Serial No. N6841 NAFO SCS Doc. 18/18

SCIENTIFIC COUNCIL – April 2018

NAFO Scientific Council Flemish Cap (NAFO Div. 3M) Cod Stock Benchmark Assessment Meeting

Instituto Português do Mar e da Atmosfera (IPMA)
09–13 April 2018, Lisbon, Portugal

1. Opening of the Meeting 3
2. Appointment of rapporteur 3
3. Adoption of agenda 3
4. Introductory presentations 3
   a) Results of the Nov. 2017 Workshop 3
   b) Data preparation SC WebEx meeting. 4
   c) Assessment 2017 6
   d) Assessment Results 7
5. Assessment Input data. 12
   a) Consider the variability in the biological parameters (i.e. age at maturity, mean weights, etc.) observed in recent years, agree on an approach to be applied in stock assessment. 12
   b) Explore alternative values for natural mortality 12
   c) Aging and Age/Length Keys (ALKs): 15
6. Potential assessment models 15
   a) Bayesian XSA 16
   b) Bayesian SCAA 23
   c) SAM 35
   d) GADGET 37
   e) SC decisions 39
7. Comparison between models 40
8. Projections 43
9. Reference points 43
10. Discussion of possible operating model elements 45
11. Conclusions 45
12. Reviewer reports 45
13. Closing 45
14. References 46

Appendix I: Provisional Agenda 47

Appendix II: List of Documentation 49

Appendix III: Participant List 50

Appendix IV: Reviewers' Reports 51

Scientific Council Flemish Cap (NAFO Div. 3M) Cod Stock
Benchmark Assessment Meeting Participants

Back row: Paul Regular, Michael Palmer, Joanne Morgan, Floor Quirijns, Jim Ianelli, Tom Blasdale, Jose De Oliveira, António Avila de Melo, Sebastian Rodriguez-Alfaro

Front row: Dayna Bell MacCallum, Ricardo Alpoim, Brian Healey, Carmen Fernandez, Fernando González Costas, Diana González-Troncoso, Agurtzane Urtizberea, Alfonso Pérez Rodriguez

**NAFO Scientific Council Flemish Cap (NAFO Div. 3M) Cod Stock
Benchmark Assessment Meeting**

Lisbon, Portugal - 09-13 April 2018

Chair: Brian Healey Rapporteur: Tom Blasdale

1. Opening of the Meeting

The Scientific Council met in Lisbon, Portugal at the Instituto Português do Mar e da Atmosfera (IPMA) to discuss the Flemish Cap (NAFO Div. 3M) Cod Stock Benchmark Assessment. The meeting was attended by representatives from Canada, and the EU (Portugal & Spain). In addition, three expert external reviewers also participated in the meeting. An observer from the International Federation of Fisheries Associations was also present. The full participants list is included as Appendix III.

The meeting was chaired by the SC Chair, Brian Healey (CAN).

2. Appointment of rapporteur

The SC coordinator, Tom Blasdale, was appointed as rapporteur.

3. Adoption of agenda

The meeting agenda (see Appendix I) was adopted without revision

4. Introductory presentations

a) Results of the Nov. 2017 Workshop

Ricardo Alpoim (EU) presented the results of the 3M cod ageing workshop, conducted under an EU project.

Otolith comparative reading showed relatively high agreement for most ages but less so (ca 33%) at the oldest ages (older than 12 years). As a consequence, there are differences between applying keys from the fishery and the survey to the same length data producing different age distributions and identifying different dominant cohorts in the fishery.

Three laboratories are involved in cod age determination. During 2016 and 2017, one set of cod otoliths was exchanged and read to compare age reading results. The exchange collection consisted on 95 broken otoliths from three different sources: Portuguese and Spanish commercial fleet and the Flemish Cap survey. The sampling dates differed among sets, which is an important factor to consider in the interpretation of the results. Otoliths were selected, in order to have a complete range of lengths, from the 2014 sampling.

A total of six readers from the three laboratories (two readers by laboratory) participated in the exchange. Age estimates varied between 0 and 12 years. The percent agreement (PA) values ranged between 0-100% (average 76.7, which are considered pretty good agreement. The results were virtually the same when modal age was estimated using the three experts (75.7 %, with a corresponding CV of 7.5%). In total 55 out 95 otoliths yielded more than 80% of agreement (with both methods). Only 8 otoliths produced an agreement smaller than 50%. The age-bias plots did not show any particular bias. The CV by modal age did not show any particular trend and was lower than 10% for most of the ages, being only high at age 1 (26%). The distribution of the age reading errors showed the absence of bias among readers.

The workshop showed significant progress in the correspondence between readers compared with previous results. Also there was a clear improvement in agreement between the exchange exercise and the readings during the workshop after discussions, especially when using only clear-pattern otoliths and comparing the most experienced readers. Good quality otoliths with good readability showed only small differences among readers and demonstrated that in general there are no differences in age reading criteria among readers. This fact highlighted the importance of using a confidence index for each reading. However, a confidence index for each otolith read during the exchange exercise was not reported, and the workshop encourages continuing the implementation of QA/QC in the different laboratories involved in age reading of cod.

It would be necessary to explore if otoliths of "low" quality are not related to a specific cohort. An otolith exchange based on cohorts more than in years should be done. Some cohorts seem, due to temporary phenomena of a specific year, to create a false first ring. There would be cohorts that are read with one year difference (such as 2009-2010 year classes) between the readers of FC survey and the commercial one, but there would be other cohorts that this problem does not exist.

Nevertheless, some differences for certain criteria were also detected, especially in the position of the first (few) annuli and the identification of several checks, especially during the first 3-4 years, and the interpretation of the transition zones.

Ageing Workshop Recommendations for future work:

-Implementation of quality assessment/quality control (QA/QC) in the laboratories involved in age reading of cod.

-Explore if otoliths of "low" quality are not related to a specific cohort. An otolith exchange based on cohorts more than in years should be done.

-An otolith exchange be made with otoliths from 2009 cohort given by both ALKs to find out what happened and why they were read differently.

-Store a reference otolith collection for future

SC encouraged continued work on the ageing workshop recommendations.

Decision on the ALKs to use in the assessment:

The problems were well identified and seem to be specific to certain years. The workshop recommended to continue applying the survey ALKs to the commercial lengths until the recommended analyzes are completed. SC ***recommended*** to use the survey ALK in the next assessment in June 2018.

b) Data preparation SC WebEx meeting.

The following recommendations were made by a data preparation SC meeting held on 13 March 2018 (SCS Doc. 18/04):

Catch

i) Total Catch

It was agreed that the best time period to run the assessment is 1988-2016 due to problems with the quality of the input data before 1988. It was also agreed that for some of the assessment models that will be presented during the Benchmark, at least for the base case both time series (from 1972 and from 1988) will be used in order to analyze the sensitivity of the results to change the time series.

It was agreed that the best way to set the SC Cod 3M total catches and the catch at age abundance for 2011-2012 is to take as the SC approved catches for these years the median of the prior for these years catches used in the last approved assessment. The reason to use the median of the prior is that that median was made with the best catch information available.

ii) Catch at age

For the benchmark, it is not necessary to have a complete catch at age abundance matrix at this point. Solutions to problems of missing data in this matrix, where they occur, will be tackled during the process of developing the models. For the model that needs the catch at age abundance for the whole period the numbers at age estimated with the method proposed in this document will be used.

Weight at age

i) Catch weight at age

For the years 1994, 1995 and 1997 it was decided that a common method would be used to estimate the mean weights at age in the catches and the catch at age abundance. The method was to use mean weights at age of the years before and after of these years and estimate the SOP with the old abundance at age. If the SOP is more than 1.10 or less than 0.90, the old numbers will be raised to the approved total catches for these years. The reason for applying this method was that big year effects can be observed in some of these years in the old catch at age in the stock figures. For the years 1994 and 1995, the weight at age was calculated as the mean of the years 1993 and 2006 and for 1997, the mean of the years 1996-1998. For these three years the old abundance figures were raised to the total catches because the new SOP estimated with this method has a difference with the total catches of more than 10%.

For 2014, it was decided to maintain the old mean weights at age and numbers estimated from the Portuguese commercial information, that give a difference between the SOP and the total catches of 1.09.

In the period 2013-2016, it was agreed to update the abundance and the mean weights at age with a modification of the UK samples and the inclusion of new available Norway samples.

ii) Stock weight at age

The corrected transform values for the stock mean weights at age and for survey abundance at age from 1988 to 2004 will be used.

Plus Group

The default approach will be to use the current 8+. The study of an older plus group is considered a secondary problem and will be carried out if there is time.

Proportion Mature at age

It was agreed in the March SC WebEx meeting to use the current FC survey maturity ogives during the Benchmark.

Further discussion on the ogives estimation method occurred during the benchmark meeting. The Flemish Cap (FC) survey is the source of the 3M cod maturity data for SSB estimation. The FC survey takes place when females are resting and discrimination between spawning active and inactive females is still possible by testing the presence of postovulatory follicles, which remain several months after spawning. However, the best moment for discriminate active spawning females is during the spawning season. For 2012 and 2014, the maturity ogives calculated during the spawning season and when females are resting were compared (SCR 18/31). Lengths in the spring fishery were transformed to make them comparable to the July survey using the von Bertalanffy growth functions (specific to the year) assuming either constant growth or seasonal grow over the six months.

Preliminary conclusions:

- Assuming constant growth through the year, the differences found, although significant, are small, and could be due to the uncertainty of other factors such as the determination of age. Differences found in maturity at length could be smoothed in the maturity at age.
- Assuming a seasonal growth pattern for cod in 3M, the reproductive state in which maturity is estimated has no effect on the ogives estimation.
- Based on the small differences observed and, until more conclusive studies were available, it is recommended to continue employing in the assessment the maturity ogives produced with the FC survey information.

SC **recommended** *to continue the present comparison study using age based ogives (possibly based on the same otolith reader).*

Natural Mortality

During the Webex meeting, it was not agreed what would be the final method to estimate M, but three options were proposed for consideration during the Benchmark: Constant for all ages and years, the median vector estimated of all Size-dependent methods varying by age and constant in time and the age/year matrix of M estimated in the update GADGET model.

Input Data agreed during March 2018 webex to be used in the Benchmark Assessment:

-*Period of assessment*: 1988-2016

-*Total catch*:

-SC estimates from 1988

-No SC estimates for 2011 and 2012. Estimated by the model.

-SC Decision: to take the prior median as SC estimated total catches for those years (approximately 12 800 tons)

-*Maturity at age*

-From the Flemish Cap survey

-SC decision: to continue with the survey maturity ogives

-*Survey data*

-Canadian: Not used (as survey pre-dates 1988)

-Problems with the sum of products in years before 2004

-SC Decision: to use corrected values from the calibration for those years

-*Plus group composition*

-Current plus group: 8+

-Last years: Number of individuals in 8+ has increased

-SC Decision: default approach still is 8+

c) Assessment 2017

Assessment methodology

The Bayesian XSA model was approved in 2009 (SCR 09/34). The assessment approved in 2015 (SCR 14/33) was updated with 2015 and 2016 data. For years with catch-at-age data, it works as an XSA, assuming priors over the survivors. For years without catch-at-age data, a prior is set over the F, incorporating total catch to the model. For 2011 and 2012 two priors over total catch with a median value of approximately 12 800 tons (9 905-16 630) were fitted.

- **Abundance at age:** 1972-2016 (45 years)
  - Years without catch-at-age: 2002-2005
  - For 2011: $\mathrm{TotalCatch(2011)} \sim \mathrm{LN(median} = 9.46, \mathrm{sd} = 0.1313)$
  - For 2012: $\mathrm{TotalCatch(2012)} \sim \mathrm{LN(median} = 9.46, \mathrm{sd} = 0.1313)$
- **Tuning:** Canadian survey, 1978-1985

  EU survey, 1988-2016
- **Ages:** 1-8+
- **Catchability:** dependent on stock size for ages 1 and 2
- **Priors:** Setting as last year assessments
- **M:** With uncertainty via a prior

Table 1. Priors used in the Bayesian XSA assessment.

Input data	Prior Model	Prior Parameters
Total Catch 2011-2012	$LN\left(median, sd\right)$	Median=9.46, sd=0.1313
Survivors(2016,a), a=1-6 Survivors(y,7), y=1972-2016	$LN\left(median = medrec \times e^{-medM - \sum_{age=1}^{a} medFsurv(age)}, cv = cvsurv\right)$	medrec=15000 medFsurv(1,...,7)={0.0001, 0.1, 0.5, 0.7, 0.7, 0.7, 0.7} cvsurv=1
F(y,a), a=1-7, y=2002-2005	$LN\left(median = medF(a),\ cv = cvF\right)$	medF(1,...,7)=c(0.0001, 0.005, 0.01, 0.01, 0.01, 0.005, 0.005) cvsurv=0.7
Total Catch 2002-2005	$LN\left(median = CW_{\text{mod}}(y),\ cv = cvCW\right)$	CW_{mod} is arised from the Baranov equation cvCW=0.05
Survey Indices: Canada and EU (I)	$I(y) \sim LN\left(median = \mu(y,a),\ cv = \sqrt{e^{\frac{1}{\psi(a)}} - 1}\right)$ $\mu(y,a) = q(a)\left(N(y,a)\dfrac{e^{-\alpha Z(y,a)} - e^{-\beta Z(y,a)}}{(\beta-\alpha)Z(y,a)}\right)^{\gamma(a)}$ $\gamma(a)\begin{cases} \sim N(\text{mean}=1, \text{variance}=0.25),\ if\ a=1,2 \\ =1,\ if\ a \geq 3 \end{cases}$ $\log(q(a)) \sim N(\text{mean}=0, \text{variance}=5)$ $\psi(a) \sim gamma(shape=2, rate=0.07)$	I is the survey abundance index q is the survey catchability at age N is the commercial abundance index α = 0.5, β = 0.58 for EU survey (survey made in July), and α = 0.08, β = 0.17 for Canadian survey (made in January-February) Z is the total mortality
M	$M \sim LN(\text{median}, cv)$	Median=0.218, cv=0.3

d) Assessment Results

Some concerns about the Bayesian model used in the assessment were raised by STACFIS during its meeting in June 2017. The appropriateness of the priors used in the model, unchanged since 2008, was discussed. The robustness of the model with regards of changing the priors over the survivors was studied during the STACFIS meeting, but a deeper review is needed before changing model settings. STACFIS approved the assessment to perform the projections despite the issues encountered taking into account that the results of the assessment are in line with the survey trends. The impact of this issue will be studied in the 2018 benchmark.

Total Biomass and Abundance: Estimated total biomass and abundance showed an increasing trend since 2006 until 2012, reaching a biomass level similar to the pre-moratorium period. Since then a decreasing trend can be observed, with the greater decrease observed in abundance. The biomass value is around the level of the early 1990´s, while the abundance is at the level of the recovery of the stock in 2009 (Fig. 1). The posterior median of M estimated by the model was 0.19.

