Figure 8**). The main signal for estimating scale is in the age composition data with the relatively flat indices uninformative for the upper limit on R_0.

Retrospective performance is acceptable with Mohn's rho values less than 0.2 (**Table 2, Figure 9**). Fits to the aggregated age composition (**Figure 10**) are good though fits to indices are rather poor (**Figure 11**).

Overall, the group-preferred model is not substantively different in scale, trend or variability in SSB, recruitment or fishing mortality from the initial benchmark model or the reference model with the revised data (**Figure 12**). The preferred model simply has several interventions that lead to more stable treatments of recent recruitments, which substantially reduces the retrospective bias which was the major concern of this IBP panel.

2.6.3 Overall evaluation of the model

Strengths: Overall model performance is solid, log-likelihood profiling indicates that the model estimates the key parameters without severe conflict among components. The Group-preferred model with asymptotic selectivity, a restriction to not estimate terminal year recruitment deviations and a loose (steepness 0.99) Beverton–Holt stock–recruitment relationship reduced the retrospective bias considerably and several interventions improved model fit.

Weaknesses: The model cannot estimate steepness, so it does not have an internal estimate of F_{MSY} from a stock–recruitment relationship. Furthermore, the selectivity, growth, natural mortality and fecundity parameterization does not lend itself to well-defined equilibrium proxy benchmarks based on yield or spawner per recruit. Maturity, though much better resolved with the new staging criteria for ages 2+, remains an uncertainty in the model for age 1 and, as it affects the interpretation of the indices, remains influential on model performance.

The issue of scale remains in that the model does not estimate a population size as large as the acoustic survey and the survey (DEPM, PELGAS and BIOMAN) catchability estimates remain above 1 which indicates that they potentially detect more fish than the model estimates are in the population. However, this makes the assumption that the surveys are actual estimates of absolute biomass. The assessment model does not treat the survey as absolute estimates and it estimates catchability, which really is not meaningful, once the survey is treated as a relative abundance index. Treating surveys such as these as relative indices is commensurate with most treatments of these in models, as estimating absolute abundance is very difficult and prone to numerous strong assumptions.

While the model estimates of catchability for the surveys are above one, they could very well have been below one if the indices were scaled to a mean of one as is often done with relative indices. The issue with them being above one only is a concern if it is considered that these estimates are absolute estimates. In this regard, the Group is aware of ongoing research into the acoustic surveys that may address concerns regarding overestimation of target strength that could be a factor in its estimation of total biomass higher than the assessment model. Also the selectivity of the acoustic survey is modelled with data from pelagic trawls, which does not appear to capture fish less than 10 cm. If the acoustic survey detects fish smaller than this, particularly in the massive quantities of 5–10 cm fish, then perhaps they are under-represented in the assumed vulnerable biomass in the model. Furthermore, given the rapid population dynamics (high natural mortality) and high seasonality of the biology and fishery, it is possible that any mismatch in the timing of the acoustic surveys relative to the assumed timing of growth, reproduction and mortality could lead to the acoustic survey to estimate higher biomass than the model estimates. For example, given natural mortality alone, about 50% of the fish die during the year. If M is seasonal, which it likely is (but is currently modelled as constant throughout the year) then a pre-natural mortality survey in the spring would clearly estimate higher biomass than under a constant M assumption. Finer scale partitioning of temporal dynamics might better resolve any mismatches in the future but, given the current uncertainties, treating the surveys as relative indices appears prudent.

Future configurations of the model should consider specifically parameterizing all of the biological inputs, so as to allow for switching back and forth between reliance on the mean weight-at-age matrix and using internally derived or estimated weight-at-age. This will also allow for inputting the mean weight-at-age as data rather than just fixed known values. This is of particular importance for fitting growth as the mean weight-at-age matrix does not contribute to the likelihood and hence, when growth was estimated in several of the trials, the model did not gain much improvement in likelihood, as it would have if it fit to these data. Further explorations of the model should increase the resolution on the length composition bins and could consider options for either iterative reweighting of the composition data using one of the standard approaches such as Francis (2011).

3 Update of stock annex

3.1 MSY and PA reference points

The group recommended an FSPR proxy of $F_{35\%SPR}$, based on considerations of life history and precautionary reference points (Myers *et al.*, 1999; Mace, 1994; Mace and Sissenwine, 1993) and proxies for F_{MSY} based on natural mortality rate (Zhou *et al.*, 2012; Francis, 2011). Recalculated reference points are included in **Table 3** based on the Group-preferred model.

From the group-preferred model developed in this Inter-benchmark (up to 2018), the sardine in 8abd shows a scatterplot of 18 pairs of stock and recruitment estimates (2000–2017), covering a narrow range of biomasses (Min/Max=51%) and with no clear indications of impaired recruitment (**Figure 13**). According to ICES guidelines for stocks in category 1 and 2, when the assessment covers a limited range of biomass and it is not informative about B_{lim} (i.e. does not show evidence of past impaired recruitment) the stock can be classified as Type 6. In these cases, ICES guidelines (ICES, 2017c) suggest that B_{loss} could be a candidate for B_{pa}, but this being dependent on considerations of the historical fishing mortality. If fishing mortality has been low (as it can be the case for most of the series assessed for sardine), "then this may actually be a stable stock for which the B_{pa} should be defined as the B_{loss} value". These guidelines were followed in WGHANSA 2017 and 2018 and B_{pa} was set as B_{loss}. In WGHANSA 2018 (ICES, 2018) B_{pa} was set at 88 000 t. Then, B_{lim} was set as a function of B_{pa} as follows: $B_{lim} = B_{pa}$ x exp(-1.645 σ_B), which resulted in $B_{lim} = B_{pa} / 1.4$ = 63 328 t ,where $\sigma_B = 0.2$ as indicated in the ICES guidelines in a deliberate precautionary approach when inferring B_{lim} from B_{pa}. The guidelines for all the remaining reference points were followed leading to F_{MSY} equal to 0.27 (ICES, 2018). The WG questioned the use of a default sigmaB (0.2) instead of the actual SigmaB estimated by the model (of 0.23 in WGHANSA 2018) as an overly precautionary approach for a type 6 stock (where the first precautionary decision was to take B_{pa} as B_{loss}). ADGHANSA (ICES, 2018b) supported the use of sigmaB=0.23 for future determinations of B_{lim} from B_{pa}.

These estimates were supported but also partly questioned in 2017 for the relatively precautionary values resulting from the ICES ad hoc procedure for this sardine stock (see Annex 2 to ICES, 2017b).

The Inter-benchmark team points out that the past sardine BRPs resulting from the default procedure for this type 6 stock might have resulted in values too precautionary for the following considerations:

- constrained to $F_{p0.05}$ (0.27) to 61%. This means that adopted F_{MSY} (0.27) is well below the typical F_{MSY} proxies at %SBR of 40% or 50% (Mace, 1994; Horbowy and Luzenczyk, 2012) (Table 1 in Uriarte *et al*. WD to IBP and see also Table 1 below in both tables assuming SSB_0 corresponding to geometric mean recruitment). This suggests that the restrictions imposed by $F_{p0.05}$ to ensure a 95% probability of being above B_{lim}, may be too high, and so B_{lim}.
- Constraining F_{MSY} by $F_{p0.05}$ leads to a F_{MSY} value of 0.27 which is well below the typical cautionary target of $F_{0.1}$ (**Table 4**) and of the alternative F_{MSY} proxy of 0.87*M (= 0.44, see last paragraph of this section) (based on life history for data-limited stocks Zhou *et al.*, 2012). B_{lim} would be above the expected biomass at $F_{0.1}$ (as calculated for this stock in the deterministic yield per recruit of Table 1 in Uriarte *et al.*, 2019, and see also Table 1, below).
- Unrestricted F_{MSY} (0.415) corresponds a %SBR of about 52%, F_{pa} (0.302) to 59% of F_{MSY}.

The IB group explored some alternative definitions of B_{lim} in terms of %SBR. Based on the assessment results, an equilibrium SPR analysis was conducted. R_0 was equal to 5 025 398 (1000s), as estimated in the model. Maturity, stock and catch weights-at-age and selectivity were taken as the average of the last five years (2013–2017). No assumption of a stock–recruitment relationship was made as the assessment itself cannot estimate steepness for the Beverton–Holt S–R model (other than suggesting that steepness values higher than 0.6 are more likely). Values of B_{lim} at 20%, 30% and 35% SBR were considered (**Table 4**). The 20% threshold has been suggested as a biomass below which recruitment can be reduced (Beddington and Cooke, 1983; Meyers *et al.*, 1994). However, Mace (1994) and Mace and Sissenwine (1993) pointed out that for stocks of unknown resilience a more prudent approach would be using $F_{30\%B0}$. Furthermore in their analysis Mace and Sissenwine (1993) found that pelagic species that reach relatively small maximum size and/or mature at small size, seem to have high replacement %SPR, and the analysis by taxonomic groups suggested a mean replacement %SPR for cupleoids of about 37.5% higher than for other taxonomic groups. Myers *et al.* (1999) also found that the median steepness of cupleoids and engraulidae were intermediate (not in the upper range of values). Therefore, it can be deduced or presumed from a precautionary approach that small pelagic fish may have relatively lower resilience to fishing (Mace and Sinsenwine, 1993). This led the IBP group to set B_{lim} at 35%B0, which was equal to 56 300 t (Table 1).

Other ways of estimating B_{lim} based, for instance on SSB leading to a Recruitment level at 50%R0 (as suggested by Mace, 1994), were not considered as they imply assuming the steepness of the S–R relationship.

Following the ICES guidelines for stocks in Category 1 and 2, the remaining reference points were derived from the former value of B_{lim} (= 56 300 t). B_{pa} was derived as $B_{pa} = B_{lim} \times \exp(1.645\ \sigma_B)$, where σ_B is the standard deviation of ln(SSB) in the terminal year (2018). The value of σ_B was approximated by the coefficient of variation of SSB in the final assessment year, which resulted to be 0.204. This value was very close to the default value of 0.2. Thus, B_{pa} was set at 78 700 tonnes.

The limit fishing mortality (F_{lim}) is the F that, in equilibrium from a long-term stochastic projection, gives 50% probability of SSB being above B_{lim}. This was computed using Eqsim for a projection based on stochastic recruitment around a segmented regression with breakpoint fixed at B_{lim} (Figure 2). Biological parameters (mean weights-at-age, maturity and natural mortality) and exploitation pattern (selectivity) were sampled from the last five years of the stock assessment (2013–2017). No assessment/advice errors were considered ($F_{cv} = F_{phi} = 0$) and no advice rule was included ($B_{trigger}$=0). The resulting limit fishing mortality F_{lim} was 0.757.

The precautionary approach fishing mortality F_{pa} is the value of the estimated F that ensures that the true F has less than 5% probability of being above F_{lim}, i.e. the 5th percentile on distribution of the estimated F if true F is at F_{lim}. Thus, F_{pa} was derived from F_{lim} as: $F_{pa} = F_{lim} \times \exp(-1.645\ \sigma_F)$, where σ_F is the standard deviation of ln(F) in the final assessment year. The standard deviation of the logarithm of F in 2017 was approximated by the coefficient of variation of the apical F in 2017, which was equal to 0.207, leading to F_{pa} at 0.539.

For the stochastic projections in Eqsim to compute F_{MSY} and MSY $B_{trigger}$, recruitments are sampled from the predictive distribution of fitted parametric stock–recruitment models. Initially, Beverton–Holt, Ricker and segmented regression stock–recruitment models were considered, and the fitted models were averaged using smooth AIC weights (Buckland *et al.*, 1997). However, the fit of the Beverton–Holt was unrealistic (a flat line) and no biological support was found for the Ricker model (all observed points in the impaired recruitment region). Alternatively, the breakpoint of the segmented regression model was slightly lower than the lowest observed SSB (which in this case was used to define B_{pa}). Therefore, it was decided to use a segmented regression model with the breakpoint fixed at B_{lim} (**Figure 14**).