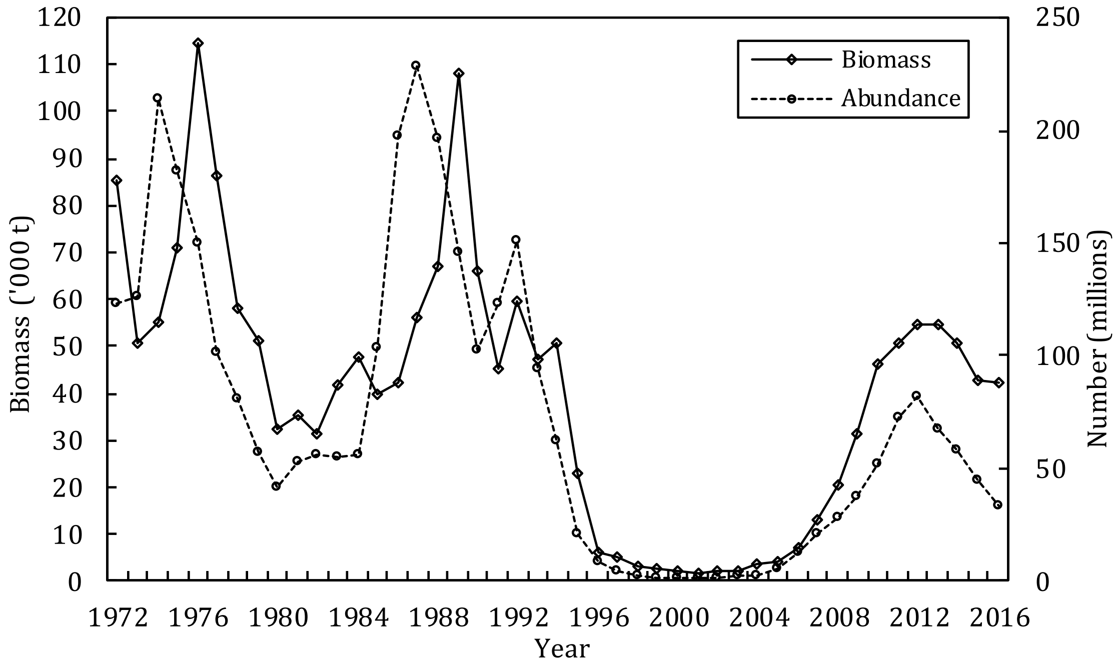


Fig. 1 2017 assessment: Biomass and Abundance estimates.

Spawning stock biomass: Estimated median SSB (Fig. 2) increased since 2005 to the highest value of the time series in 2013. This increase is due to several abundant year classes and their early maturity. Since then it has declined but with a very low probability (<1%) of being below B_{lim} (14 000 t).

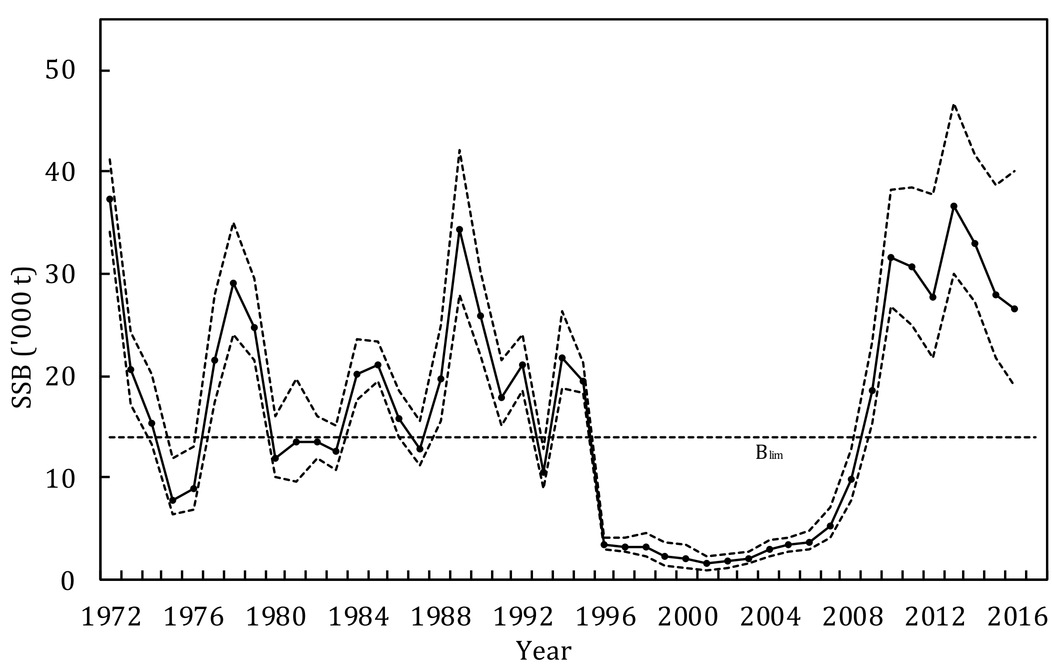


Fig. 2 2017 assessment: Median and 90% probability intervals SSB estimates. The horizontal dashed line is the B_{lim} level of 14 000 t.

Recruitment: After a series of recruitment failures between 1996 and 2004, values of recruitment at age 1 in 2005-2012 were higher, especially the 2011 and 2012 values. Since 2013 recruitment has decreased, reaching in 2016 low values as observed during the period 1996 to 2004 (Fig. 3).

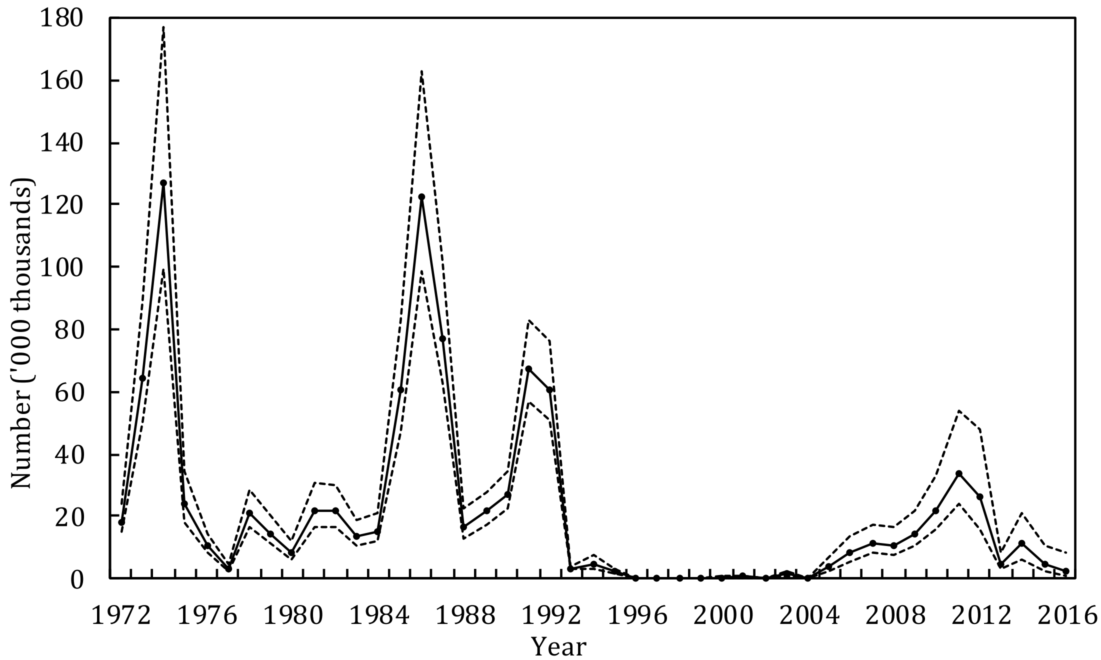


Fig. 3 2017 assessment: Recruitment (age 1) estimates and 90% probability.

Fishing mortality: F increased in 2010 with the opening of the fishery and it has remained stable since then at two times F_{lim} (mean $F_{2010\text{-}2016}$ =0.253) and below historical average (0.471) (Fig. 4).

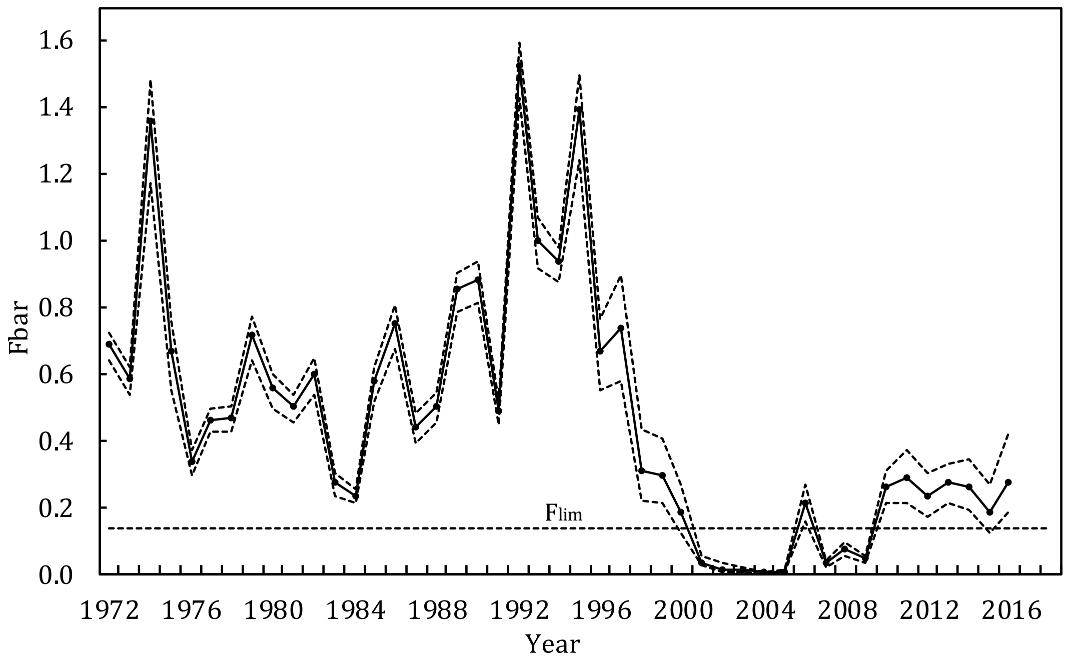


Fig. 4 2017 assessment: F_{bar} (ages 3-5) estimates and 90% probability intervals. The horizontal dashed line is the F_{lim} (0.139).

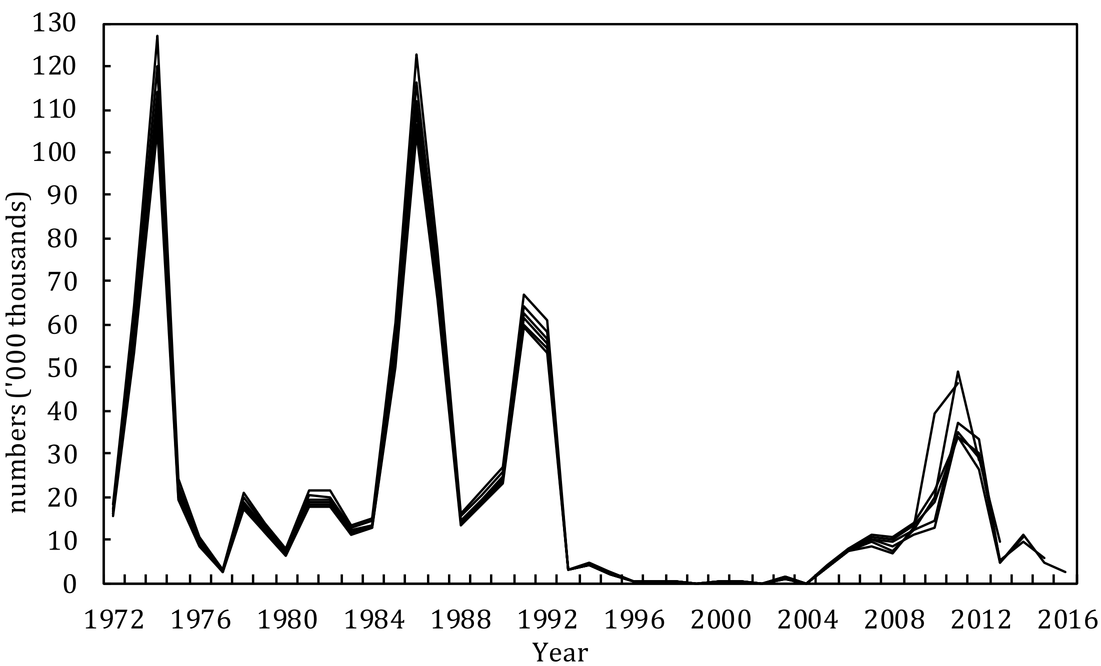


Fig. 5. 2017 assessment: Retrospective results for recruitment.

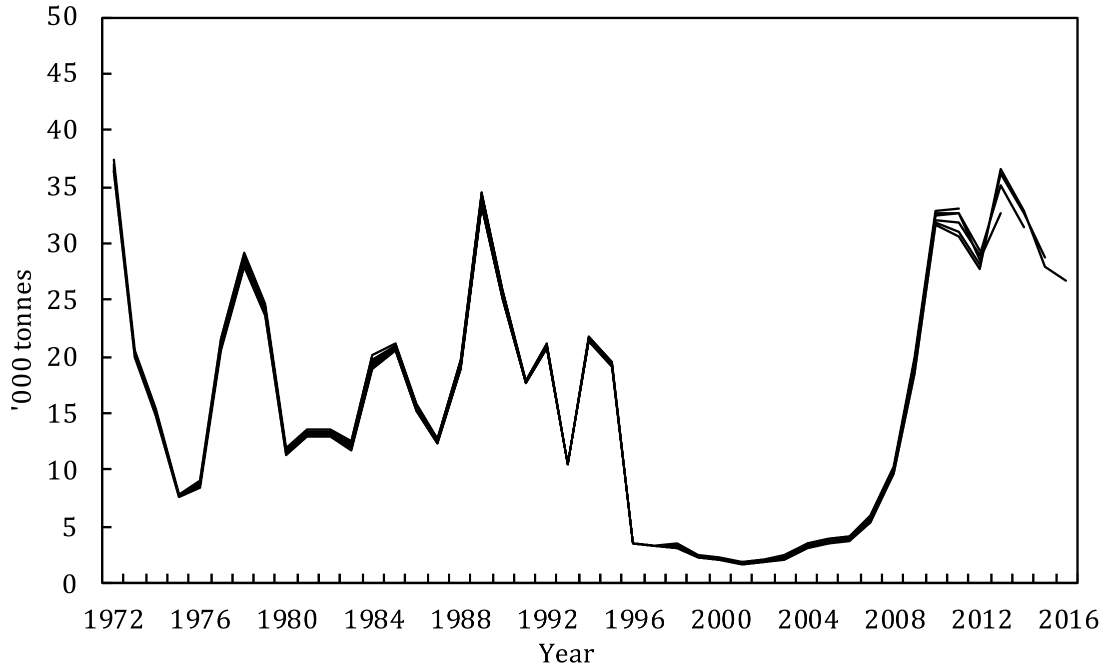


Fig. 6. 2017assessment: Retrospective results for SSB.