Biological parameters (weights-at-age, natural mortality and maturity) and the exploitation pattern (selectivity) were resampled at random from the last five years of the assessment (2013–2017). This makes the explicit assumption that the declines in mean weight-at-age (**Figure 1**) observed in the population in the last five years will continue in the future. Assessment/advice errors could not be estimated for this stock, since the model was not used in the latest years to provide advice. Therefore, assessment/advice errors were set according to the default option in WKMSYREF4 (ICES, 2016). The conditional standard deviation in the log domain was FCV=0.212 and the parameter of autocorrelation in the AR (1) process for fishing mortality was Phi=0.423. The biomass trigger point ($B_{trigger}$) was fixed at 0, indicating that the ICES MSY advice rule (fishing mortality is linearly reduced if the biomass in the TAC year is predicted to be lower than MSY $B_{trigger}$) was not applied. All the settings for the base case run in Eqsim are given in **Table 5**.

F_{MSY} was computed as the F maximizing the median landings yield curve and was equal to 0.621. Since this value was larger than F_{pa}, for consistency with the precautionary approach F_{MSY} was reduced to F_{pa} (0.539) (**Figure 15**).

MSY $B_{trigger}$ in the ICES MSY advice rule is defined as the 5th percentile of the distribution of SSB when fishing at F_{MSY} and could be calculated via stochastic simulation in Eqsim. From 2002 to 2011, fishing mortalities were below 0.2, increased around 0.4 in 2012–2017. In the absence of fishing at F_{MSY}, MSY $B_{trigger}$ was set at B_{pa} (78 700 tonnes).

The effect of including the ICES MSY advice rule was evaluated by running Eqsim with $B_{trigger}$ equal to MSY $B_{trigger}$ at 78 700 tonnes. $F_{p.05}$, the F that leads to SSB >B_{lim} with probability 0.95, resulted in 0.453 when including the ICES MSY advice rule. However, this value was still below F_{MSY}, indicating that the F_{MSY} and MSY $B_{trigger}$ combination do not fulfil the precautionary criterion (**Figure 16**). Therefore, F_{MSY} was further reduced to $F_{p.05}$ at 0.453. The PA and MSY reference points are summarised in **Table 6**.

The final estimate of F_{MSY} (over ages 2–5) (= 0.453) has the property of being consistent with the ideas of Zhou *et al*. (2012) of setting F_{MSY} equal to 0.87·Natural Mortality. Making use of the weighted M (to the expected equilibrium populations at ages 2–5 at R_0, wM=0.502) leads F_{MSY} proxy = 0.87 * 0.502= 0.44. In addition, the current F_{MSY} value is also more consistent with suggested values in literature for %SBR as it seems to fall between $F_{50\%SPR}$ and $F_{40\%SBR}$ (closer to the former one).

3.2 Projection specifications

Given the short duration of the meeting, the Working Group could not fully specify the projection specifications for the preferred model structure but anticipates this will be more fully covered in the assessment update. Key considerations for projections such as time period for biological assumptions, selectivity and fishing mortality specifications could (and, to be consistent with the benchmark calculations, must) follow the specifications outlined in **Table 5**.

4 Recommendations

1. Consider Group-preferred model, defined above.
2. Evaluate methods for estimating natural mortality, particularly considering the work conducted for Bay of Biscay Anchovy (Uriarte *et al.*, 2016). Consider moving to a single parameter for natural mortality (but with a Lorenzen-scaling) so that it can be profiled.
3. Explore incorporation of length and conditional age at length data into the model to allow estimation of growth internally as well as to estimate time-varying growth processes.
4. Consider explicit incorporation of environmental covariates on growth. This may not be critical for modelling historical growth but is highly influential when parameterizing the expected growth in the future, e.g. one needs to assume what growth (or weight-at-age) defines the 'benchmark' conditions.
5. Incorporate mean weight-at-age as 'data' in the data file, rather than as fixed, assumed-known values.
6. Given the fact that most of the catch is from ages 1 and 2 and the likely two year time-lag between a stock assessment model and the provision of catch advice much of the catch that constitutes the TAC is a product of model projections. This makes the assumption of future recruitment highly influential in the TAC advice (Rice and Browman, 2014). As this fishery has high quality annual surveys that should serve as useful management procedures, the fishery could benefit from an MSE-tested management procedure.
7. Evaluate the historical time-series of Basque/Spanish catches. If even a quarter of these historical catches were from the Bay of Biscay, they would scale the overall population substantially higher.
8. Carry out further studies to clarify the stock identity of the catches coming from the Douarnenez Bay (in Subarea 7), currently assumed to belong to the sardine stock in Subarea 8, which was shown to affect the scaling of the assessment. Particularly conduct biological sampling in Douarnenez to obtain annual ALKs for this area and to evaluate stock mixing and potential growth differences.
9. BIOMAN surveys should estimate maturity within the survey so that they are not reliant on PELGAS maturity.
10. Recommend exploring method such as the Francis/McAllister or other reweighting for composition data be employed which might give more weight to the composition data.

5 Conclusions

Overall the group-preferred model is not substantively different than the benchmark model, it simply has several interventions (asymptotic selectivity and some structure on unconstrained recruitment deviations) that greatly improve diagnostic performance. The data revisions appear justifiable and further do degrade model performance. The other major issue was concern over the different population scales as estimated by the acoustic and egg surveys what were designed to provide absolute biomass estimates. However several possibilities such as mismatch in assumed timing, overestimation of target strength or incorrect selectivity assumptions for the surveys could all lead to survey biomasses higher than the model. Given the extreme challenges in obtaining absolute biomass estimates, particularly for a highly temporally dynamic biomass such as sardine treating these surveys as relative indices appears justified. Once the indices are treated as relative then the absolute scaling of catchability does not matter.

Hence the main concerns that lead to the assessment being classified as a category 2 stock (stocks with analytical assessments and forecasts that are only treated qualitatively): poor retrospective performance and differences in scale between the surveys and the model appear resolved at the current point for which the working group recommends elevation of the assessment to category 1 and provides a series of reference point calculations based on an MSY proxy of $F_{35\%SPR}$.

6 Literature cited

Beddington, J.R. and Cooke, J. 1983. The potential yield of previously unexploited stocks. FAO Fisheries Technical Paper No. 242, 63 pp.

Buckland, S., Burnham, K., and Augustin, N. 1997. Model Selection: An Integral Part of Inference. *Biometrics*, *53*(2), 603–618. doi:10.2307/2533961.

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

Cass-Calay, S. L., Tetzlaff, J., Cummings, N., Isely, J. 2014. MODEL DIAGNOSTICS FOR STOCK SYNTHESIS 3: EXAMPLES FROM THE 2012 ASSESSMENT OF COBIA IN THE U.S. GULF OF MEXICO. Collect. Vol. Sci. Pap. ICCAT, 70(5): 2069–2081.

Duhamel, E., Doray, M., Huret, M., Sanchez, F., Marie-Lepoittevin, T., Peltier, H., Autthier, M. 2018. Direct assessment of small pelagic fish by the PELGAS18 acoustic survey. Working Document for WGHANSA. Lisbon.

Francis, R. C. 1974. Relationship of Fishing Mortality to Natural Mortality at the Level of Maximum Sustainable Yield Under the Logistic Stock Production Model. Journal of the Fisheries Research Board of Canada. 31(9): 1539–1542.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68: 1124–1138.

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.

Horbowy, J., and Luzeńczyk, A. 2012. The estimation and robustness of FMSY and alternative fishing mortality reference points associated with high long-term yield. Canadian Journal of Fisheries and Aquatic Sciences, 69: 1468–1480.

Huret M., Duhamel E., Gatti P., Lebigre C., Le Mestre S., Petitgas P., Rabiller M., Guigue T., Joguet M., Maillet J. 2019. Final report of the CAPTAIN project: Connaissances Améliorées à l'aide des Pêcheurs sur la sArdIne et l'ANchois de la façade Atlantique - Bilan 2015–2018.

ICES. 2008. Report of the Workshop on Small Pelagics (*Sardina pilchardus*, *Engraulis encrasicolus*) maturity stages (WKSPMAT), 10–14 November 2008, Mazara del Vallo, Italy. ICES CM 2008/ACOM:40. 82 pp.

ICES. 2016. Report of the Workshop to consider FMSY ranges for stocks in ICES categories 1 and 2 in Western Waters (WKMSYREF4), 13–16 October 2015, Brest, France. ICES CM 2015/ACOM:58. 187 pp.

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

ICES. 2017b. Report of the Working Group on Southern Horse Mackerel, Anchovy and Sardine (WGHANSA), 24–29 June 2017, Bilbao, Spain. ICES CM 2017/ACOM:17. 640 pp.

ICES. 2017c. ICES Advice Technical Guidelines. ICES fisheries management reference points for category 1 and 2 stocks. ICES Advice, Book 12. (Published 20 January 2017). DOI: 10.17895/ices.pub.3036.

ICES. 2018. Report of the Working Group on Southern Horse Mackerel, Anchovy and Sardine (WGHANSA), 26–30 June 2018, Lisbon, Portugal. ICES CM 2018/ACOM:17. 639 pp.

ICES. 2018b. ADGHANSA 2018 Minutes ICES Offices, 04 and 05 July 2018.

Lavialle, G., Duhamel, E., Véron M. 2019. Preliminary results on the comparison of the sardine growth between the channel part (>48°N; Douarnenez Bay) and the south part of the bay of Biscay Sardine. ICES Working document, Sardine IBP.

Mace, P.M. 1994. Relationships between common biological reference points used as thresholds and targets of fisheries management strategies. Can. J. Fish. Aquat. Sci. 51(1): 110–122. doi:10.1139/f94-013.

Mace, P.M. and Sissenwine, M.P. 1993. How much spawning per recruit is enough? Risk Evaluation and Biological Reference Points for Fisheries Management (eds S.J. Smith, J.J. Hunt and D. Rivard). Canadian Special Publication in Fisheries and Aquatic Sciences No. 120, National Research Council of Canada, Ottawa, 101–118.

Methot, R.D. and Taylor, I.G. 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. *Can. J. Fish. Aquat. Sci.*, 68:1744-1760.

Myers, R.A., Bowen, K.G., and Barrowman, N.J. 1999. Maximum reproductive rate of fish at low population sizes. Can. J. Fish. Aquat. Sci. 56: 2404–2419. doi:10.1139/f99-201.

Pawlowski *et al*. 2019. WD xx. Investigating the causes of low biomass estimates in assessment and retrospective patterns using exploratory assessments using SS3 on Bay of Biscay sardine stock. ICES Working document, Sardine IBP.

Rice, J. and H. I. Browman. 2014. Where has all the recruitment research gone, long time passing? ICES Journal of Marine Science 71(8): 2293–2299.

Uriarte, A., Ibaibarriaga L., Pawlowski L., Massé J., Petitgas, P, Santos, M., Skagen, D. 2016. Assessing natural mortality of Bay of Biscay anchovy from survey population and biomass estimates. 2016. Canadian Journal of Fisheries and Aquatic Sciences, 2016, 73(2): 216–234.

Uriarte, A., Lavialle, G., Pawlowski, L., Ibaibarriaga, L., Citores, L. and Erwan Duhamel, E. 2019. Preliminary results on reference points for sardine in 8abd. ICES Working document, Sardine IBP.

Zhou, S., Yin, S., Thorson, James T., Smith, Anthony D. M., Fuller, M., and Walters, C. J. 2012. Linking fishing mortality reference points to life history traits: an empirical study. Canadian Journal of Fisheries and Aquatic Sciences, 69: 1292–1301.

Table 4. Yield per recruit analysis for a constant recruitment at Virgin Recruitment estimated by SS3 (R0 = 5 025 398) leading to Virgin Biomass (B_0) around 160 800 t. (Yield had a monotonic increasing shape and hence F_{max} was not encountered). The small difference of SSB at $F_{35\%B0}$ (56 371 t) with the actual 35%B_0 (56 277 t) arises from the step of increasing F while covering the wide range of Fs for the analysis).

	Fmax	Fbar.0.1	F20%SBR	F30%SBR	F35%SBR	F40%SBR	F50%SBR
Fbar(2–5)	**0.000**	**0.810**	**3.035**	**1.385**	**1.025**	**0.780**	**0.480**
Fbar(1–3)	**0.000**	**0.637**	**2.387**	**1.089**	**0.806**	**0.613**	**0.377**
HR(B1+)	0.000	0.416	0.980	0.604	0.492	0.404	0.280
HR(SSB)	0.000	0.514	1.414	0.788	0.621	0.498	0.332
SSB	0	63,315	32,182	48,327	56,371	64,477	80,430
Yield	0	32,547	45,500	38,079	35,031	32,140	26,724
%SBR	0%	39%	20%	30%	35%	40%	50%

Table 5. Settings for the Eqsim runs for sardine in 8abd.

Data and Parameters	Setting	Comments
SSB-recruitment data	Full time-series (2002–2015)	
SR models	Segmented regression with breakpoint at B_{lim}	
Mean weights, maturity and natural mortality	2013–2017	
Exploitation pattern	2013–2017	
Assessment error in the advisory error (Fcv)	0.212	Default value
Autocorrelation in assessment in the advisory year (Phi)	0.423	Default value

Table 6. PA and MSY reference points for sardine in 8abd.

Framework	Reference point	Absolute value	Technical basis
MSY approach	MSY $B_{trigger}$	78 700	B_{pa}
	F_{MSY}	0.453	$F_{MSY} = F_{p.05}$
Precautionary approach	B_{lim}	56 300	35%SPR, i.e. equilibrium biomass at F that leads to 35% of spawner of recruit without fishing
	B_{pa}	78 700	$B_{pa} = B_{lim} \times \exp(-1.645 \times sigma)$, where sigma=0.2
	F_{lim}	0.757	F that results in 50% probability that SSB is above B_{lim} in the long term, using segmented regression with B_{lim} (EqSim)
	F_{pa}	0.539	$F_{pa} = F_{lim} \times \exp(-1.645 \times sigma)$, where sigma=0.28
Management plan	SSB_{MGT}	Not applicable	
	F_{MGT}	Not applicable	

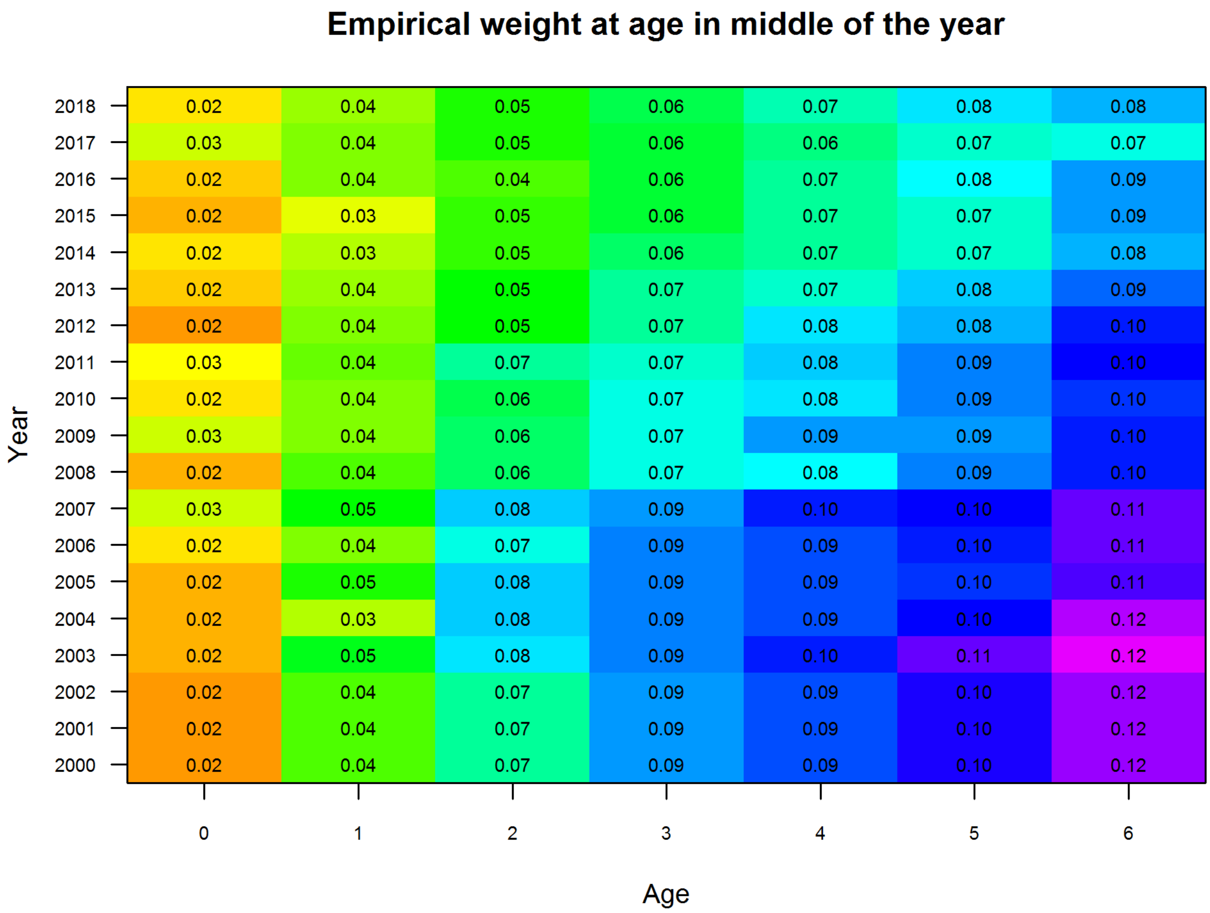

Figure 1. Empirical weight-at-age of population in the middle of the year.

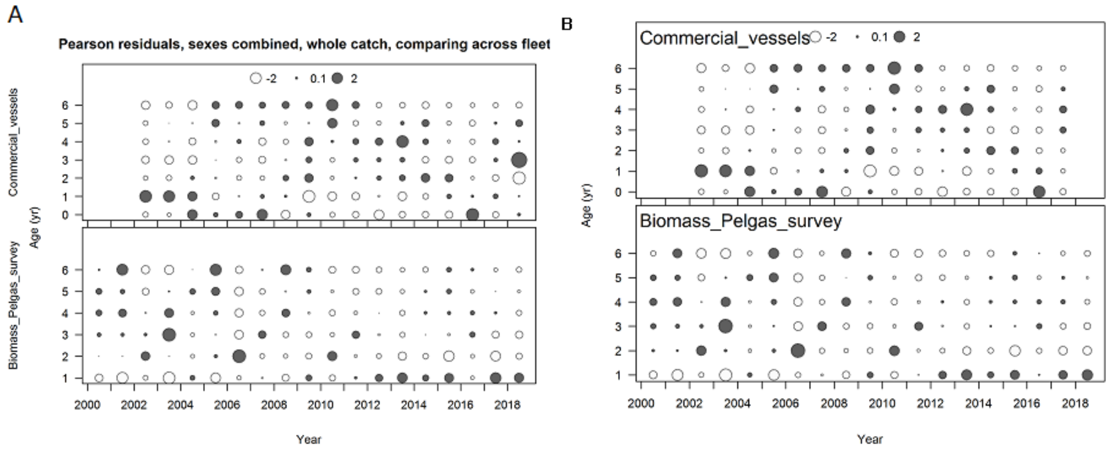

Figure 2. Pearson residuals for A. Initial Reference model, Commercial_vessels (max=3.96), and Pelgas_survey (max=2.85), B. Group preferred model, Commercial vessels (max=2.52) and, Pelgas survey (max=3.03).

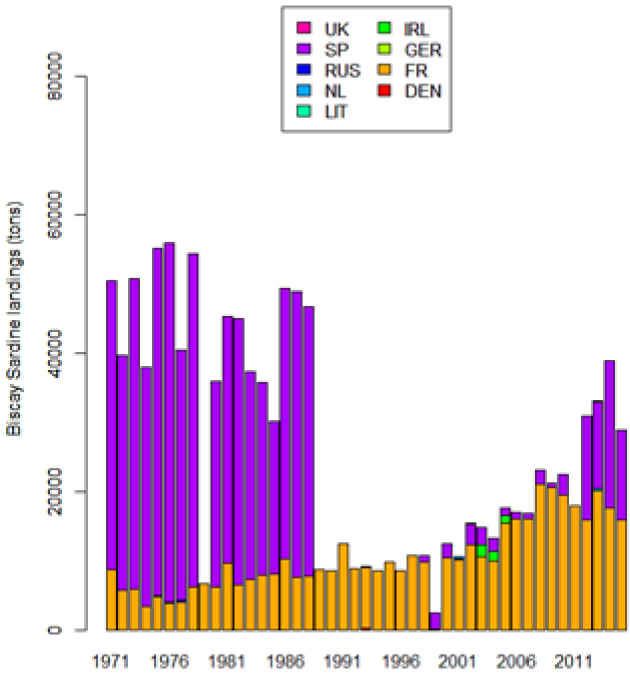

Figure 3. Historical landings of sardine in 8abd by country.