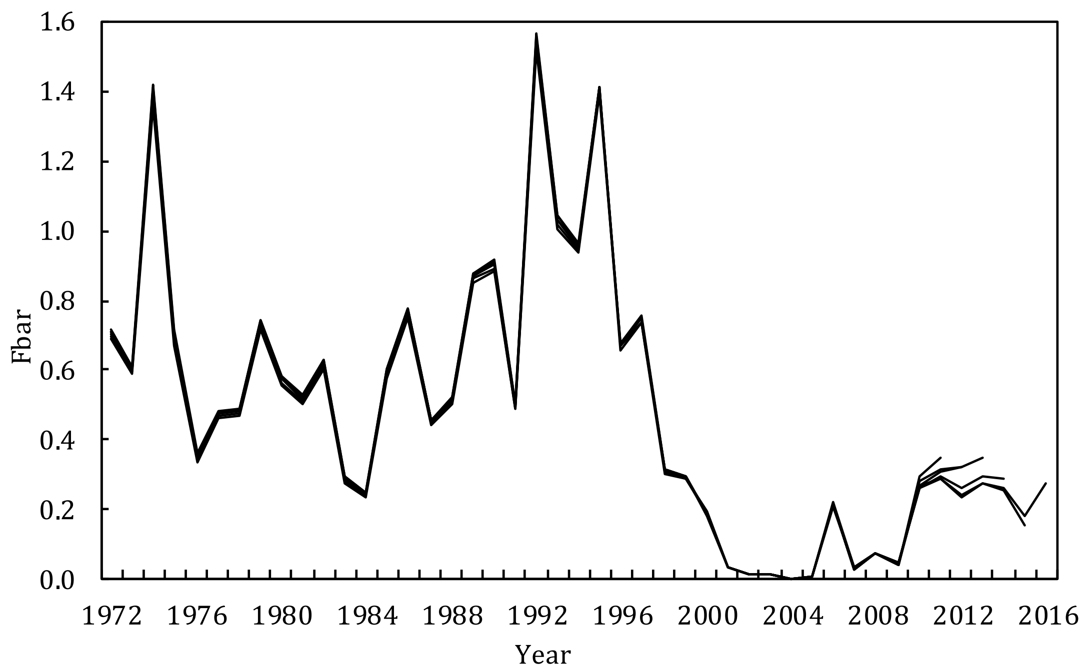


Fig. 7. 2017 assessment: Retrospective results for average fishing mortality.

5. Assessment Input data.

a) Consider the variability in the biological parameters (i.e. age at maturity, mean weights, etc.) observed in recent years, agree on an approach to be applied in stock assessment.

Variability in weight at age and maturity at age was discussed in the March 2017 3M cod workshop held in Vigo (NAFO SCS-Doc. 17/07); see also section 4.b) of this report.

b) Explore alternative values for natural mortality

In the previous 3M cod assessments, M was estimated (constant across years and ages) by the assessment model. At the SC WebEx meeting it was decided to investigate methods to estimate values of M independent of the assessment model. A SCR was presented during the benchmark with the results of the Natural Mortality (M) of NAFO cod Div. 3M estimated by different methods. The methods presented in this document to estimate M are age-independent and age-dependent methods that take into account the biological characteristics of this species. The results of M of the GADGET multispecific model, that takes into account the trophic relationships between different species, were also presented.

The results of three M scenarios have been analyzed: M constant for all ages and years, M variable by age and constant by year and M variable by age and year. The results show significant differences when assuming variability in M.

i) Estimation of M from biological models

Age-independent methods: During the 2017 workshop the values of M for the 3M cod estimated by different methods were presented (Table 2). These methods are based on fitting regressions through the relationship between estimated and measured M values and a range of life history parameters.

Table 2. The M values estimated with Age-independent methods for the 3M Cod.

Method	**M**
Pauly (1980) - Length Equation	0.113
Pauly (1980) - Weight Equation	0.136
Hoenig (1983) - Joint Equation	0.261
Hoenig (1983) - Fish Equation	0.246
Alverson and Carney (1975)	0.307
Roff (1984)	0.603

The life history parameters were estimated from the FC survey data for the period 2010-2015 (Table 3).

Table 3. The life history parameters used to estimate the M in the 3M Cod case. *Linf* and *Kl* are the Von Bertalanffy length equation. *Winf* and *Kw* are the Von Bertalanffy weight equation. *T*=mean temperature. *Tmax*=Maximum observed age and *tm*=age of first maturity.

Parameter	*Linf*	*Winf*	*Kl*	*Kw*	*T*	*tmax*	*tm*
Value	140	26000	0.121	0.16	3.5	17	3.9
Estimated	No	No	Yes	Yes			

ii) Size/Age-dependent models

These models calculate M based on the observation that, in general, natural mortality rate decreases with increasing size due to reduced predation. Whereas some methods use von Bertalanffy growth parameters to estimate size-dependent natural mortality, others calculate natural mortality as a function of weight. Both these weight-based methods for estimating M exhibit power relationships consistent with metabolic theory of ecology where biological rates, such as mortality, should scale with body mass to the power of -1/4. The results for M of the size-dependent method and the last approved assessment estimated M is shown in Table 4.

Table 4. The M values estimated with Size-dependent methods for the 3M Cod. This table also includes the mean natural mortality at age estimated in the GADGET model (Table 5) and the mean M at age estimated by all the methods.

Method	A1	A2	A3	A4	A5	A6	A7
M (Gislason)	2.19	0.99	0.56	0.41	0.32	0.25	0.2
M(Charnov)	2.33	1.11	0.66	0.48	0.39	0.31	0.25
M(Peterson and Wroblewski)	0.7	0.48	0.37	0.32	0.29	0.26	0.23
M(Lorenzen General)	0.94	0.61	0.45	0.38	0.33	0.29	0.26
M(Lorenzen Fish)	1.08	0.69	0.5	0.42	0.36	0.32	0.27
M (Chen&Wata)	2.06	0.67	0.39	0.26	0.19	0.15	0.16
M (Gadget) (mean (1988-2016))	0.57	0.48	0.39	0.36	0.35	0.35	0.35
2017 assessment	0.19	0.19	0.19	0.19	0.19	0.19	0.19
Mean All methods	1.26	0.65	0.44	0.35	0.3	0.27	0.24

Estimation of M via Gadget: The multispecies model GadCap (Pérez-Rodríguez *et al.* 2017) indicates that cannibalism in cod is a very important driver determining the survivorship or juvenile stages, especially when high recruitment events are coincident with high numbers of large individuals in the stock.

A direct output of GadCap model is the estimation of predation mortality (M_{pred}), which is estimated using diet composition, consumption estimate, predator-prey length relationship, number of predators and number of prey. However, the residual natural mortality (M_{resid}) is still a portion of remaining M that has to be provided to the model as fixed values. Estimating M internally during model optimization is extremely difficult, and often impossible, due to the interaction of M with the optimization of recruitment, growth and fishing catchability at age. For this reason, M_{resid} has to be estimated externally using an alternative option. Methods to estimate the M_{resid} based in the catch curve, the longevity and the loglikelihood methods were explored (SCR Doc. 18/025). Ultimately, a matrix with values of total M (M_{pred} and M_{resid}) was produced (table 5) with M_{resid} fixed at 0.35 for all ages and years.

Table 5 Total M (M_{resid} + M_{pred}) estimated with the model GadCap once M_{resid} is fixed as 0.35 for all ages and years.

age	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
1	0.766	1.125	0.91	0.455	0.479	0.406	0.41	0.471	0.392	0.373
2	0.397	0.842	0.656	0.41	0.374	0.389	0.395	0.419	0.385	0.362
3	0.358	0.388	0.581	0.367	0.355	0.355	0.36	0.357	0.362	0.358
4	0.352	0.356	0.368	0.361	0.352	0.351	0.351	0.351	0.351	0.353
5	0.35	0.351	0.353	0.351	0.351	0.35	0.35	0.35	0.35	0.35
6	0.35	0.35	0.351	0.35	0.35	0.35	0.35	0.35	0.35	0.35
7	NA	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
8	0.35	NA	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
9	0.35	0.35	NA	0.35	0.35	0.35	0.35	0.35	0.35	0.35
10	0.35	0.35	NA	NA	0.35	0.35	0.35	0.35	0.35	0.35
11	0.35	0.35	0.35	NA	NA	0.35	0.35	0.35	0.35	0.35
12	0.35	0.35	0.35	0.35	0.35	NA	0.35	0.35	0.35	0.35
age	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
1	0.362	0.367	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2	0.359	0.363	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
3	0.351	0.353	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
4	0.351	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
5	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
6	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
7	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
8	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
9	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
10	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
11	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
12	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
age	2008	2009	2010	2011	2012	2013	2014	2015	2016	
1	0.35	0.35	0.876	0.822	0.581	0.592	1.441	1.425	0.809	
2	0.35	0.35	0.692	0.683	0.622	0.656	0.693	0.894	0.789	
3	0.35	0.35	0.412	0.457	0.506	0.497	0.517	0.48	0.527	
4	0.35	0.35	0.365	0.37	0.392	0.403	0.384	0.415	0.392	
5	0.35	0.35	0.352	0.354	0.356	0.363	0.361	0.364	0.373	
6	0.35	0.35	0.35	0.351	0.352	0.353	0.353	0.356	0.356	
7	0.35	0.35	0.35	0.35	0.35	0.351	0.351	0.352	0.352	
8	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	
9	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	
10	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	
11	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	
12	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	

For ages older than 3 years, predation M appears to be unimportant. The high inter-annual variability in recruitment success and the relatively low number of strong trophic interactions suggest that variability in natural mortality due to predation is probably a frequent issue in the Flemish Cap at least for ages 1-3. Under these conditions, the use of a multispecies model as source of annual values of natural mortality by age, used as input for single species models could be an option.

iii) Other (including Bayesian estimates)

The option of allowing the Bayesian models to estimate M internally was also discussed and will be considered further under each of the Bayesian models.

A separable approach to the computation of M

An alternative approach to computation of M was proposed, in which natural mortalities over the period 2005 to 2017 are considered to vary simultaneously with time and age following a similar framework to that assumed in the computation of separable fishing mortalities and catch/stock projections under several average F options. Average natural mortality (ages 1-7) would be allowed to vary every two years (there is no evidence of inter annual changes on environmental conditions so dramatic to justify the need of M fit on an annual basis) while relative M (proportion of average M at each true age) is age dependent but is constant with time. According to the results of all size dependent methods, M is expected to be higher on the very young ages, basically justified by predation, including cannibalism, and gradually decline with individual growth. This variation of M with age is obviously also reflected in the relative M at age vector. For each year and age, M at age is the product of the relative M for that age and the average M for that year.

It was suggested that average M should be computed through a sequence of reruns of the Bayesian XSA model ending on alternate years from 2006 to 2017, all using the input data and assessment settings approved on the SC Cod Benchmark WebEx Preparatory Meeting March 2018 (SCS-Doc 18-04), and M constant at 0.2 for ages 1-8+ over the 1988-2004 interval. This set of runs would allow the step by step search for a best fit average M, from 2005-2006 till 2015-2017 (in order to include last assessment year). The approach is very similar to the one implemented on the 3M beaked redfish stock unit from 2011 assessment onwards as regards the search for variability on the average level of natural mortality over time (SCR Doc. 17-032REV2). Relative M at age could be obtained either from an average M at age vector from the results of size dependent methods, or from the M at age vector from the Gad Cap results (Pérez-Rodríguez *et al*, 2017).

SC Decisions:

- After analyzing the results, the SC considered that the best option for the assessment of this stock is to use a vector of M variable by age as the median of a prior distribution with a coefficient of variation of 15% within a Bayesian model. It was decided to use the mean of all Size/Age-dependent models presented in Table 4 as the prior median.
- The reasoning for this choice was that the results of a vector (M variable by age) and of a matrix of M (M variable by age and years) are quite similar and that the final matrix of M is quite sensitive to the estimate that is made of the residual M. The Benchmark decided to continue estimating M inside of the assessment Bayesian model by providing a prior with enough information that is reflected in the value of the low coefficient of variation (15%).

c) Aging and Age/Length Keys (ALKs):

See section 4.a

6. Potential assessment models

Four different models were presented during the benchmark: Bayesian XSA, Bayesian SCAA, SAM and GADGET. Several scenarios changing the inputs of each model were run.

a) Bayesian XSA

Some concerns about the Bayesian model used in the assessment were raised by STACFIS during its meeting in June 2017. The appropriateness of the priors used in the model, unchanged since 2008, was discussed. The robustness of the model with regards of changing the priors over the survivors was studied during the STACFIS meeting, showing that in some cases changing some of the parameters make a big difference in the results. A deeper review of these issues was asked to be done during the benchmark.

Ages in the Bayesian XSA model are from a=1 to A=8+ and years are from y=1 (1988) to Y=29 (2016). The cohorts are modelled backwards in time, starting from survivors of the last true age (age 7) in each year and survivors from each true age (1 to 7) in the last assessment year, taking into account the natural and fishing mortality. The model equations can be seen in González-Troncoso, 2017.

Different scenarios with this model changing some of the settings and/or input data were run. The scenarios run are listed in Table 6. The rationale of these parameters is:

-*medrec*: recruitment value used to set the prior median on the last year survivors.

-*qs*: catchabilities of the surveys by age. They can be all different or can be grouped by groups of ages.

-*psi.EU*: this parameter controls the variability (CV) of the observation equation for the survey abundance index at age. It can be estimated including a prior distribution, it can be different for each age or grouped by groups of ages, or it could be fixed.

-*cv(surv)*: this parameter controls the variability (CV) of the prior distribution on survivors.

-*M*: it can be one value, a vector or a matrix, and it can have a prior or be fixed.

-*adep*: this setting indicates the ages for which the survey catchability depends on population abundance

-*Zeros*: If the zeros in the catch at age and in the survey index are included or not.

-*Years*: years used in the input data of the assessment,

Table 6. Scenarios run for the Bayesian XSA model. Differences between successive model runs highlighted in red text.

Run	Base	medrec	Age groups survey catchability	CCV survey observation equation	CV prior on survivors	M	CV prior on M	*adep*	Treatment of zeros	Years	Retro
1	Approved	15000	All different	Prior	1	1 prior	0.3	1,2	0.0005	1988-2016	No
2	Approved	15000	All different	Prior	1	1 prior	0.3	1,2	0.0005	1972-2016	No
3	1	15000	All different	Prior	1	Vector		1,2	0.0005	1988-2016	No
4	1	15000	All different	Prior	1	Matrix		1,2	0.0005	1988-2016	No
5	1	15000	1,2,3,4+	Prior	10	8 priors	0.15	1	0.0005 in survey, 0 in catch	1988-2016	Yes
6	5	45000	1,2,3,4+	Fix (30%)	10	8 priors	0.15	1	Min for EU, 0 in catch	1988-2016	Yes

All the models reach the convergence after 50,000 iterations, from which 5000 are taken for the results.

The method followed to run the different scenarios was to start running the approved assessment but with the revised data. In the first run, the settings are those from the approved assessment, starting the assessment in 1988 instead of 1972 (and using the revised data). In order to check the influence of the revised data approved by the SC input data meeting, R2 is the same as the approved assessment, starting in 1972, and using the revised data. Then, based on the first run, R3 was run with M fixed as a vector, and R4 with M fixed as a matrix. In R5 we change the CV of the priors of the survivors and priors over M (one by age) were set. In R6 the median of the prior of the recruitment was changed, and the variability of the observation equation for the survey abundance index was fixed.

The recruitment value used to set the prior median for survivors was 15000 in the approved assessment. During the Benchmark, several SCAA runs (next Section) were conducted using a prior median for survivors of 45000. For comparison purposes, an XSA run (R6) using this value was conducted for comparison purposes.

With regards to **catchability** in the survey, we tried two different settings. In R1-R4 we have catchability different for all ages, but looking to the posterior distribution of the catchabilities of R1 (Figure 8), it can be observed that some of the ages have similar catchabilities. These results are robust with the EU survey and fishery information available. Based on the biological and survey gear information we have chosen to group the ages in catchability as follows: 1, 2, 3 and 4-8+. This is applied starting from R5.