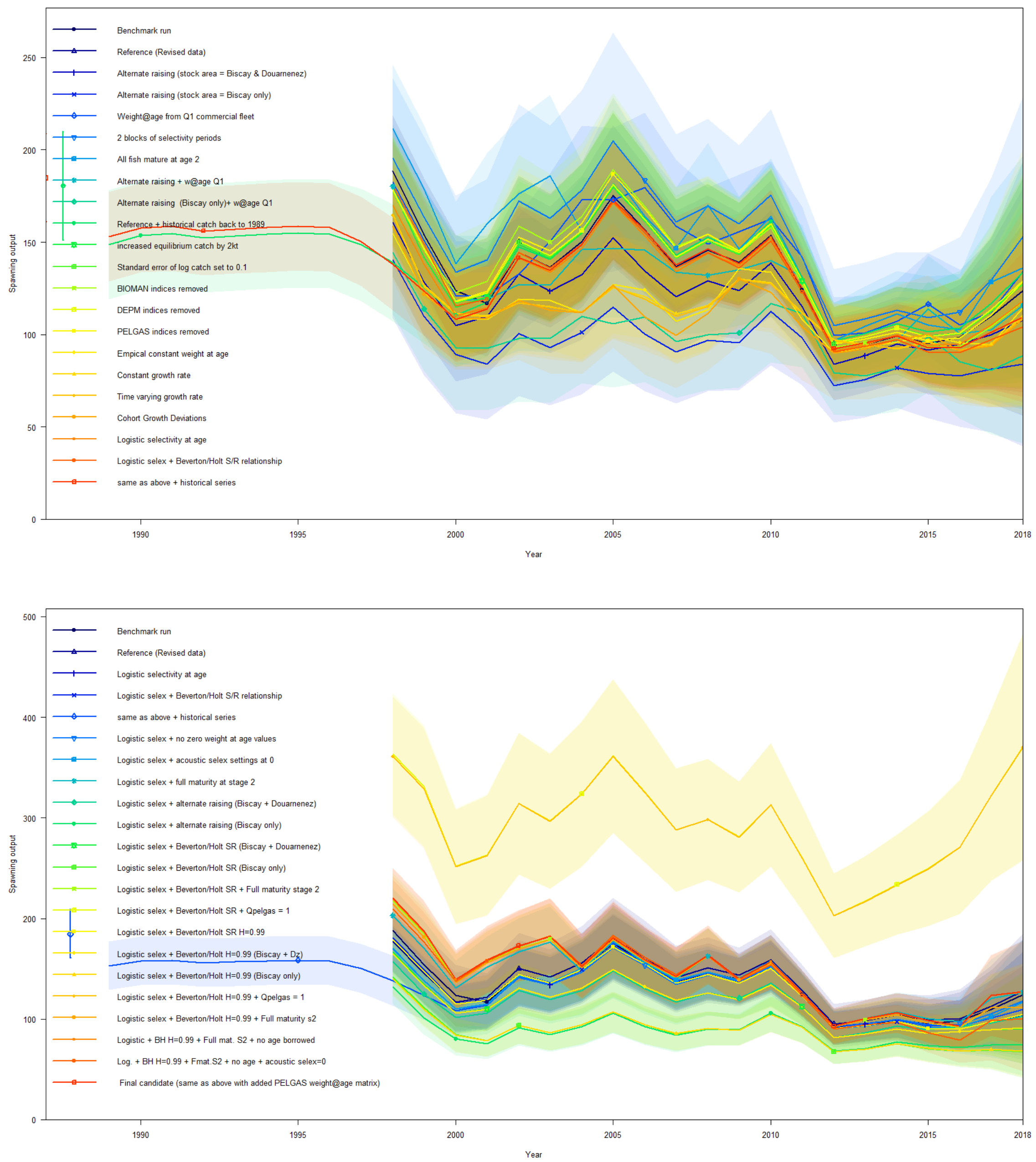

Figure 4. Spawning output for all model runs.

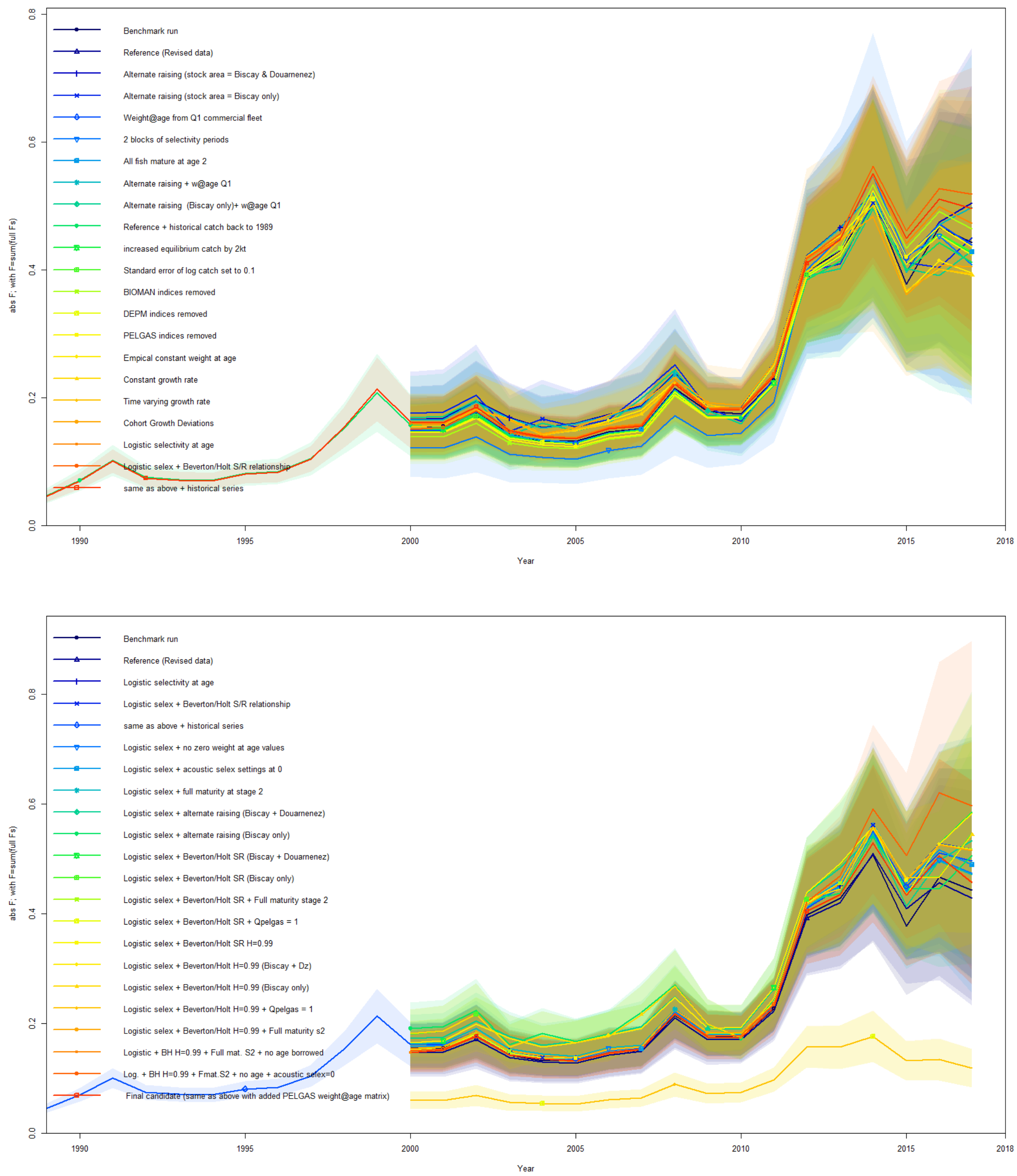

Figure 5. Fishing mortality (apical) output for all model runs.

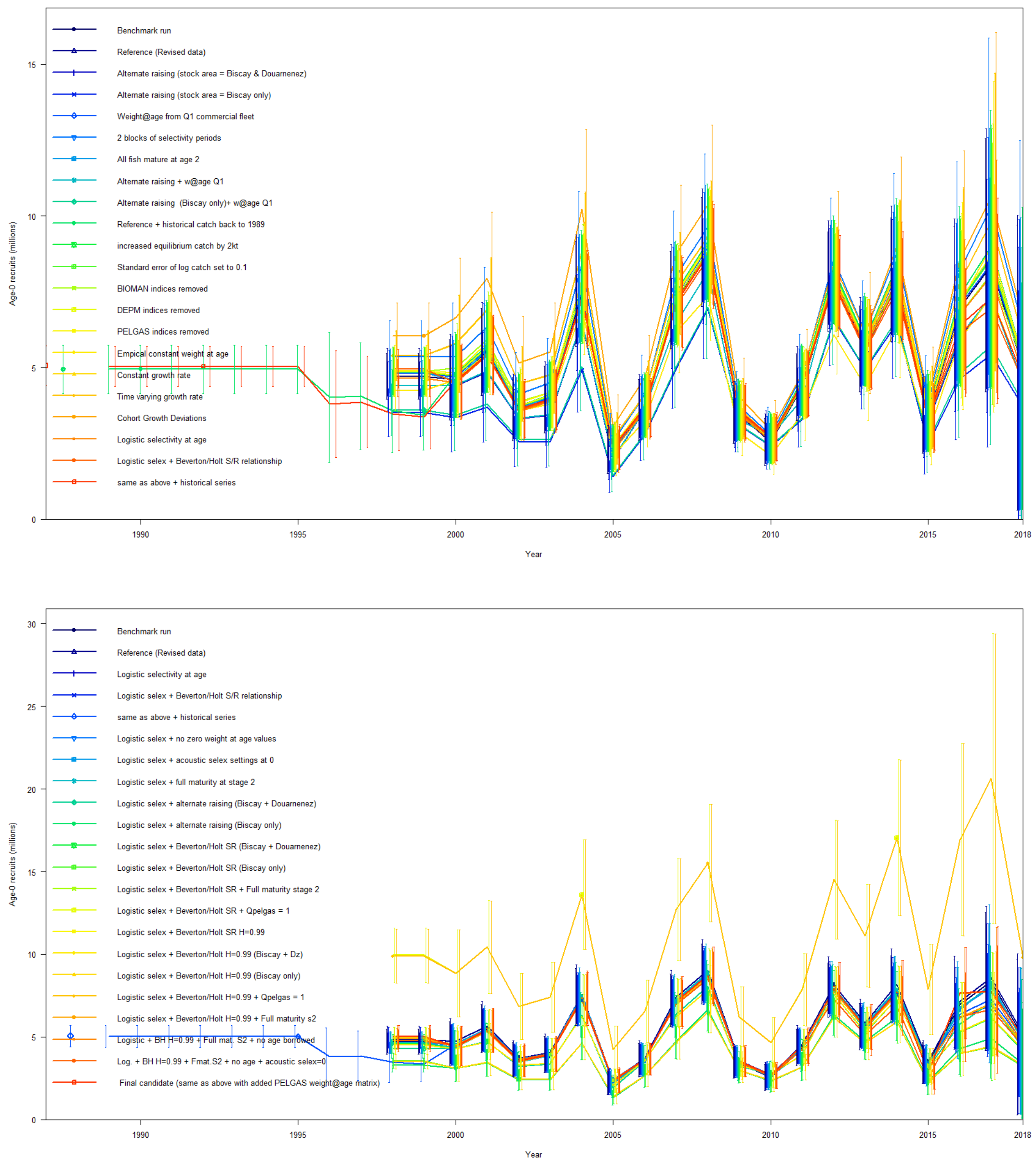

Figure 6. Recruitment output for all model runs.

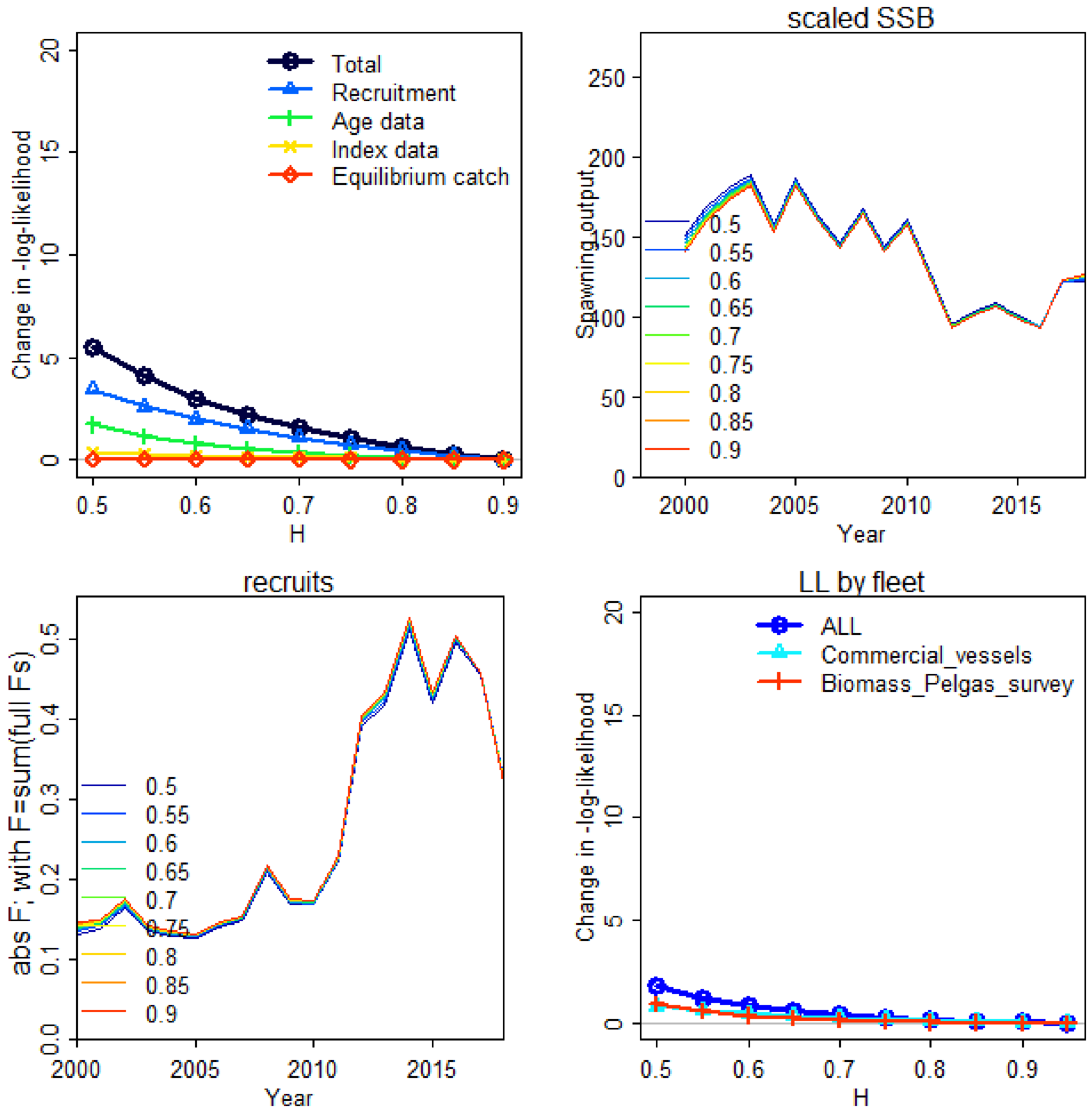

Figure 7. Steepness (H) likelihood profile for group-preferred run.