The variability (CV) of the observation equations for the **EU survey indices** can be controlled by a prior or can be fixed. In R1-R5, priors over this CV were set, allowing different values of the CV for each age. But as we can see in Figure 9, the posterior estimates for the CV are too high for ages 1, 7 and 8+. Due to that, it was decided to fix this CV to be 30% in R6. A CV of 30% is reasonable for the survey abundance index of this stock taking into account their variability.

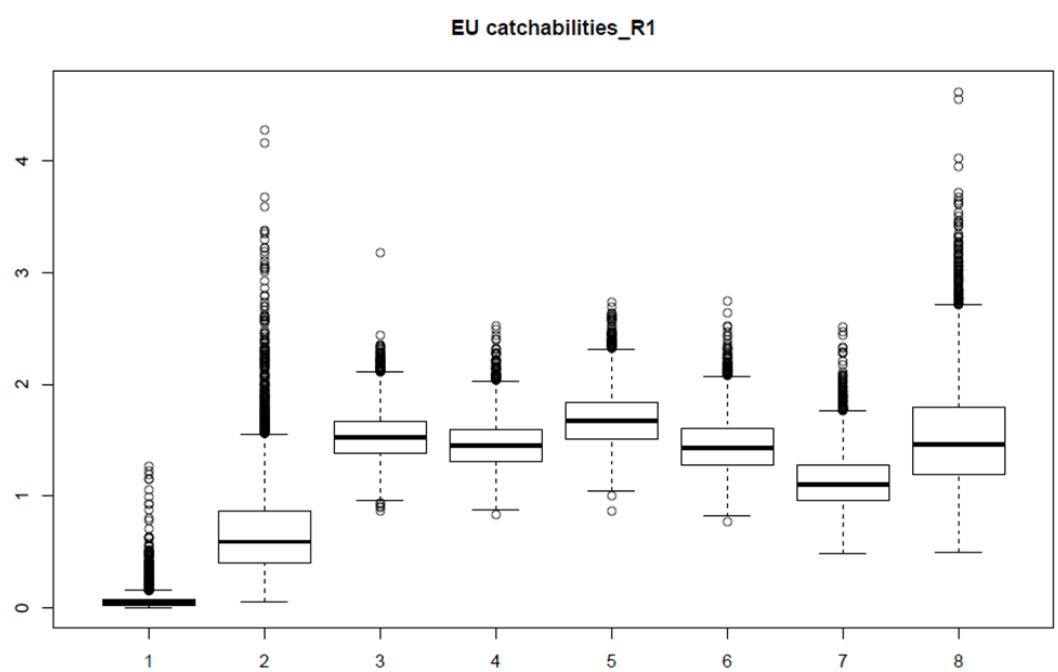


Fig. 8. Posterior EU catchabilities in R1

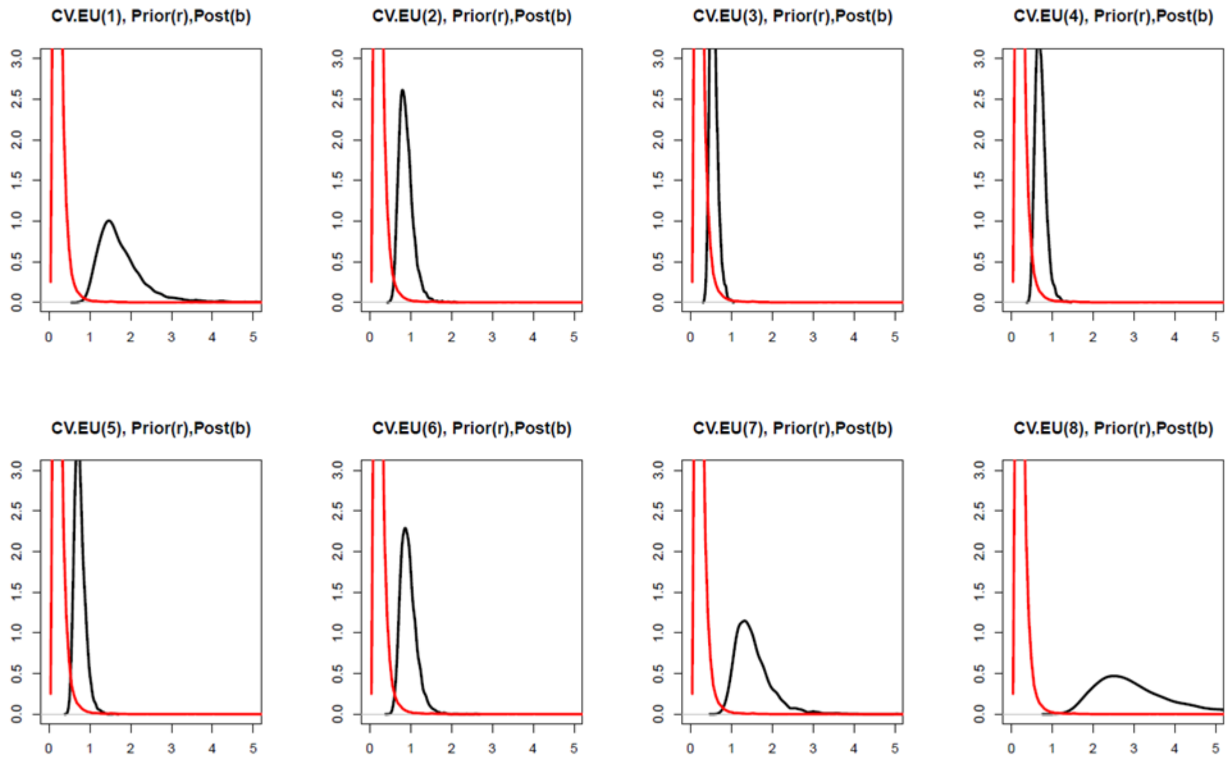


Fig. 9. Prior (red) and posterior (black) of the CV (%) of the abundance EU survey index in R1

In the case of the CV of the prior distribution on **survivors**, we start with a CV of 1, as in the last approved assessment, but also a value of 10 was tried in R5 and R6 in order to check the sensitivity of the model to this parameter to allow the survivors to have a higher variability if justified.

In the case of **M**, four different approaches were tried: to have a prior over the M (constant over years and ages), as in the last approved assessment (R1 and R2); to have a vector of Ms as input (constant through years but different by age) (R3); to have a matrix of Ms as input (varying over ages and years) (R4), and to have 8 priors, one for each age, with the median of the prior equal to the vector used as input in R3 (R5 and R6). The values used for the vector and the matrix are in Table 8 and comes from González-Costas and González-Troncoso (2018) and the results of the GADGET model (Table 2). The posterior median of the scenarios with prior (R1, R2, R5-R6) are presented in Table 7. It can be seen that, when we allow the M to be different between ages, the value for age 1 is much higher than for the rest of the ages. For ages 2 and 3 it is still high, and then it decreases.

Table 7. Results of the posterior median over M for R1, R2, R5 and R6.

	1	2	3	4	5	6	7	8
R1	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
R2	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
R5	1.24	0.62	0.42	0.33	0.28	0.27	0.26	0.24
R6	1.25	0.58	0.40	0.29	0.24	0.26	0.34	0.24

With regards to the **first age** in which the catchability is independent of population abundance (so, the exponent gama=1), we have tried two different approaches: catchabilities of ages 1 and 2 dependent on abundance (R1-R4) and only age 1 dependent on abundance (starting from R5). It seems more logical to use only age 1 as dependent on abundance, as when we use ages 1 and 2, the posterior median of the distribution of the exponent (gama) at age 2 is almost equal to 1.

There are some zeros in the observed catch-at-age in numbers and in the observed EU index. In the case of the catch-at-age the XSA model allows us to have 0, but not in the case of survey numbers at age. A first approach with a value of 0.0005 instead of 0 in both inputs was taken. Differences between the results are not evident.

So, we leave the 0 in the catch at age and two more attempts were made in the EU survey index: 0.0005 and the mean of the minimum value in that age for all the years. This last approach, tried in R6, seems to be the most reasonable.

Figures 10 to13 show the SSB, the R, the F_{bar} and the number at age for each of the scenarios and for the approved assessment, respectively.

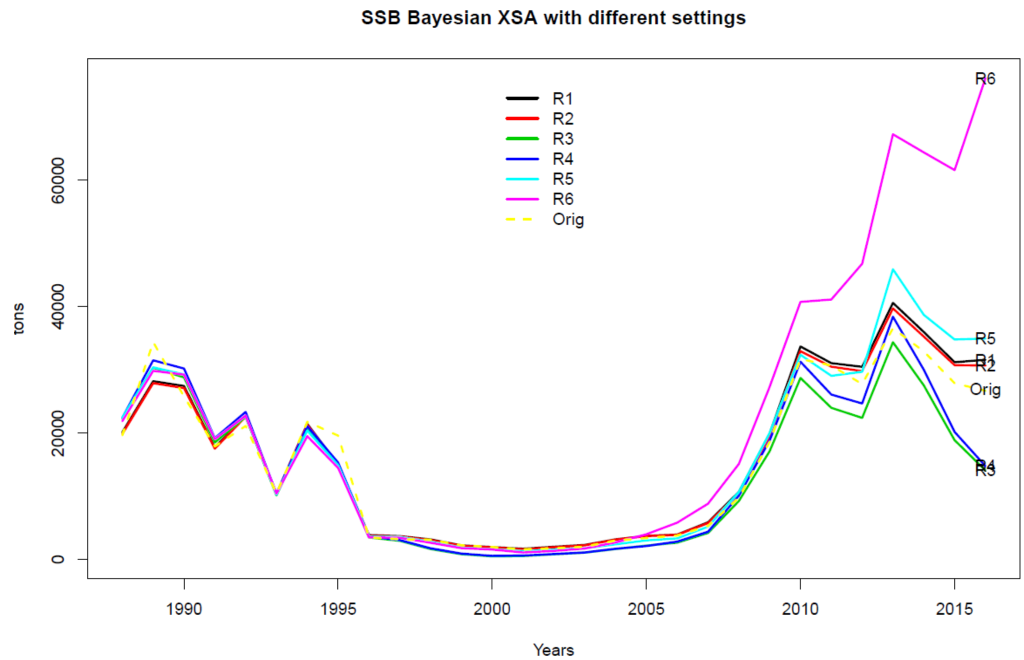


Fig. 10. Results of the posterior median SSB for the different runs of the Bayesian XSA (R1-R6) as well as the approved assessment (labelled as "Orig").

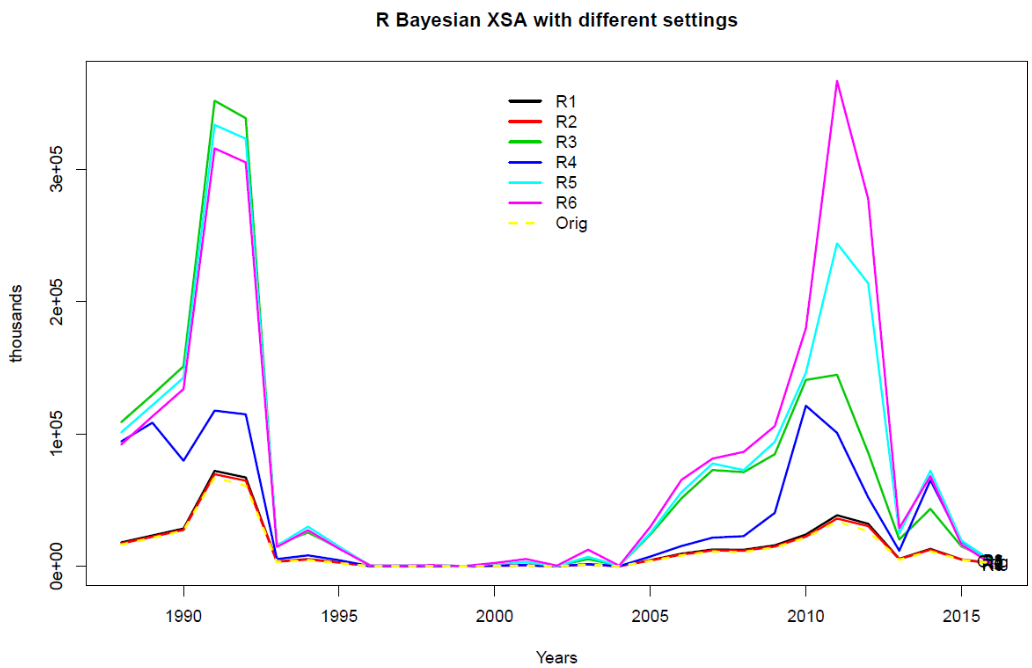


Fig.11. Results of the posterior median recruitment for the different runs of the Bayesian XSA (R1-R6) as well as the approved assessment (labelled as "Orig").

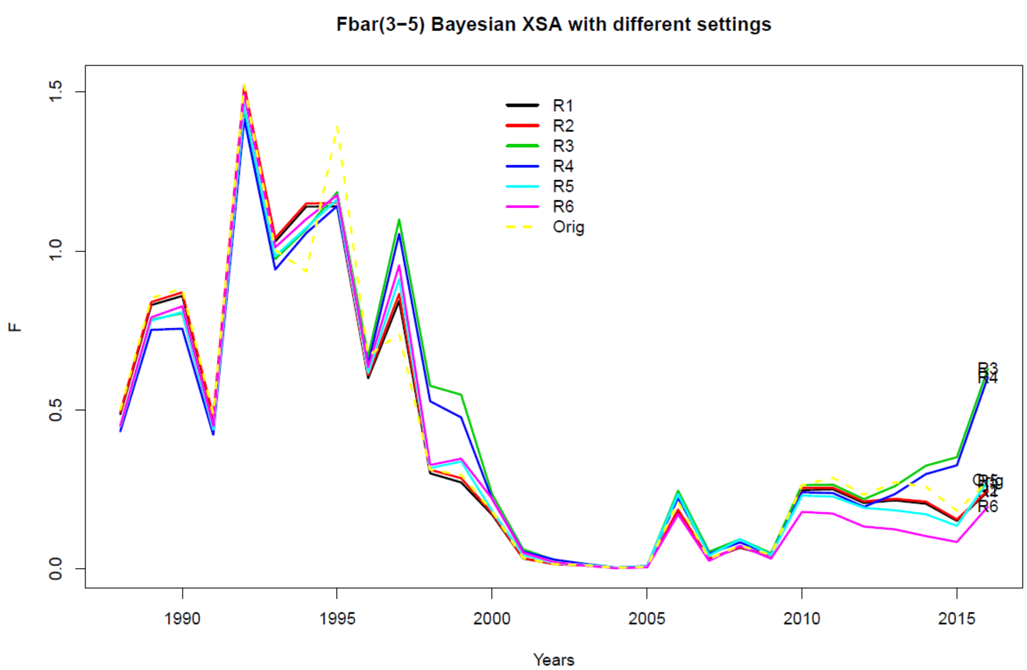


Fig. 12 Results of the posterior median Fbar(3-5) for the different runs of the Bayesian XSA (R1-R6) as well as the approved assessment (labelled as "Orig").

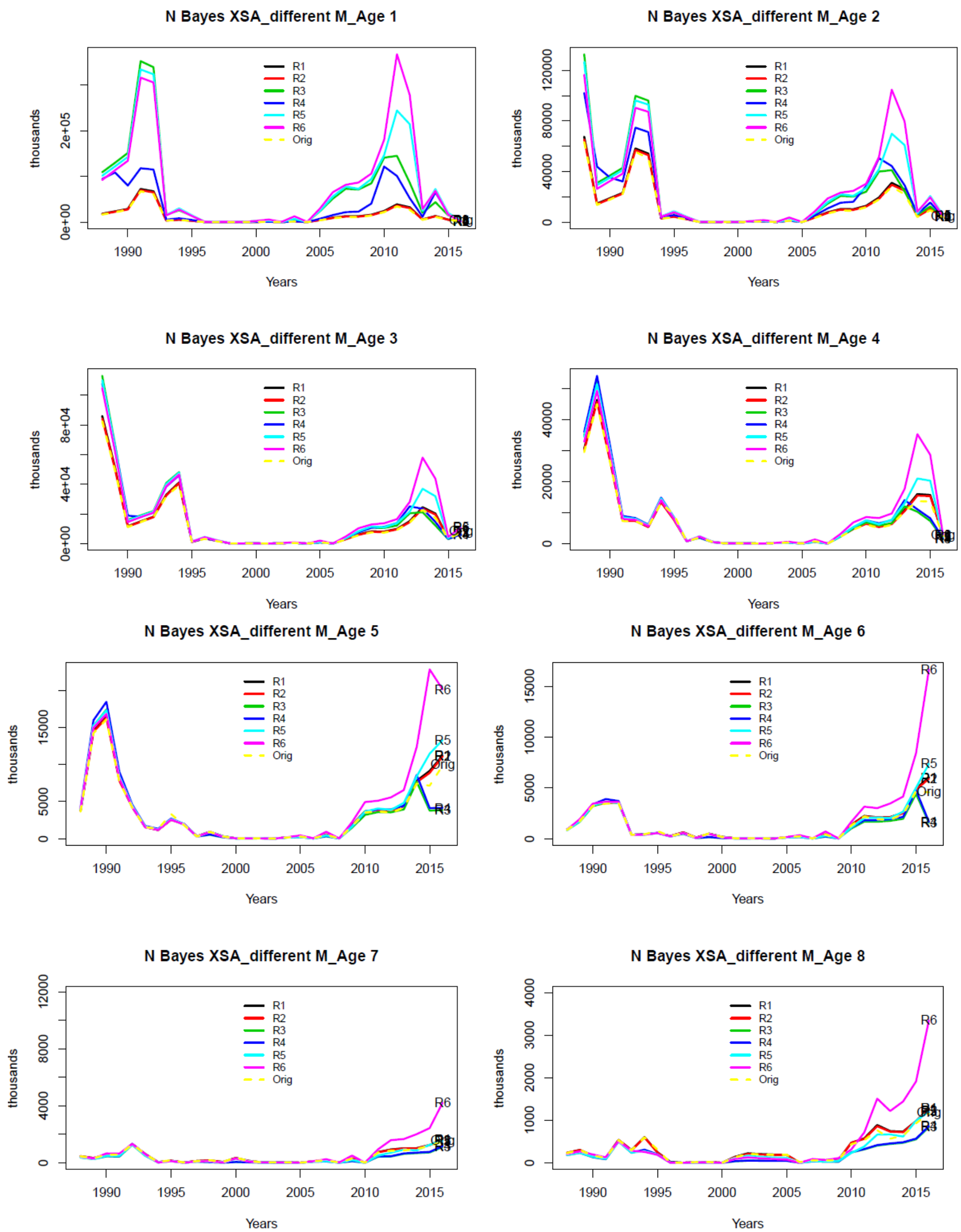


Fig. 13. Results of the posterior median numbers for the different runs of the Bayesian XSA (R1-R6) as well as the approved assessment (labelled as "Orig").

In the results, we have three different groups: R3 and R4 show the lowest SSB and R6 the highest. The difference between R6 and the rest of the models is remarkable. The differences come mainly from the numbers at age in ages 5+. R1 and R2 are very similar, which means that the revised data and shortened assessment period do not affect the results of the assessment. R5 is fairly similar to R1 and R2, except for the recruitment estimates, so changing settings as having M variable over ages, the CV of the prior of the survivors or to have age 2 independent of abundance have not a great impact affect the results. Instead of that, fixing the value of M has a great impact in the subsequent numbers, mainly in ages 5 and 6, that are much lower that for the rest of the scenarios. The main difference arises when we change the value of the median of the prior over the recruitment and when we fix the CV of the EU survey index. In that case, the SSB is much higher than in the rest of the models. The SSB of the last approved assessment is the lowest in the last years if we take out R3 and R4. SSB result in the SSB of 2016 is more than three times the SSB in 2016 of the last approved assessment.

One of the parameters contributing to the difference in the level of SSB between runs R6 and R1 (R1 was the approved assessment from June 2017) is the survey catchability. Figure 14 show the catchabilities by age for both scenarios. Take into account that R1 estimates the catchabilities to be different for all ages (1-8+), while the R6 groups the ages (1, 2, 3 and 4-8+). We can see that while the median catchability in the R6 is around 1 for ages older than 3, in the case of the R1 they range between 1.19 at age 7 and 1.83 at age 5 with bigger variability. For that, the abundance in the last year are bigger for the R6 that for the R1, giving bigger SSB.

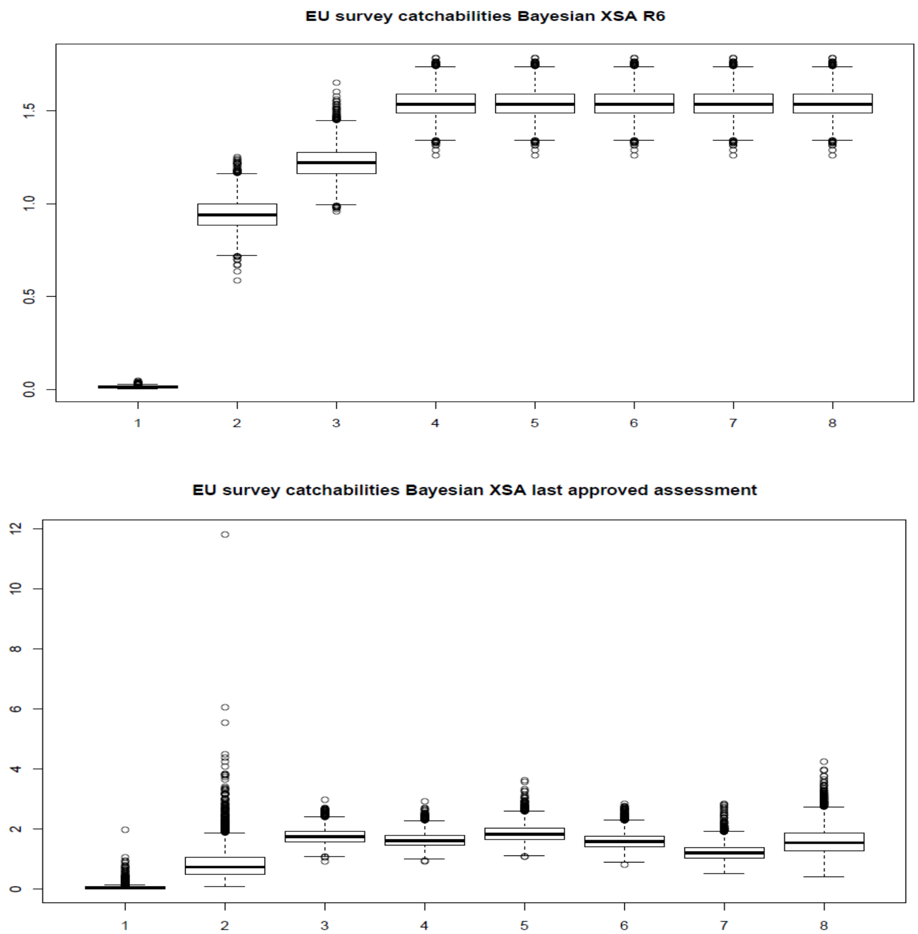


Fig. 14. Survey catchabilities for R6 (top) and for the XSA last approved assessment in 2017 (bottom).

After looking to all the scenarios run and with the modifications made during the benchmark and based on the knowledge and information about the fishery and the survey, the SC decided that the base case for the Bayesian XSA to be presented during the June SC meeting is R6, as the value of the prior median of recruitment, as well as the CV over the EU indices and the groups in the catchabilities are more appropriate than in the rest of the scenarios. Due to the lack of time, the analyses of this model could not be completed during the Benchmark. Therefore, more work will be done to be presented to the June SC meeting, mainly regarding the sensitivity of the model to the parameters chosen in R6 and to run the retro of the final model.

b) Bayesian SCAA

Model structure

A Bayesian statistical catch at age (SCAA) model was applied to the data. Ages are from a=1 to A=8+ and years are from y=1 (1988) to Y=29 (2016). The cohorts are modelled forwards in time, starting from the recruits (age 1) in each year and abundance of each age 2-8+ in the first year, taking into account the natural and fishing mortality. The main specifications of the model are:

Recruits (Age 1) each year, N[y,1], for y=1,...,Y.

$$N[y,1] \sim logN(median = medrec, CV = cvrec),$$

Numbers at age in the first year, N[1,a], for a=2,...,A.

$$N[1,a] \sim LogN$$

$$(median = medrec \times e^{-\sum_{i=1}^{a-1}(M[1,i]+medF[i])}, CV = cvyear)$$

$$N[1,A+] \sim LogN(median = medrec \times \frac{e^{-\sum_{i=1}^{A-1}(M[1,i]+medF[i])}}{1-e^{-(M[1,A+]+medF[A+]}}, CV = cvyear1)$$

Forward population each year and age, N[y,a], for y=2,...,Y and a=2,...,A.

$$N[y,a] = N[y-1,a-1]e^{-Z[y-1,a-1]}$$

$$N[y,A+] = N[y-1,A-1]e^{-Z[y-1,A-1]} + N[y-1,A+]e^{-Z[y-1,A+]}$$

$$Z[y,a] = M[y,a] + F[y,a].$$

F[y,a] **=**f[y]*rC[y,a] semi separable, y=1,…,Y and a=1,…,A.

it is assumed that *rC(y,A+) = rC(y,A-1)* and that *rC(y, a=aref) =1*

a. ln(*f[y]*) is modelled as an AR(1) process over the years, with autocorrelation parameter *rhof*. The median and CV of the marginal prior distribution of *f[y]* in each year are *medf* and *cvf*, respectively.

   rhof is assigned a Uniform(0,1) prior distribution,

b. For each age different from *aref* and A+, ln(*rC[y,a]*) is modelled as random walk over the years, independently from age to age.

$$rC[1,a] \sim LogN(median = medrC[a] = cvrC[a])$$

The distribution in subsequent years (y>1) is given by a random walk in log scale:

$$ln(rC[y,a]) \sim N(mean = ln(rC[y-1,a]) \frac{1}{variance} = taurCcond[a])$$

$$taurCcond[a] \sim Gamma(shape = s1.Ccond, rate = s2.Ccond)$$

Total catch, Cton[y], y=1,…,29

$$Cton[y] \sim Log(median = \sum_{a=1}^{A+} mu.C[y,a] \times wcatch[y,a], CV = cvCW)$$

$$mu.C[y,a] = N[y,a]\left(1 - e^{-Z[y,a]}\frac{F[y,a]}{Z[y,a]}\right)$$

Catch Number at-age, *C[y,a], for y=1,...14,19,...,Y and a=1,...,A.*

$$\ln(C[y,a]) \sim N(mean = \ln(mu.C[y,a]), \frac{1}{variance} = psi.C[a])$$

$$psi.C[a] \sim Gamma(shape = s1.C, rate = s2.C)$$

Survey indices*, CPUE.EU[y,a], y=1,...,Y and a=1,...,A*

$$\ln(CPUE.EU[y,a]) \sim N(mean = \ln(mu.CPUE.EU[y,a]), \frac{1}{variance} = psi.EU[a])$$

$$mu.CPUE.EU[y,a] = phi.EU[a]\left\{N[y,a]\frac{\exp(-alpha.EU * Z[y,a] - \exp(-alpha.EU * Z[y,a]}{(beta.EU - alpha.EU) * Z[y,a]}\right\}^{gamma.EU[a]}$$

$$\ln(phi.EU[a]) \sim N(mean = medlogphi, \frac{1}{variance} = taulogphi)$$

$$gama.EU[a] \sim N\left(mean = medgamma, \frac{1}{variance} = taulogphi\right)\ for\ a < aindep$$

$$gamma.EU[a] = 1\ for\ \geq aindep$$

$$psi.EU[a] \sim gamma(shape = shpsi, rate = rtpsi)$$

Model parameters

The unchanged parameters used in all the scenarios of the model are summarized in table 8

Table 8. Parameters in the priors of the Bayesian SCAA that are common to all the runs (for which the parameter is applicable).

Parameter	***Value***	***Parameter***	***Value***
medF	C(0.0001,0.1,0.5,0.7,0.7,0.7,0.7)	*rtpsi*	0.07
cvCW	0.08	*alpha.EU*	0.5
S2.C	0.345	*beta.EU*	0.58
mMedlogphi	0	*medf*	0.2
taulogphi	1/5	*rhofmin*	0
mMedgama	1	*aref*	5
tTaugama	1/0.25	*medrC*	c(0.001,0.3,0.6,0.9,1,1,1)
sShpsi	2	*S2.Scond*	*0.04*

Different scenarios (37) with this model changing some of the settings and/or input data were run. Model settings common to all the runs are presented in Table 8.

In the model, M can be treated either as common value over years and ages, a vector or a matrix and it can be fixed or be assigned a prior distribution. In the configuration favoured by the benchmark, M is a vector (i.e. age-dependent, constant over the years), which follows a log-Normal prior distribution for each age. The prior median for each age was assigned based on life-history considerations and outcomes from explorations of the multispecies GadCap model (Table 5); a 15% prior CV was used.

Given the very low catch numbers observed at age 1, the benchmark decided to set the catch at age 1 data equal to zero in all years and to assume in the model that F at age 1 is equal to zero. The benchmark also decided to treat both the zeros observed in the survey abundance indices at age and those observed in the catch at age matrix for ages > 1 as missing values. This procedure on the data was not applied in the initial set of runs presented at the benchmark (in those earlier runs, zeros had been replaced by very low values) and it was agreed to include it in runs conducted after the benchmark. The results are not altered by this change.

The complete list of scenarios run is provided in Table 9. The rationale of each of the parameters is the following:

- $taurCcond[a]$: this parameter controls the variability (CV) of the prior of the selectivity (rC) between years. It can be selected to be different for each age or the same for groups of ages. If *S2.Ccond* is fixed at 0.04, then *S1.Ccond*=4 results in the median of the prior distribution of this CV being 10%, whereas *S1.Ccond*=0.75 results in the prior median of the CV being 30%.

-*psi.C[a]*: this parameter controls the variability (CV) of the observation equations for catch numbers-at-age. It can be selected to be different by age or grouped by groups of ages. The higher the value of this parameter, the closer the model must follow the abundance catch-at-age (so, closer to an XSA model).

-*psi.EU[a]*: this parameter controls the variability (CV) of the observation equation for the survey abundance index at age. It can be selected to be different by age or grouped by groups of ages.