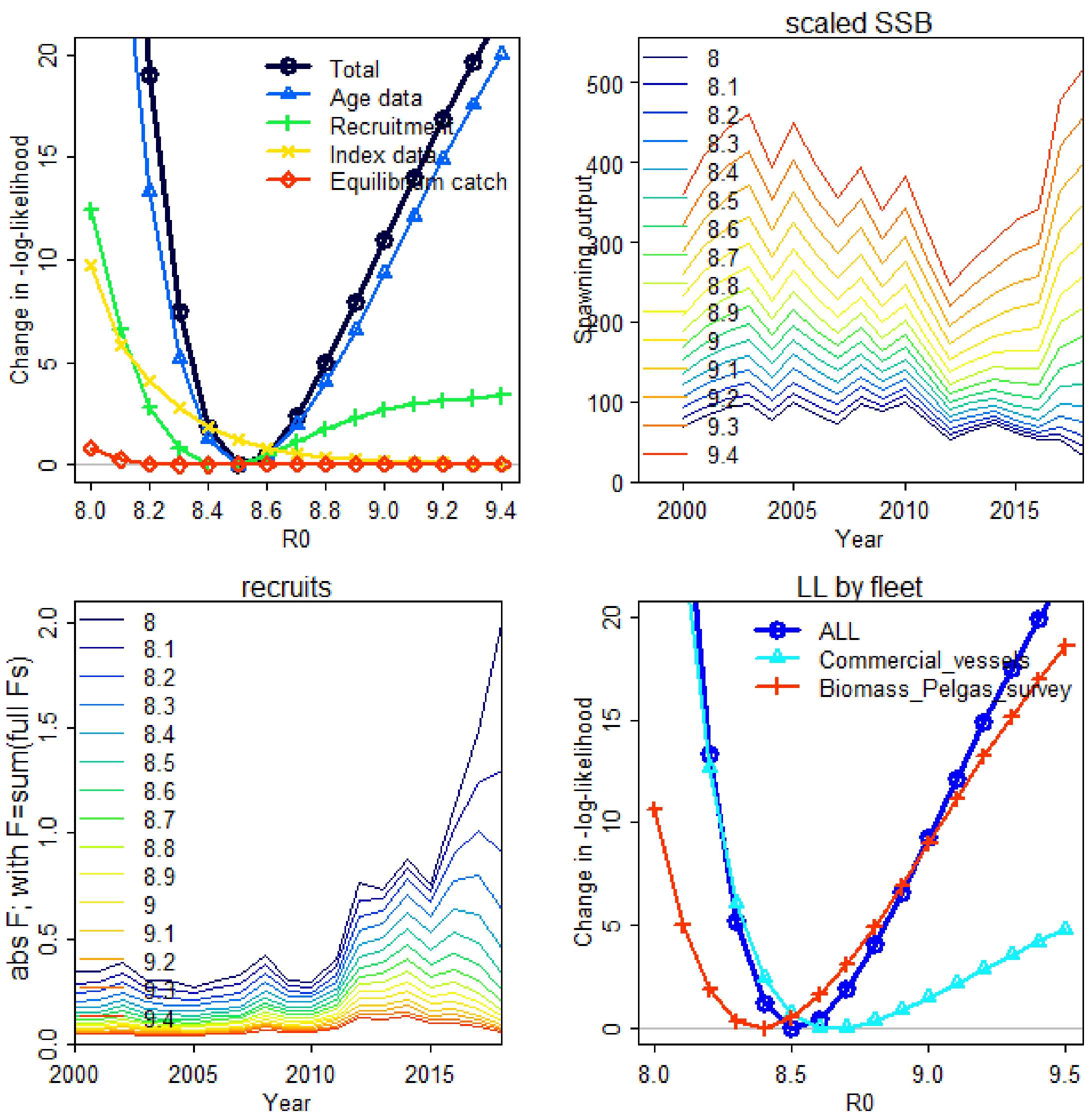

Figure 8. R_0 likelihood profile for group-preferred run.

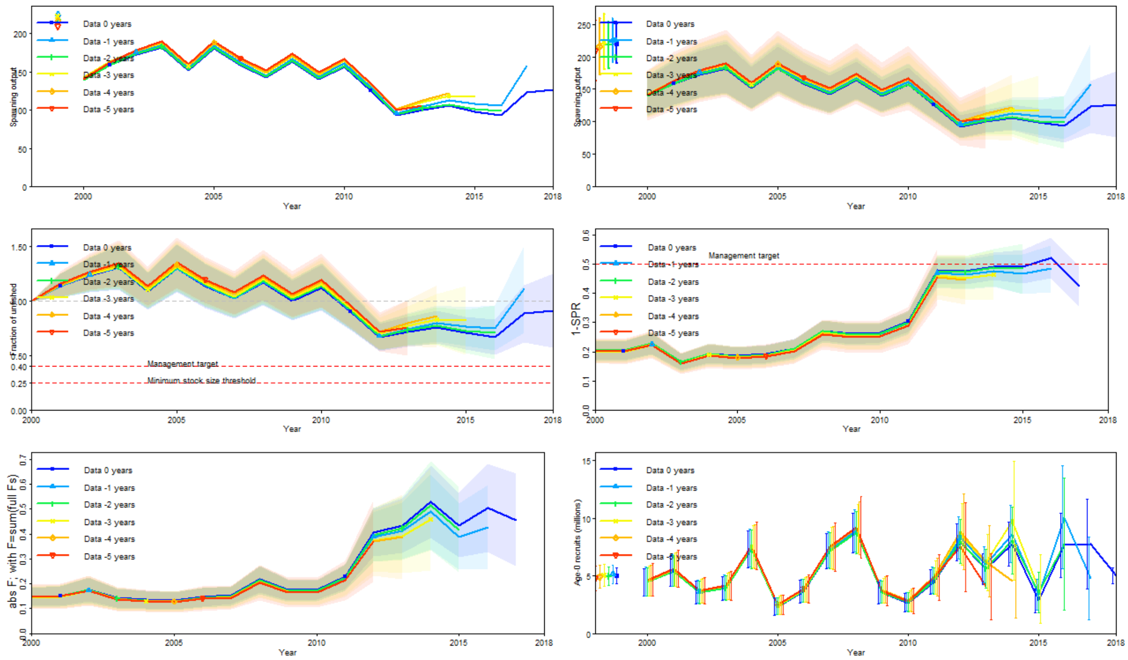

Figure 9. Retrospective patterns for the group-preferred run. From top to bottom and from left to right: spawning output, spawning output with confidence intervals, fraction of unfished, 1-SPR, apical F and age 0 recruits.

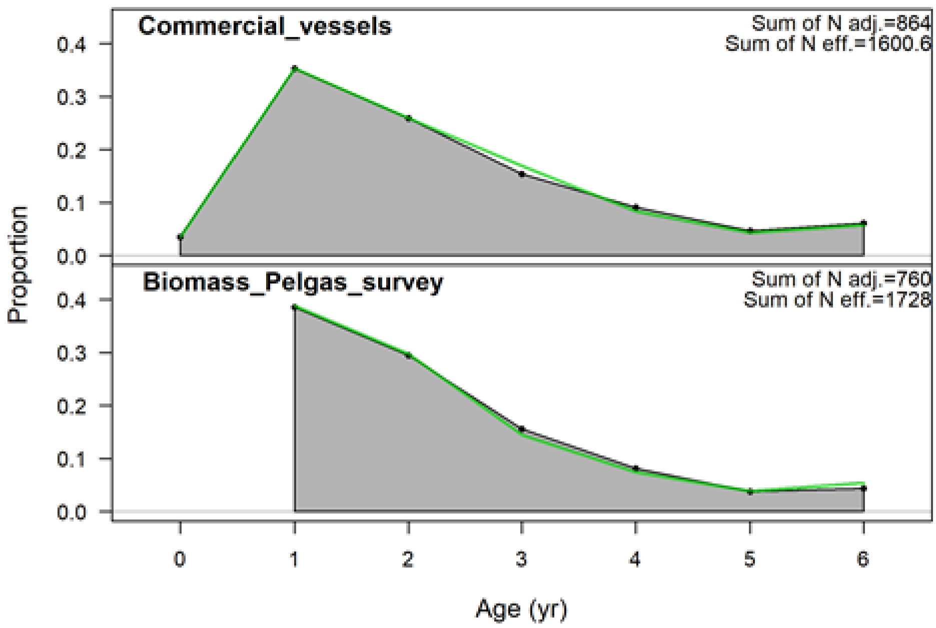

Figure 10. Age composition fits, aggregated across time by fleet.

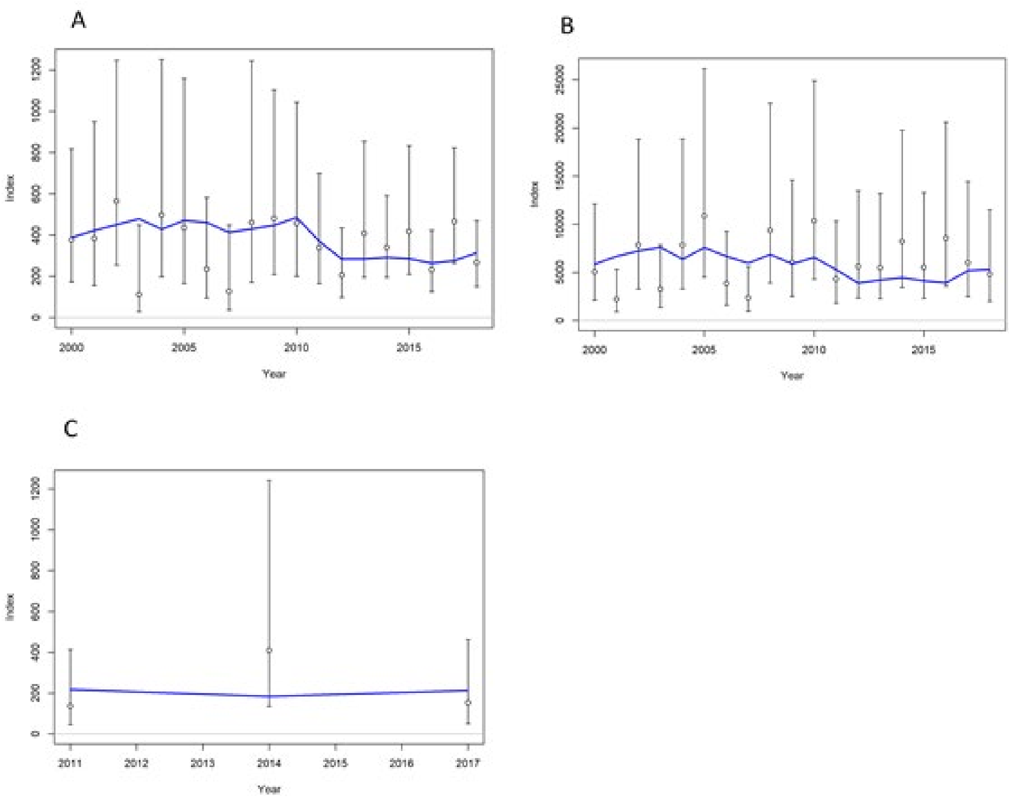

Figure 11. A. Fit to PELGAS survey. B. Fit to BIOMAN survey. C. Fit to DEPM survey.

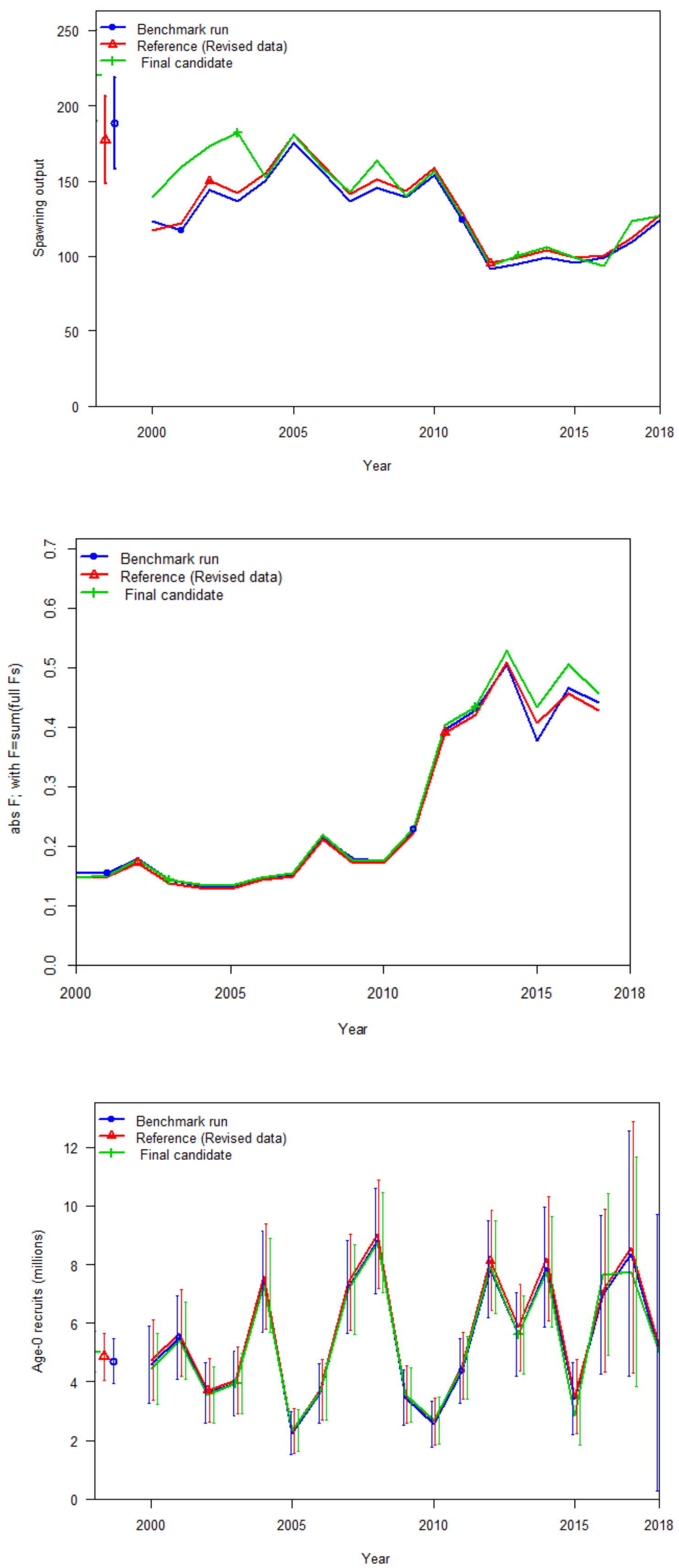

Figure 12. Summary outputs for of the group-preferred run (in green) in comparison with the previous model settings from the Benchmark 2017 (in blue), and the previous model settings with the French data revision (in red).

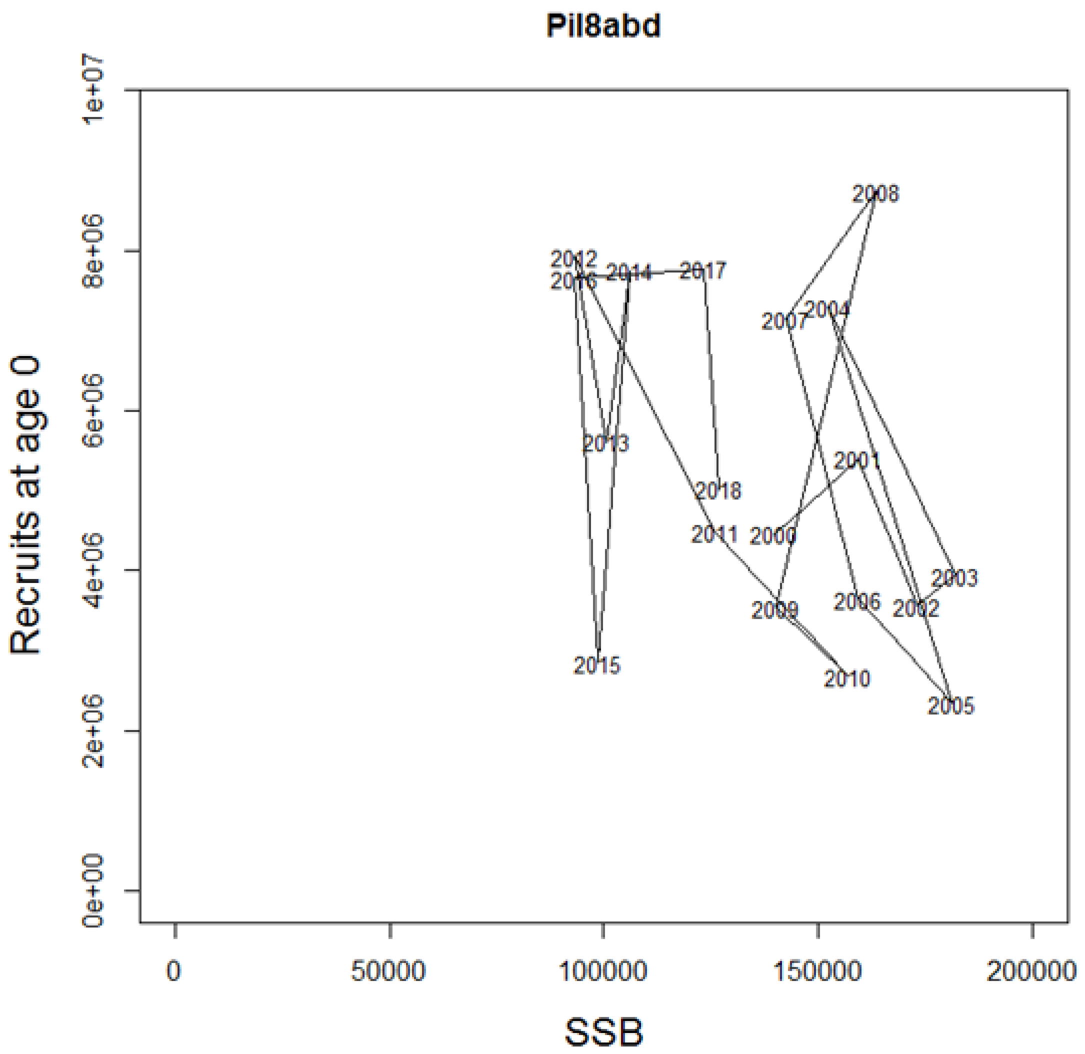

Figure 13. Stock–recruitment plot for sardine in 8abd.

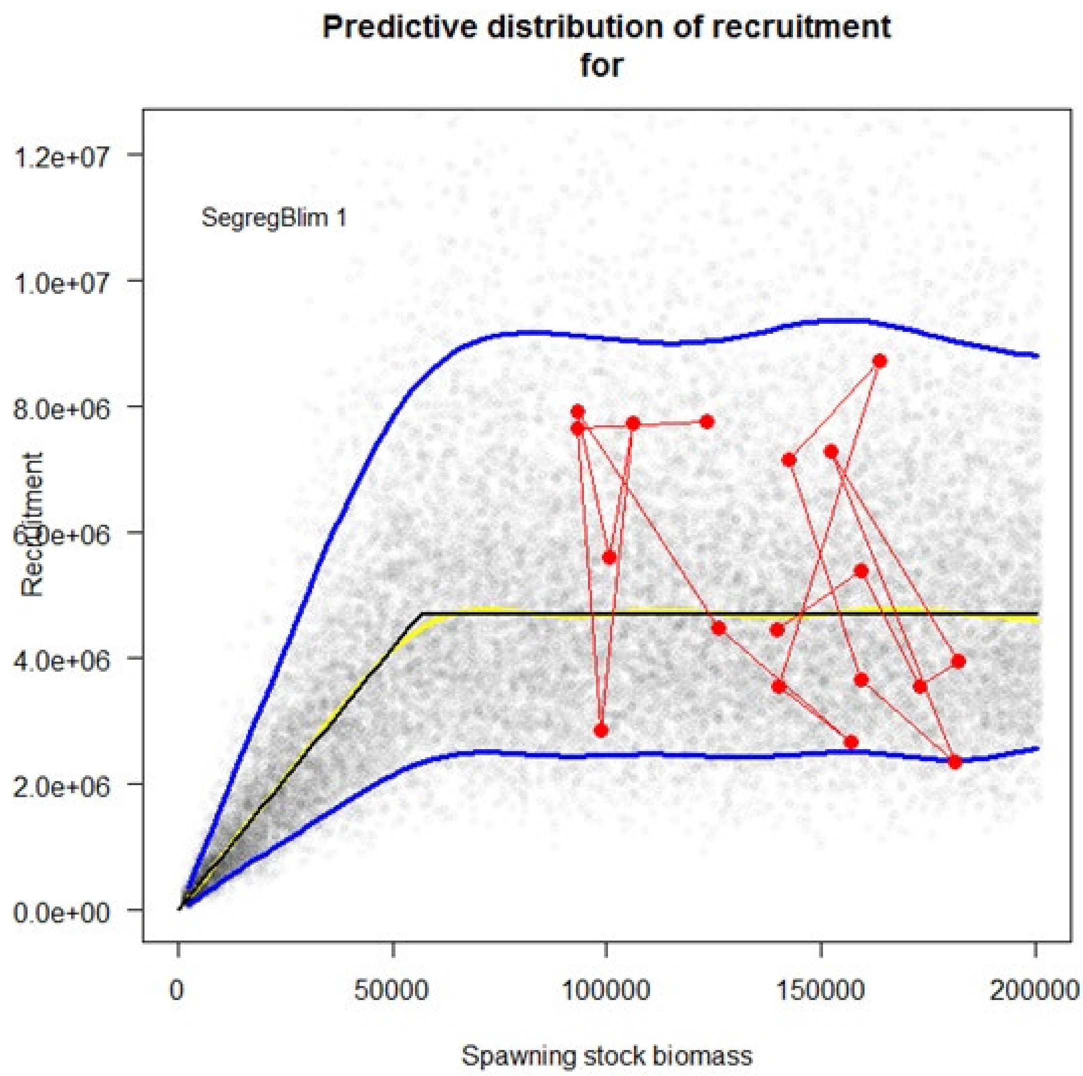

Figure 14. Segmented regression model for stock–recruitment (externally derived outside of SS) with the breakpoint set at B_{lim} (56 300 t) for sardine in 8abd.