-*phi.EU[a]* (qs): these are the catchabilities of the surveys by age. They can be all different or can be grouped by groups of ages.

-*CVs*: four different CVs were changed in the runs conducted at the benchmark: cvrec (CV over the annual recruitment), cvyear1 (CV over the numbers by age in the first year), cvf (the CV over f) and cvrC (the CV over rC, so, selectivity by age and year; this is related to the model parameter $taurCcond[a]$).

-*adep*: this is the set of ages for which the survey catchability depends on population abundance (see Table 10); so gama.EU=1 for all ages not belonging to the set *adep*.

-*Y/Y-1*: last year used in the calculation of the catch-at-age (see Table 10; it should have been Y in all runs, but was Y-1 in some of them due to an oversight).

-*Zeros*: if the zeros in the catch at age and in the survey index are included as zeros (actually, they were replaced by very low values) or treated as missing values (see Table 10 and the discussion earlier in this document about benchmark decision).

-*medrec*: this parameter is the median of the prior distribution for the recruitment.

-*cvrec*: this parameter controls the variability (CV) of the prior distribution for the recruitment.

Table 9. Settings of Bayesian SCAA runs. In the column labels *s1.Ccond* relates to interannual variability in selectivity and *s1.C* relate observation equation for catch at age; *cvf* and *cvrC* are prior CVs on f(y) and selectivity-at-age in initial year. In each row, values in red indicate how that run differs from the corresponding "Base run". R37 (highlighted row) was the scenario accepted as the new assessment method.

Run	**Base Run**	***S1.C cond***	**Age groups survey catchability**	**Age groups CV of catch-at-age**	**Age groups CV of survey**	**Age groups interannual variability in selectivity**	***S1.C***	***adep***	***cvf & cvrC***	**M**	**Y/Y-1**	**Zeros**	***medrec***	***cvrec***	***cvyear1***	**DIC**	**Penalty**
1		4	All different	All different	All different	All different	4	1,2	1	0.19	Y-1	Incl	15000	2	1	1376	94.3
2	1	0.75	All different	All different	All different	All different	4	1,2	1	0.19	Y-1	Incl	15000	2	1	1128	329.5
3	2	0.75	1,2,3-6,7-8	All different	All different	All different	4	1,2	2	0.19	Y-1	Incl	15000	2	2	1010	206.5
4	2	0.75	1, 2-8	All different	All different	All different	4	1,2	1	0.19	Y-1	Incl	15000	2	1	1085	284.6
5	3	0.75	1,2,3-6,7-8	1-8	1-8	All different	4	1,2	2	0.19	Y-1	Incl	15000	2	2	1580	138.8
6	3	0.75	1,2,3-6,7-8	1-8	1-8	1-8	4	1,2	2	0.19	Y-1	Incl	15000	2	2	1646	98.9
7	3	0.75	1,2,3-6,7-8	1-2,3-6,7-8	All different	All different	4	1,2	2	0.19	Y-1	Incl	15000	2	2	1029	235.5

8	3	0.75	1,2,3-6,7-8	All different	All different	All different	100	1,2	2	0.19	Y-1	Incl	15000	2	2	511	222.7
9	2	0.75	1,2,3-8	All different	All different	All different	4	1,2	2	0.19	Y-1	Incl	15000	2	2	1000	221.1
10	9	0.75	1,2,3-8	All different	All different	All different	4	1	2	0.19	Y-1	Incl	15000	2	2	1092	269.4
11	10	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	2	0.19	Y-1	Incl	15000	2	2	1037	189.7
12	11	0.75	1,2,3-8	1-2,3-6,7-8	All different	1,2-8 (-5)	4	1	2	0.19	Y-1	Incl	15000	2	2	1154	182.6
13	12	0.75	1,2,3-8	1-2,3-6,7-8	1,2,3-8	1,2-8 (-5)	4	1	2	0.19	Y-1	Incl	15000	2	2	1191	183.5
14	12	0.75	1,2,3-8	1-2,3-6,7-8	1,2-8	1,2-8 (-5)	4	1	2	0.19	Y-1	Incl	15000	2	2	1217	168.8
15	14	0.75	1,2,3-8	1-2,3-6,7-8	1,2-8	1,2-8 (-5)	4	1	4	0.19	Y-1	Incl	15000	4	4	1350	292.4
16	14	0.75	1,2,3-8	1-2,3-6,7-8	1,2-8	1,2-8 (-5)	4	1	16	0.19	Y-1	Incl	15000	16	16	1298	259.4
17	11	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	2	0.19	Y	Incl	15000	2	2	1119	233.7
18	11	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	4	0.19	Y-1	Incl	15000	4	4	1085	229.3
19	18	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	4	0.19	Y	Incl	15000	4	2	1135	276.3
20	15	0.75	1,2,3-8	1-2,3-6,7-8	1,2-8	1,2-8 (-5)	4	1	4	0.19	Y	Incl	15000	4	4	1272	246.0
21*	19	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	4	0.19	Y	Incl	15000	4	4	1449	176.0
22	19	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	4	Vector	Y	Incl	15000	4	4	1108	269.8

23	19	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	4	Matrix	Y	Incl	15000	4	4	1052	194.4
24	19	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	2,2	0.19	Y	Incl	15000	4	2	1131	250.7
25	19	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	2,2	0.19	Y	Incl	15000	4	4	1072	204.8
26	19	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	4,2	0.19	Y	Incl	15000	4	4	1157	276.5
27	19	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	4,4	0.19	Y	Incl	15000	4	4	1913	896.1
28	19	NA	1,2,3-8	Fix (20%)	Fix (30%)	Fix (20%)	NA	1	4	1 prior	Y	Incl	15000	4	4	1050 1	170.1
29	28	NA	1,2,3-8	Fix (20%)	Fix (30%)	Fix (20%)	NA	1	4	0.19	Y	Incl	15000	4	4	1048 1	125.8
30	19	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	4	1 prior	Y	Incl	15000	4	4	994	219.3
31	19	0.75	1,2,3-8	All different	All different	1,2-8 (-5)	4	1	4	8 priors	Y	Incl	15000	4	4	1022	229.3
32	28	NA	1,2,3-8	Fix (20%)	Fix (30%)	Fix (20%)	NA	1	4	1 prior	Y	NA	15000	4	4	1612	127.0
33	32	NA	1,2,3-8	Fix (20%)	Fix (30%)	Fix (20%)	NA	1	4	8 priors	Y	NA	15000	4	4	1639	148.8
34	28	NA	1,2,3-8	Fix (20%)	Fix (30%)	Fix (20%)	NA	1	4	8 priors	Y	Incl	15000	4	4	9736	130.1
35	34	NA	1,2,3-8	Fix (20%)	Fix (30%)	Fix (20%)	NA	1	4	8 priors, medM	Y	Incl	15000	4	4	9693	128.4
36	34	NA	1,2,3-8	Fix (20%)	Fix (30%)	Fix (20%)	NA	1	4	8 priors, cvM	Y	Incl	15000	4	4	9693	161.8
37	33	NA	1,2,3,4-8	Fix (20%)	Fix (30%)	Fix (20%)	NA	1	4	8 priors, cvM=0.15	Y	NA	15000	4	4	1596	121.7

In order to compare between scenarios, an attempt to calculate the deviance information criterion (DIC) was made. DIC is a hierarchical modeling generalization of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The idea is that models with smaller DIC should be preferred to models with larger DIC. However, we found the calculation of the DIC inside the model to be unstable: running a given model twice with the same settings, the DIC changed between runs. Although in general the change is not too high (less than 10%, albeit the difference is the double in some cases), it changes sometimes the order of the preferred models. It is not clear that this is working reliably to choose model settings on its own, and it could perhaps be better to let the choice be guided by the biological rationale underlying the settings considered in each run for similar DIC values.

Model outputs

All the models reach the convergence after 100,000 iterations, from which 1000 are taken for the results.

The approach followed to explore potential scenarios was to start the first runs allowing the model to estimate all the parameters independently (resulting in many parameters to estimate) and then, looking at the results and using the available biological and fishery information, to try and reduce the number of parameters fitted by the model.

With regards to **catchability** in the survey (*phi.EU[a])*, we tried four different age groupings: 1,2,3-6,7-8+; 1,2-8+; 1,2,3-8+; 1,2,3,4-8+. Based on the biological and survey gear information, we have chosen to group the ages as follows: 1, 2, 3 and 4-8+. This setting was used in R37.

The CV of the observation equation of the **catch-at-age** (relates to *psi.C[a]*) can be estimated (including a prior) or can be assigned a fixed value, and it can be different by age or grouped by groups of ages. We tried here four different approaches: CVs different for all ages, CVs equal for all ages, CVs grouped by ages 1-2, 3-6, 7-8+, or CV fixed for all the ages. When we take all CVs to be equal, the SSB increases incredibly (R5 and R6). Examining the prior and posterior of the runs we can see that for ages 3 to 7 the posterior of the CV is around 30%, but for ages 1, 2 and 8+ the estimated CV value is not consistent as it is too high (depending on the scenario, the CV was as much as 2000% in some of the ages). It is, therefore, necessary to force the model to reduce the CV in those ages, and the best way to do that seems to be to force the CVs to be equal by groups of ages. The age groups 1-2, 3-6 and 7-8+ seem to be quite logical, as the bulk of the catches is always between 3 and 6, so the CV of those ages probably is smaller, while ages 1 and 2 and ages 7 and 8+ are less represented in the catch and so the CV must be different. But it is still a problem with the CVs of ages 1, 2, 7 and 8+, that are still too high. To deal with this, the benchmark decided to fix this CV at 20% for all ages, and this was the value used in the all runs starting from R32. A CV of 20% seems reasonable for the catch-at-age of this stock taking into account their variability.

A parameter related to the CV of the observation equation of catch-at-age is *S1.C* (see equations earlier in this document for technical detail). We think that 4 is a sensible value for this parameter, as it results in a CV of around 30% for the observed catch-at-age, but we made a run with a much higher value of this parameter (100) in order to force the model to follow the observed catch-at-age, similarly to what the XSA does. The results in SSB for this run (R8) are similar to the Bayesian XSA approved in June 2017. In the last runs, as the CV of the catch-at-age is fixed, this parameter does not exist in the model.

For the CV of the observation equation of the **EU survey indices** (relates to *psi.EU[a]*), we tried different age groupings: all different, all equal, 1,2,3-8+ and 1,2-8+. If we take all different, we can see that the CVs between 2 and 6 are quite stable, but high for ages 1, 7 and 8+. The next step was to try to reduce the number of parameters. After consideration of several results, and the fact that the survey covers quite well the age composition and stock distribution, the meeting agreed that the adopted approach would be to have a common CV of 20% (fixed) across all ages.

With regards to the interannual variability in the prior of the **selectivity by age** (related to *taurCcond* parameter) we have tried different age groups: all different, all equal, and 1,2-8+ (always excluding age 5, the reference age for selectivity in the model). If we allow the variability to be different for all the ages, the CV of most ages is extremely high, particularly for ages 1 and 2. If we instead take just two groups, 1,2-8+, we can see that, although the CV is still very high, particularly for age 1, it is one of the lowest if we look to all the runs. Based on these findings, the benchmark decided that in order to have CVs in a range of values that seem logical, it was better to fix this value at 20%.

A parameter related to the interannual variability in selectivity is *S1.Ccond* (see equations earlier in this document for technical detail). Although we started the runs with a value of *S1.Ccond* corresponding to a prior distribution for the CV centred at 10% (R1), taking into account the great variation over the years in the selectivity we think that, in the case of having a prior over the CV (i.e. over the interannual variability in selectivity), it is more logical to center such a prior at around 30% (R2-27, 30-31). In the last runs, as the CV of the selectivity is fixed, the *S1.Ccond* parameter does not exist in the model. The greatest interannual variability in the selectivity is mainly due to the closure of the fishery between 1999 and 2009, when the catches were mainly by-catch of other fisheries.

With regards to the **first age** for which the survey catchability is independent of population abundance (so, gama.EU=1), we have tried two different approaches: catchabilities of ages 1 and 2 dependent on abundance and only age 1 catchability dependent on abundance. It was decided to use only age 1 dependent on abundance. The rationale for this decision is that when we use ages 1 and 2, the posterior median of the distribution of gama.EU at age 2 is almost equal to 1 and the distribution is quite narrow, so it seems reasonable to fix gama.EU for age 2 at 1, as for the ages 3-8+.

We have tried four different values for two different **CVs** (*cvf* – variability in f[y] and *cvrC – recruitment variability*), namely 1, 2, 4 and 16 (100%, 200%, 400% and 1600%). We think that the values 1 and 2 are a bit small for our data and 16 a bit high, so the best one seems to be 4.

During the March SC meeting on input data (by WebEx), it was decided to try three different approaches for **M**: a single value for all ages and years, a vector by age constant by year, and a matrix variable by age and year. The values used are the same as used in the XSA model. During the benchmark, it was decided to run some scenarios for which the M was not a fixed input but an output of the model (i.e. estimated) via a prior, as in the approved Bayesian XSA model (González-Troncoso, 2017). In this context, two different settings were analyzed: M with one prior with median equal to 0.19 (M equal for all ages and years), and M with eight priors, one for each age, with median equal to the fixed vector used in previous runs (M equal for all years but different by age). The posterior median of two of the scenarios, R32 and R37, are presented in Table 10. It can be seen that, when we allow the M to be different between ages, the value for age 1 is much higher than for the rest of the ages. For ages 2 and 3 it is still high, and then it decreases to increase again a bit at age 8. Based on the biological information available for this stock (cannibalism, age composition, etc), this seems more logical than having a single value for all ages, so the benchmark decided to use this approach in the last run (R37).

Table 10. Results of the posterior median over M for R32 and R37

Age	1	2	3	4	5	6	7	8
R32	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
R37	1.17	0.57	0.36	0.26	0.27	0.34	0.30	0.38

In the first scenarios run, the catch-at-age data were used in the model until year Y-1 instead of Y. This was an oversight and was fixed after R19.

There are some zeros in the catch-at-age in numbers and in the EU index-at-age, particularly in the older ages. As having these zeros in the data seems to be a sampling issue more than the reality, the benchmark decided to replace them in the data by missing values. This approach was used in the last scenarios (R32, R33, R37).

With regards to the prior median of the annual **recruitment**, a value of 15000 was set in the first runs, as it was the value used in the last approved assessment. This value was chosen in 2008 taking into account the recruitments of the previous years. During the benchmark, examining the results of the Bayesian SCAA model, a more logical value of 45000 was set taking into account the recent recruitments of this stock that in almost all scenarios and almost all years are well above 15000, and the fact that M is now centred around considerably larger values than used in previous assessments, particularly for age 1. To try and prevent the prior distribution for recruitment to have undue impact on model results, it was considered appropriate to increase the CV of the prior distribution from the originally considered values (2 and 4) to 10, i.e. 1000%.

The CV of the prior distribution of the **numbers-at-age in the first year** was set at various values between 1 and 16. The benchmark recommended exploring a scenario with this CV set to 10 (1000%) to determine whether the prior values were constraining results in the runs which have a lower CV value.

Figure 15 show the SSB for all the scenarios run. As there are too many runs, we select some of them which are considered more representative of the different behaviors encountered. To select them, the run settings and the results in the SSB were taken into account. Therefore, we have the rest of the plots for the following runs: R5, R8, R16, R19, R21, R22, R23, R28, R31 and R37. Figures 16-19 show the the SSB, the R, the F_{bar} and the Number at age for each of those scenarios, respectively.