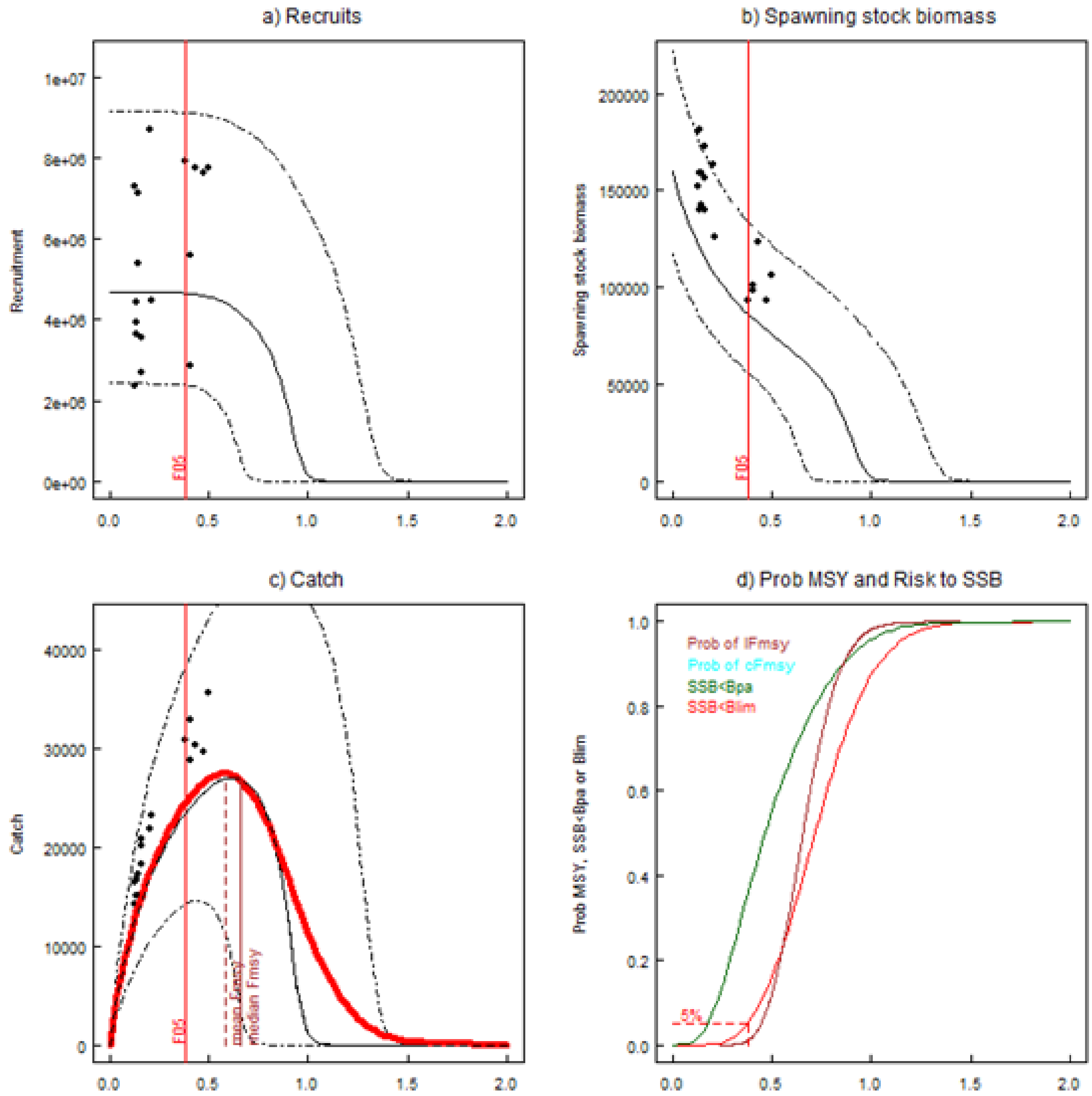

Figure 15. Eqsim summary plots without the ICES MSY advice rule for sardine in 8abd.

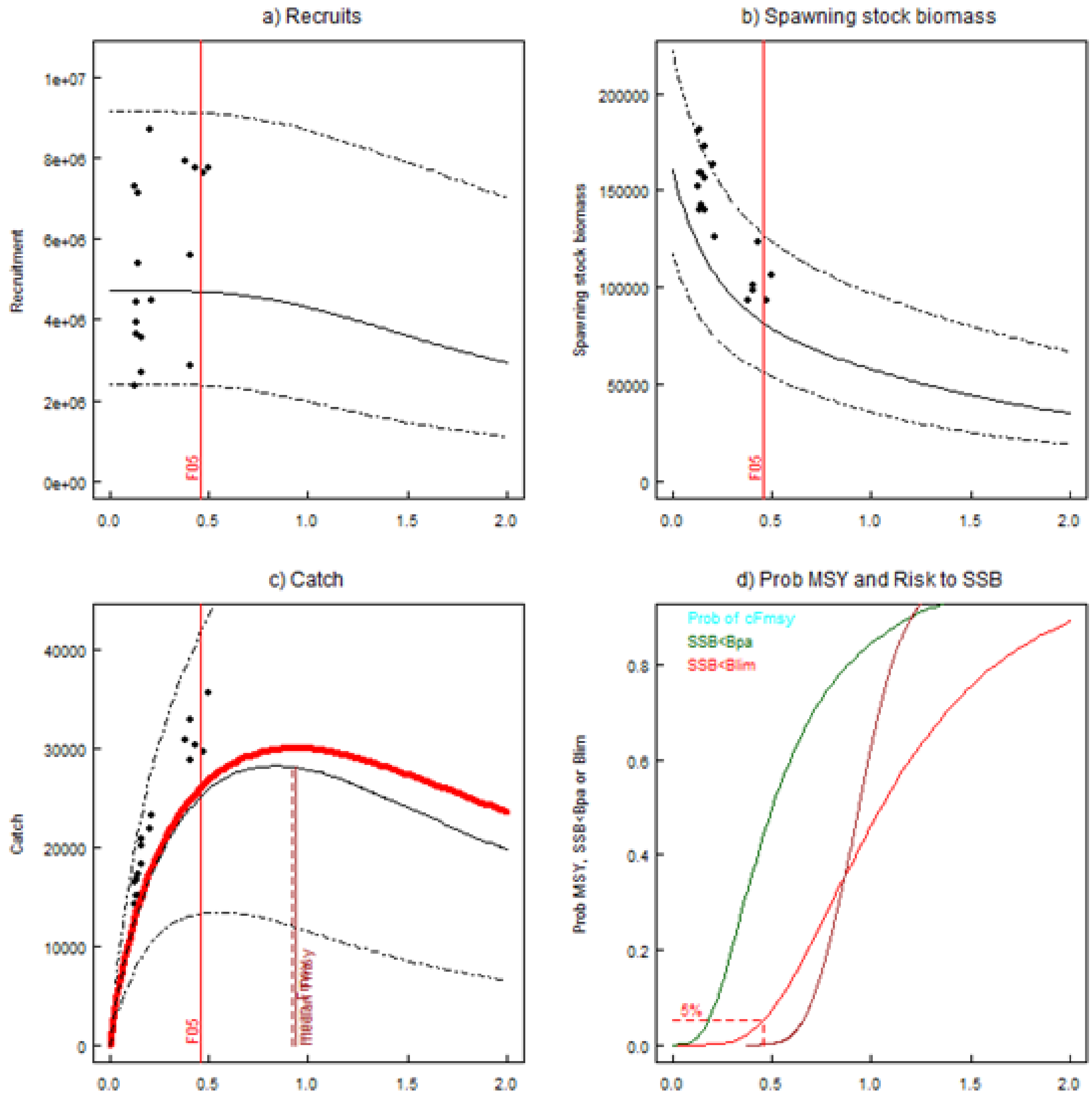

Figure 16. Eqsim summary plots with the ICES MSY advice rule for sardine in 8abd.

Annex 1: List of participants

Name	Institute	Country (of institute)	E-mail
Larry Alade, Reviewer	NOAA	US	Larry.alade@noaa.gov
Leire Citores	AZTI	Spain	lcitores@azti.es
Kiersten Curti, Reviewer	NOAA	US	Kiersten.curti@noaa.gov
Erwan Duhamel	Ifremer	France	Erwan.duhamel@ifremer.fr
Ruth Fernandez	ICES	Denmark	Ruth.fernandez@ices.dk
Leire Ibaibarriaga	AZTI	Spain	libaibarriaga@azti.es
Gaël Lavialle	Ifremer/Les Pêcheurs de Bretagne PO	France	gael.lavialle@pecheursdebre-tagne.eu
Lionel Pawlowski	Ifremer	France	Lionel.pawlowski@ifremer.fr
Alexandra Silva	IPMA	Portugal	asilva@ipma.pt
Andrés Uriarte	AZTI	Spain	auriarte@azti.es
Matthieu Veron	Ifremer	France	matthieu.veron@ifremer.fr
John Walter, Chair	NOAA	US	john.f.walter@noaa.gov

Annex 2: Review of the Inter-benchmark of Sardine in the Bay of Biscay

Larry A. Alade, NOAA, Northeast Fisheries Science Center, Woods Hole Laboratory, MA

Kiersten L. Curti, NOAA, Northeast Fisheries Science Center, Woods Hole Laboratory, MA

Background

This report provides an independent scientific peer review of the sardine in the Bay of Biscay (8.abd) presented at the Inter-benchmark review meeting conducted October 21–23, 2019.

The goal of the inter-benchmark was to address retrospective bias, data revisions, a scaling issue and update the biological reference points. The review group felt that all ToRs for the inter-benchmark were adequately addressed with substantial intervention to the existing model formulation. A range of model formulation changes were explored, including changes in fishery and survey selectivity, growth, and the assumed stock–recruitment relationship, as potential causes of the retrospective pattern in the model. Through a series of standard model diagnostics, there have been important advances in the provision of the fishery and survey data for Sardine in the Bay of Biscay. As such, the review group (RG) recommended that the preferred model formulation be accepted with the following changes:

i. Assume an asymptotic selectivity for both the surveys and the commercial fleet;
ii. Impose a Stock–recruitment relationship with a steepness close to 1 (h = 0.99) with a bias correction on recruitment deviations;
iii. Terminal year recruitment was not estimated in the model to minimize parameter uncertainty since recruitment in the terminal year tends to be poorly informed.

The preferred model formulation resulted in a substantial reduction of the retrospective pattern. Furthermore, indices included in the final model were treated as relative inputs, thereby removing the interpretation of catchability as an absolute value and eliminating any scaling issues. As such, the RG recommended that the assessment be upgraded from a category 2 to a category 1 assessment.

The meeting proceeded efficiently and was concluded on schedule. The assessment leads provided clear presentations for the various terms of reference, which were followed by constructive dialogue between reviewers, assessment scientists and other participants of the inter-benchmark. The terms of reference were clearly defined and were adequately addressed by the expert group. Additional recommendations are given in this review report and also available in the main report.

Summary review

The assessment is conducted in Stock synthesis. Prior to any interventions made during the workshop, the model formulation was based on a single commercial fleet configuration with catch-at-age and annual mean weight-at-age data. The assessment was tuned to three survey indices (PELGAS for biomass estimate, BIOMAN egg count and a triennial Daily egg production (DEPM survey)). Fecundity was derived from the maturity ogive and weight-at-age from the

PELGAS survey. Maturity assumed that 50% of individuals are mature at stage 2. The assessment assumed a rescaled age-specific and time-invariant natural mortality, derived from Gislason *et al.*, 2010. Natural mortality was rescaled to 0.8 to be consistent with total mortality derived from the PELGAS survey catch-at-age. PELGAS survey selectivity was set to 1 for ages 2 to 5 and estimated for ages 1 and 6+. DEPM and egg count survey selectivity parameters were set at 0 for age 0 and assumed asymptotic at 1 for ages 1 to 6+. For the commercial fleet, selectivity was set at 1 for ages 3 to 5 and estimated for ages 1, 2 and 6+.

As indicated earlier, one of the primary objectives of the inter-benchmark was to address the retrospective pattern in the assessment. The retrospective pattern has resulted in the overestimation of SSB and underestimation of fishing mortality and thus leads to a downward revision of absolute biomass. The retrospective pattern has resulted in the allocation of the assessment in category 2 because the entire time-series is re-scaled as years of data are removed. The implication of such allocation is that the estimated quantities from the assessment (F and SSB) were treated in relative terms.

A series of runs were brought forward by the EG during the inter-benchmark to examine the cause of the retrospective pattern in the assessment. Runs included revision to the French data, alternative catch data inputs, changes in selectivity time blocks, an alternative assumption about maturity at stage 2 and the application of Q1 weights-at-age from the commercial fleet (See Table 1). Retrospective patterns were estimated for all runs based on ICES guidelines using the Mohn's rho estimate. The revision of the French data led to lower values of Mohn's rho compared to the reference run. This became the basis for further explorations. It should be noted that the initial model explorations presented by the EG did not resolve the retrospective pattern. Therefore, several additional model explorations were considered during the inter-benchmark meeting to resolve the retrospective pattern. Additional explorations included starting the model time-series in 1989, assuming flat-top selectivity for the fishery and the survey, testing the influence of each survey by removing each index one at a time, and various treatments of growth (1. fixed mean weight-at-age, 2. assuming a von Bertalanffy growth and length–weight relationship, allowing the model to internally derive mean-weights-age, 3. allowing deviation in the K and deviations in growth among the cohorts to be estimated).

a) Evaluating Causes of Retrospective Patterns

The RG questioned the assumption of a dome-shape selectivity in the PELGAS survey, particularly for age 6+. Besides improved model diagnostics, there was no strong biological justification to support a lower selection of age 6+ in the survey. As such, a flat-top formulation was explored. This resulted in a lower AIC, reduced the retrospective pattern, and was kept as a candidate formulation for further exploration.

The effect of removing one survey at a time demonstrated the impact of each survey and resulted in a reduction of the retrospective pattern, with the largest reduction in Mohn's rho occurring with the exclusion of the BIOMAN survey. Although the survey exclusion exercise showed improvement in the retrospective diagnostics, none of the explorations fell within the acceptable guidelines of rho <0.2 to be further pursued. The RG also expressed concern about the double use of the egg data in the BIOMAN survey and the egg count.

The various growth configurations did not result in improved model fits nor improved retrospective diagnostics. Further, it was noted that the mean weights-at-age for age 0 were inputted as zeros in the model. To address RG concerns that this setting could be problematic, the zero weights-at-age-0 were replaced with mid-year weights. However, this exploration did not result in improved model diagnostics and therefore was not considered further.

The lack of a stock–recruitment relationship and its potential cause of the model's retrospective pattern was a major point of discussion among the WG and RG. Although there was no clear stock–recruitment relationship, it was hypothesized that imposing a weak relationship could potentially add structural integrity to the model by minimizing the magnitude of variability in the recruitment deviations. The WG further recommended that σ_R be estimated with a bias correction according to Methot and Taylor (2016). Recognizing the lack of S–R relationship, a Beverton–Holt formulation was considered with a steepness fixed at 0.9 and σ_R estimated with a bias correction. This resulted in a significant reduction in the retrospective pattern while reducing the AIC of the model. A follow-up run with steepness fixed at 0.99 was explored and thought to be consistent with the benchmark formulation that did not impose a stock–recruitment relationship. This further improved the model AIC and retrospective diagnostics.

Following a re-evaluation of the maturity ogive from the PELGAS survey, fecundity was recalculated assuming 100% of stage 2 fish were mature. This is a departure from the benchmark formulation that assumed 50% maturity at stage 2. The effect of the 100% maturity assumption was mainly on the age-1s, resulting in a higher proportion of age-1s mature in the early part of the time-series. A model run was considered to test the effect of this new maturity ogive assumption. Although, there was a modest increase in retrospective bias, the WG felt that this model run should be maintained among candidate runs because the new maturity assumption is consistent with observations from the spring PELGAS survey.

The initial treatment of the terminal year age composition was based on the age composition in year t-1. The RG deemed this approach to be problematic and unnecessary in the model settings. The issue with such setting is a mischaracterization of the age composition, contributing to lack of proper cohort tracking. As such, age composition in the terminal year was omitted.

b) Investigation of fleet selection patterns and noise in cohort tracking

Initial model explorations evaluated time blocks in selectivity and determined little to no improvement in model fit and AIC. Evaluation of age composition residuals did not indicate a systematic lack of fit that is indicative of a change in selectivity.

c) Update of historical French catch time-series

Historical French catches were revised for years 2013–2016 to reflect actual production catch data from the various French organizations. The revised catches are included in this updated assessment. In conjunction with the revisions to the French catch data, weights-at-age and numbers-at-age were also revised to ensure internal consistency between the total catch and French biological data. The WG fully supported the revised data and agreed that it should be the basis for the reference assessment model.

Two statistical rectangles from area 7e (Dournenez Bay) were explored for inclusion in the total catch estimates. The Dournenez Bay represents almost 20% of total catch from 2014–2017 from the coastal French purse seine fleet. In previous assessments, Dournenez Bay catches were included in the stock but based on very little evidence to support a separate stock area. Due to the magnitude of catch from both rectangles, difference in growth and poor sampling, the WG felt the need to address biological or spatial differences between these two areas. However, spatial differences based on growth information was equivocal and did not provide strong basis to include or exclude Dournenez Bay catches from the stock. Several model runs were explored to investigate the impact of Dournenez Bay catches in the model. There was very minimal improvement in retrospective bias in the model diagnostics when Dournenez Bay catches were omitted. Given the lack of strong evidence for the inclusion or exclusion of Dournenez Bay catches in the total stock catches, the WG recommended to maintain the Dournenez Bay catches. Genetic work currently being completed from monthly sampling of the two areas may better inform future stock delineation between the two areas.

d) Historical Spanish purse seine

The WG considered whether to include historical Spanish purse seine catches from 1971–1986. Due to data uncertainty and the likelihood that these data may have included catches from outside the Bay of Biscay, the WG agreed with the EG decision to exclude these catches from the time-series. However, the WG did consider an initial equilibrium catch (2000 mt) prior to 2000 in the model to allow for an initial estimation of fishing mortality rates to start the model. The assumed CV was increased from 0.05 to 0.10 to account for uncertainty in the catch during this period. This formulation did not substantially improve fits to the data and retrospective diagnostics did not improve over the reference model. As such, this model was not pursued among candidate model runs.

e) WG preferred model formulation

The following outlines the list of assumptions in the preferred model and also listed in the main report:

- Asymptotic selectivity (fix parameter 6 selectivity at 1) for both PELGAS and the fishery;
- Beverton–Holt stock–recruitment relationship, fixed steepness at 0.99, estimated sigmaR with Methot and Taylor bias correction;
- Exclusion of selectivity deviation parameterization;
- Retain start of the model time-series in 2000;
- Restrict estimation of terminal year recruitments as these are little informed by any data (this is commensurate with general practice of other modelling approaches to replace terminal year recruitments in VPAs or to employ some type of 'shrinkage' to the mean penalty);
- Do not carry over terminal year age composition;
- Full maturity at stage 2.

Despite the improvements made to achieve the WG's preferred model, estimated trends were similar to those of the original benchmark formulation. Three changes lead to a substantial improvement in the preferred model: 1) Implementation of an asymptotic selectivity in the fishery and the survey 2) Stable treatment of terminal year recruitment, and 3) The inclusion of a stock–recruit relationship but with a steepness (h =0.99) to account for the lack of S–R relationship in the model. These changes lead to a substantial reduction in the retrospective pattern (Mohn's rho less than 0.2) and strong model diagnostics in terms of model fit to the observations and AIC. With these model improvements, there are still a couple of limitations including the issue of population scaling (catchability estimates greater than one) and the inability to estimate steepness (based on the likelihood profile of steepness).

f) Biological reference points

The WG recommended an F_{MSY} proxy of $F_{35\%}$ SPR, based on considerations of life history and precautionary reference points (Myers *et al.*, 1999; Mace, 1994; Mace and Sissenwine, 1993), and proxies for F_{MSY} based on natural mortality rate (Zhou *et al.*, 2012; Francis, 2011). Recalculated reference points are updated in the report. The WG agrees that the reference points are in accordance with ICES guidelines for category 1 and 2 stocks, when the assessment covers a limited range of biomass and it is not informative about B_{lim} (i.e. does not show evidence of past impaired recruitment).

Recommendations

The following is a list of recommendations from the Inter-benchmark and also presented in the main report:

1. Evaluate methods for estimating natural mortality, particularly considering the work conducted for Bay of Biscay anchovy (Uriarte *et al.*, 2016). Consider moving to a single parameter for natural mortality (but with a Lorenzen-scaling) so that it can be profiled.
2. Explore incorporation of length and conditional age-at-length data into the model to allow estimation of growth internally as well as to estimate time-varying growth processes.
3. Consider explicit incorporation of environmental covariates on growth. This may not be critical for modelling historical growth but is highly influential when parameterizing the expected growth in the future, e.g. one needs to assume what growth (or weight-at-age) defines the 'benchmark' conditions.
4. Incorporate mean weight-at-age as 'data' in the data file, rather than as fixed, assumed-known values.
5. Given that 1) most of the catch is from ages 1 and 2, and 2) the likely two year time lag between a stock assessment model and the provision of catch advice, much of the catch that constitutes the TAC is a product of model projections. This makes the assumption of future recruitment highly influential in the TAC advice (Rice and Browman, 2014). As this fishery has high quality annual surveys that should serve as useful management procedures, the fishery could benefit from an MSE-tested management procedure.
6. Evaluate the historical time-series of Basque/Spanish catches. If even a quarter of these historical catches were from the Bay of Biscay, they would scale the overall population substantially higher.
7. Carry out further studies to clarify the stock identity of Douarnenez Bay catches (in Subarea 7), currently assumed to belong to the sardine stock in Subarea 8, which was shown to affect the scaling of the assessment. In particular, conduct biological sampling in Douarnenez Bay to obtain annual age–length keys for this area and to evaluate stock mixing and potential growth differences.
8. BIOMAN surveys should estimate maturity within the survey so that they are not reliant on PELGAS maturity.
9. Recommend exploring methods such as McAllister and Ianelli or other reweighting methods for composition data, which might give more weight to the composition data.

Annex 3: Stock Annex

The table below provides a link of the stock annex for sardine in divisions 8.a–b and 8.d (Bay of Biscay). 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 updated	Link
pil.27.8abd	Sardine (*Sardina pilchardus*) in divisions 8.a–b and 8.d (Bay of Biscay)	November 2019	Sardine 8abd