Concerning the SSB estimates for recent years, we note that R5 and R6 result in substantially larger estimates and R8 and R29 in substantially smaller estimates than the rest of the runs. Excluding these 4 runs, all others produce estimates of SSB in 2016 ranging between 66000 and 127000 tons. The base case at the time the benchmark ended, R37, estimates this SSB at 93000 t.

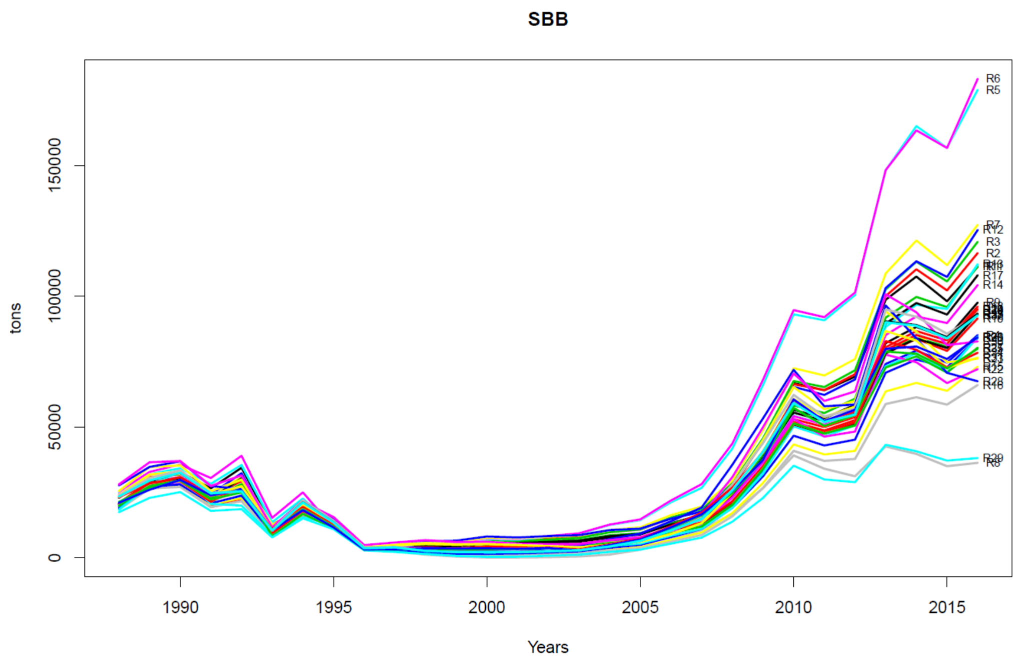


Fig. 15 Spawning stock biomass (SSB) estimates from all SCAA runs conducted during the meeting

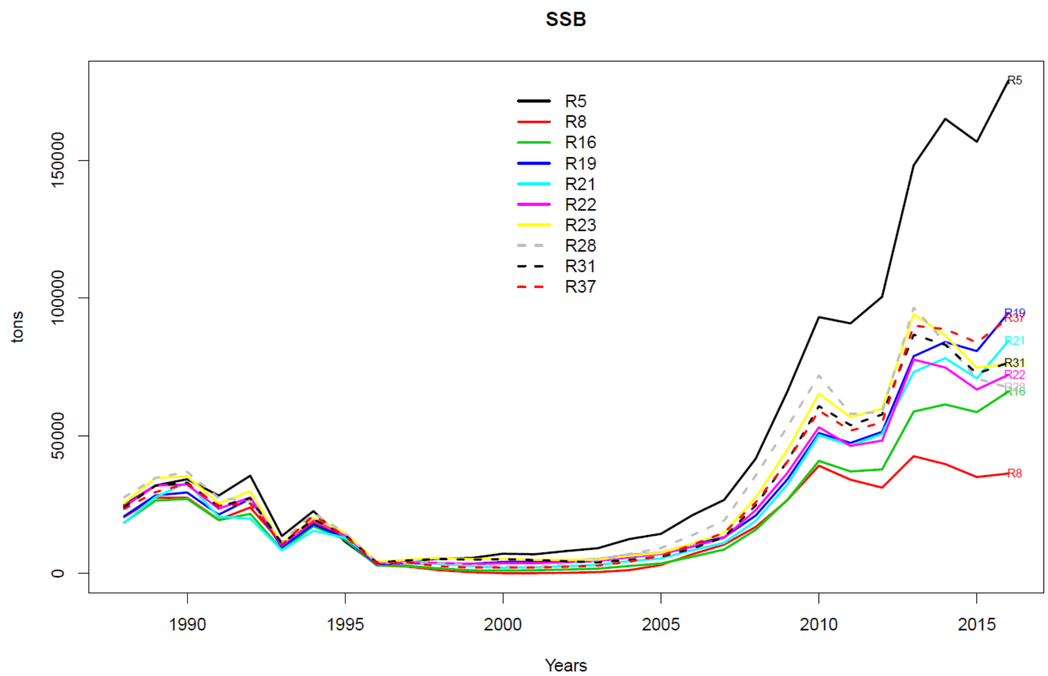


Fig. 16 Spawning stock biomass (SSB) estimates from the selected SCAA runs

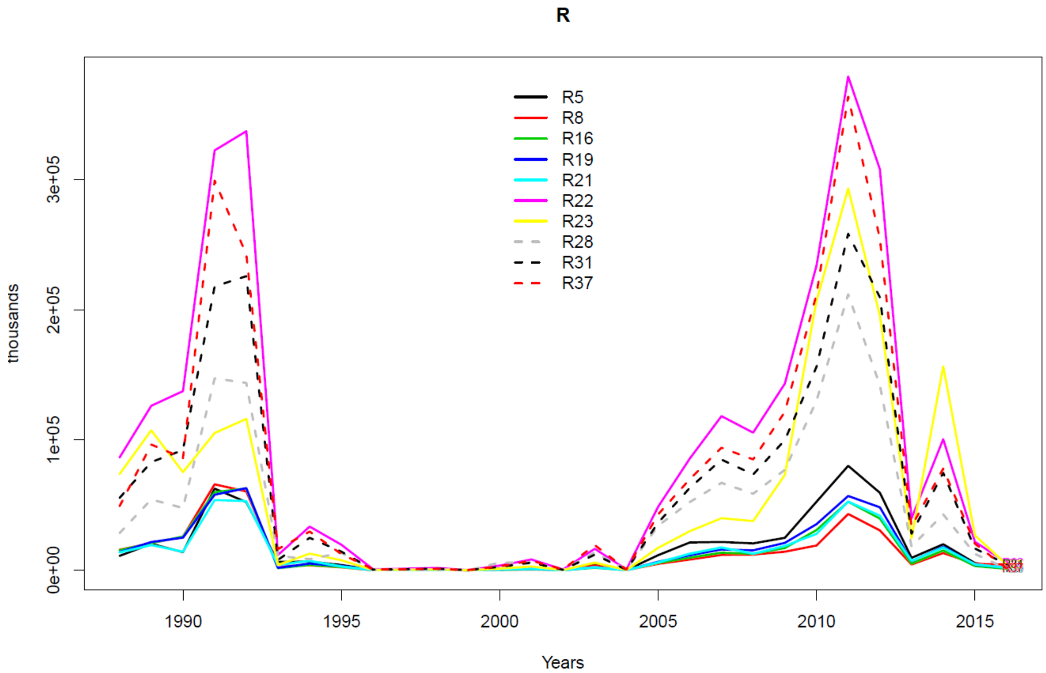


Fig. 17. Recruitment age 1 (R) estimates from the selected SCAA runs

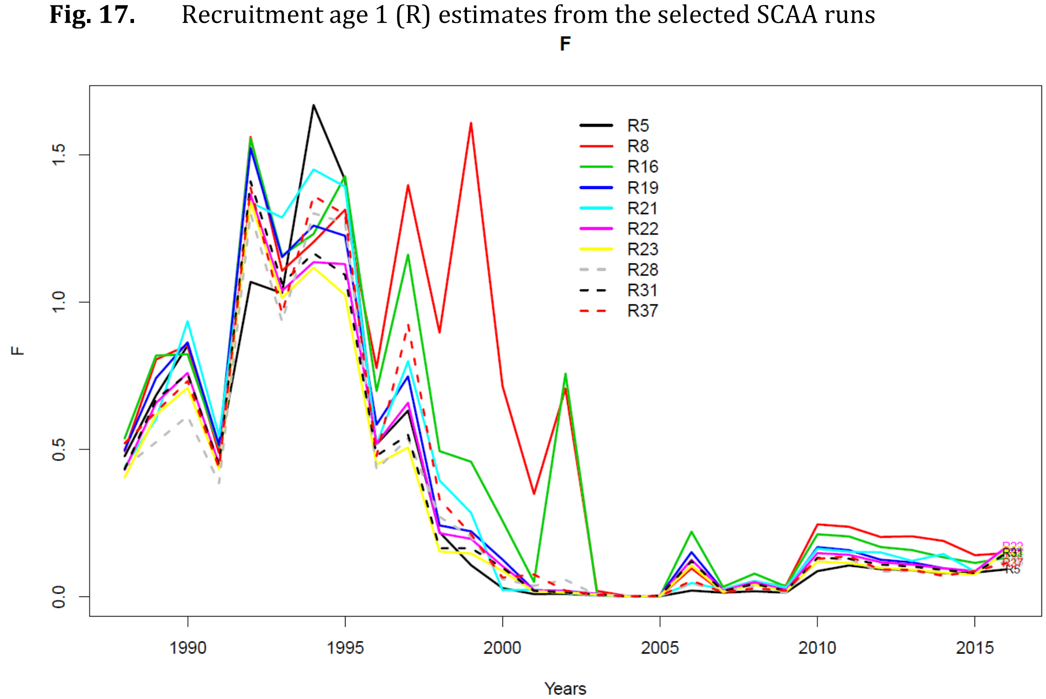


Fig. 18. fishing mortality (F) estimates from the sleceted SCAA runs

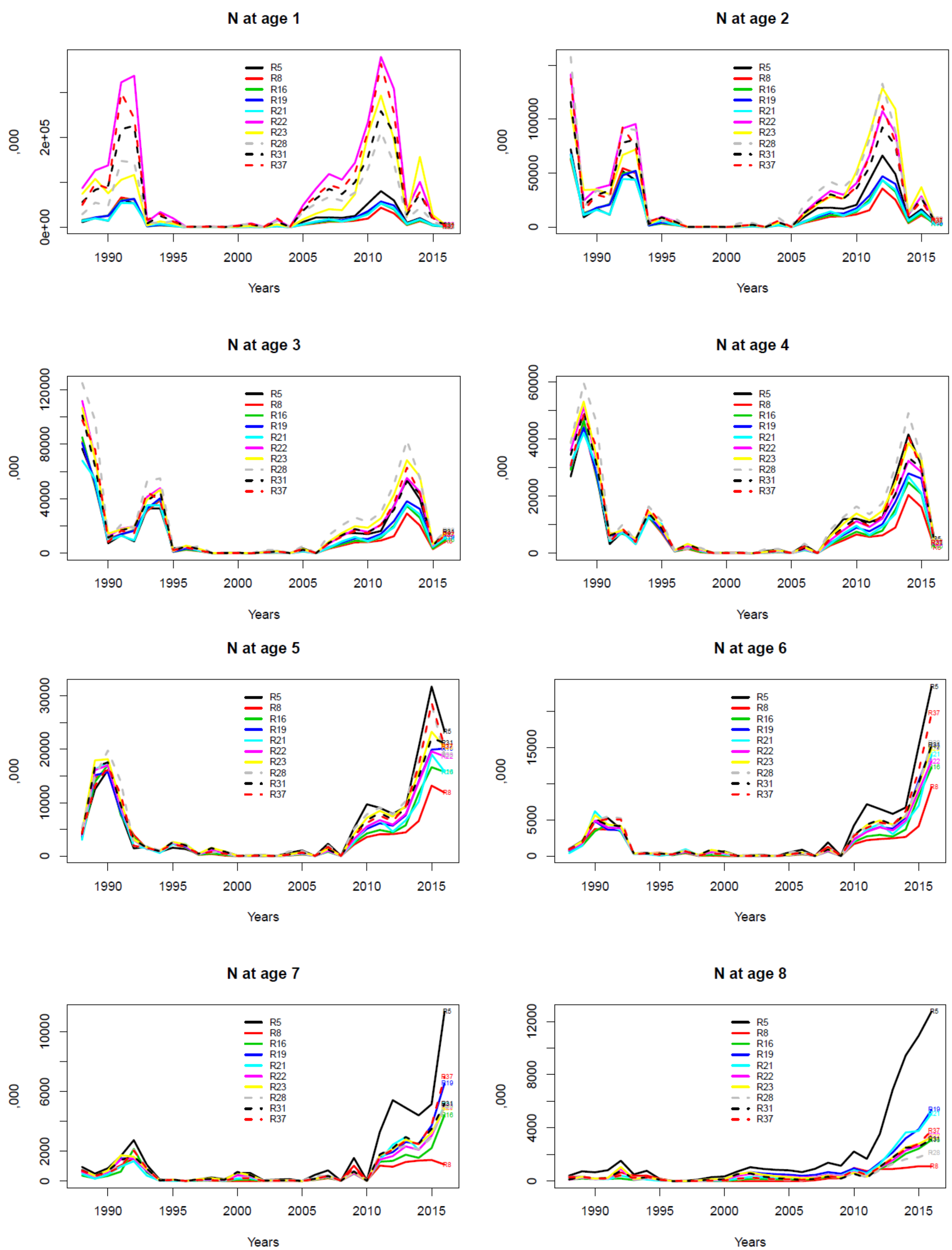


Fig. 19. numbers at age estimates from the selected SCAA runs

After looking to all the scenarios run and with the modifications made during the benchmarkand based on the knowledge and information about the fishery and the survey, the SC decided that the base case to be presented to the June SC meeting should have a structure similar to R37; however, it was noted that it would be more appropriate to use medrec=45000 and cvyear1=10 and that a full set of diagnostics would be needed for the modified R37 run.

So, the final model settings agreed during the meeting were run 37:

- Fix cvcaa (30%), cvEU (20%), cvSelectivity (30%)
- qs: 1, 2, 3, 4-8
- Zeros as N/A
- Estimated M by age via a prior (constant over time)
    - Vector of Ms and cv=0.15
- q dependent on N: only at age 1

Outputs from this final run are presented in Fig. 19

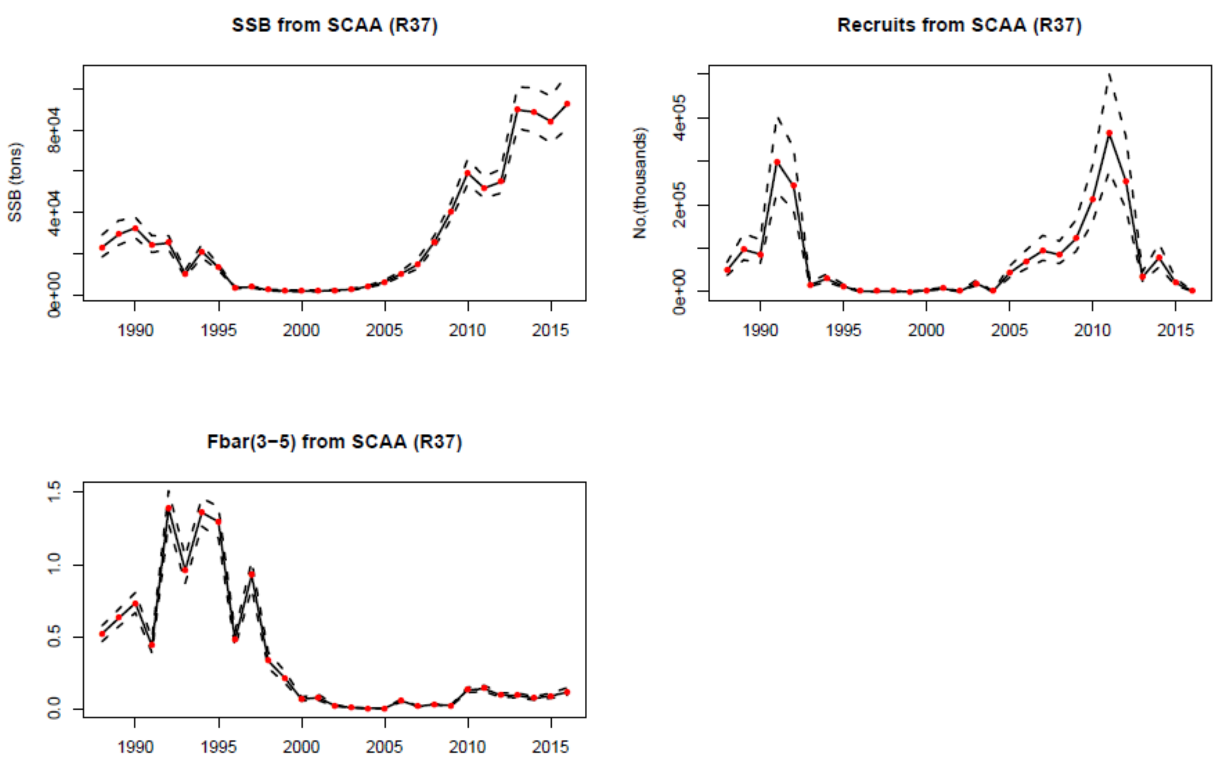


Fig. 20. Outputs of the final run of the Bayesian SCAA (run 37)

c) SAM

A series of runs with the state space assessment model (SAM) (Nielsen, A and Berg, 2014) were presented, in order to provide a comparison with the runs from other assessment models (Bayesian XSA and Statistical Catch at Age, SCAA) presented at the meeting, and to explore alternative configurations.

Model description

In SAM, the "states" (fishing mortalities and abundances at age) change from year to year according to a transition matrix representing the processes in the stock (survival equation and Baranov catch equation). Process have their associated deviations: recruitment and fishing mortality at age are modelled as random walks (with the corresponding variances being parameters estimated by the model). Fishing mortality random walk can be set completely independent across ages, correlated across ages (the correlation between adjacent ages being a parameter estimated) or entirely correlated. Abundances-at-age can deviate from the survival equation through a process error log-normally distributed (which variance is also a parameter estimated).

SAM is a fully statistical model in which all data are treated as observations. The model estimates observation variances (log normal error model) for each data sources (catches and surveys), which can be used to describe how well each data source is fitted in the model (a low observation variance indicating a strong influence on the model fit).

Uncertainties (standard errors) are estimated for all parameters and for all states (F and Ns) and the quantities derived (SSB and Fbar). This uncertainty is naturally incorporated in the short-term forecast.

SAM offers a fully statistical framework and model selection can be done based on model likelihood. This is particularly convenient in the context of a benchmark, where a range of different model configurations can be statistically compared.

In its current version, parameter estimation is carried out using the minimiser TMB. For this work, SAM was run using the R library "stockassessment" provided on GitHub ("fishfollower/SAM/stockassessment") by the model developer. For a more statistical description of the model, please refer to (Nielsen, A and Berg, 2014).

Model parameters

The model was run using the base SCAA parameters (SCAA R19) and the preferred settings at the beginning of the benchmark meeting (SCAA R15). And 4 further variants:

Table 11. SAM model parameters

	catchabilities		Process stdev		Observation stdev		Nr. params	AIC	Log-likelihood ratio test
Run name	**EU survey**	**exponent**	**Var F random walks**	**Var logN**	**Catches**	**EU survey**			
SAM R19	1,2,3-8	1	1,2-8	1,2-8	All ages	All ages	24	1130.69	
SAM R15	1,2,3-8	1	1,2-8	1,2-8	1-2,3-6,7-8	1,2-8	13	1130.17	
Compare different configurations for observation stdev									
SAM mod2	1,2,3-8	1	1,2-8	1,2-8	1-2,3-6,7-8	1-8	12	1128.30	P = 0.71 (modR15 vs mod2)
SAM mod2.2	1,2,3-8	1	1,2-8	1,2-8	1-2,3-6,7-8	1-2,3-8	13	1127.72	P = 0.10 (Mod2.2 vs 2)
SAM mod3	1,2,3-8	1	1,2-8	1,2-8	1-8	1-8	10	1130.08	P = 0.035 (mod2 vs mod3)
Compare different configurations for catchabilities									
SAM mod4	All ages	1	1,2-8	1,2-8	1-2,3-6,7-8	1-8	17	1114.71	P= 0.0002 (mod 2 vs mod4)
SAM mod5	1,2-5,6,7,8	1	1,2-8	1,2-8	1-2,3-6,7-8	1-8	14	1109.27	P= 0.90 (mod 4 vs mod5)
Compare runs with observation error structure									
SAM mod6	As model 5 but with unstructured correlation matrices for observation errors catches and survey						70	1053.33	P=0 (mod6 vs mod5)

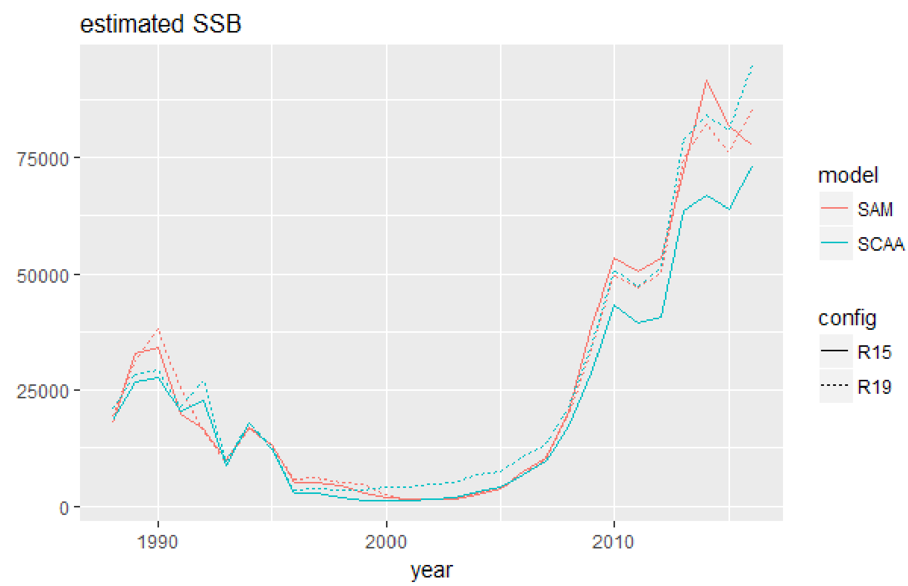


Fig. 21. Estimated SSB(t) using SAM configured with the preferred SCAA configurations R15 and R19.

With the R19 parameters it was observed that M at age 6 was anomalously high in years where the fishery was closed. The reason for this is not fully understood but this was not observed in the R15 variant or other model variants. It is possible that it results from difficulty in the model identifying signals in years where there wasn't much catch. Similar (although smaller) peaks are seen in age 3 in model run 5. It is suggested that adding auto correlation between years in the random walk estimation of F could be tried and that this could eliminate the peaks.

Run 5 used independently estimated catchabilities. The results showed reduced survey catchability of the oldest age groups which resulted in lower F in the oldest ages (flat topped rather than continuously rising F). It was considered unlikely that this pattern of survey selectivity would be realistic in terms of the biology of the stock although it was noted that the patterns that are observed in selectivity may be dependent on assumptions on M.

Run 8 introduced process error correlation matrices.

Final run used the M matrix from the Gadget model. This appeared to result in a slight increase in process error but there appears to be slightly less structure in the process error.

d) GADGET

GadCap (Pérez-Rodríguez et al. 2017) is a GADGET multispecies model for the Flemish Cap cod, redfish and shrimp developed under an EU project a as part of the NAFO roadmap for the EAF. The effect of fishing, trophic interactions (including cannibalism) and water temperature in the dynamic of these three major fishing resources has been modeled. The results highlight the interdependent dynamic of these stocks, and reveals strong interactions between recruitment, fishing and predation (including cannibalism), with marked changes in their relative importance by species-age-length over time. A full description of the model is given in SCR-Doc 16-035 and more recent improvements to the model are detailed in SCR-Doc 18-024.

Estimates of cod abundance, biomass and recruitment are presented in figures 22 to 24.

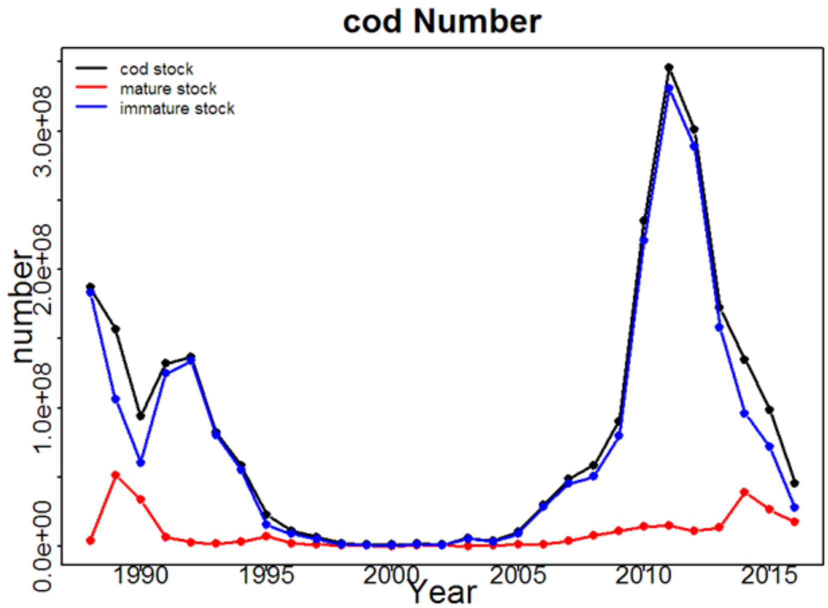


Fig. 22. Abundance estimates (total stock, and mature and immature components) from the GadCap model.

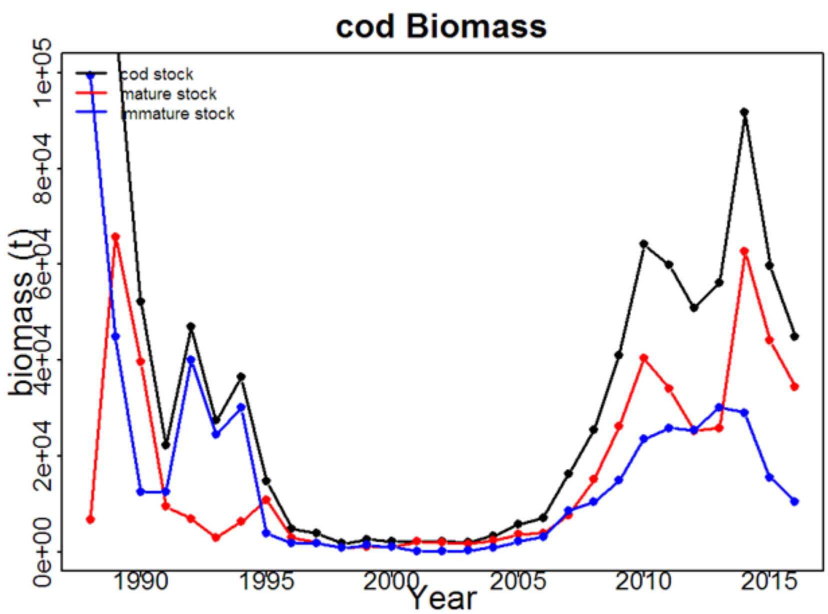


Fig. 23. Biomass estimates (total stock, and mature and immature components) from the GadCap model.

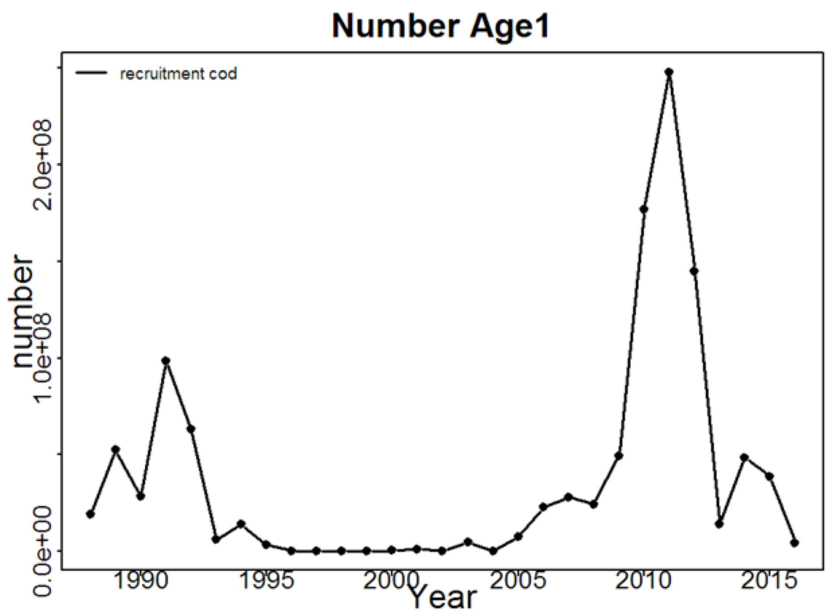


Fig.24. Recruitment (numbers at age 1) estimates from the GadCap model.

e) SC decisions

The following decisions were taken regarding future development of these models:

Bayesian XSA

- The benchmark decided to set R6 as the base case for the Bayesian XSA to be presented during the June SC meeting.
- Due to the lack of time, the analyses of this model could not be completed. More work will be done during the June SC meeting

Bayesian SCAA

- Base case to be presented during the June SC meeting: R37
- Further analysis for SCAA:

  prior median of recruitment of 45000,

  CV on prior for recruitment and abundance at age in the first year as 10

  full set of diagnostics.

- The results of these scenarios will be presented during the SC June meeting

SAM

- For practical reasons, (person doing the work is not present in the meeting) it was agreed that the SAM model will not be the final assessment model from this meeting although it is likely to remain in consideration as one operating models going into the MSE.

7. Comparison between models

A comparison between the models presented was performed. The SAM R19, the GADGET, the SCAA R37 and the XSA R6 were compared. In the case of the SCAA and the XSA, the benchmark preferred runs are compared. In the case of SAM, the most similar run to the SCAA base case at the beginning of the meeting (R19) was considered. In addition, the results of the approved assessment are plotted in order to compare the results with the new models. In order to compare the results, only the results of period 1988-2016 for the last approved assessment are shown, taking into account that this assessment was performed for the period 1972-2016.

It seems that we have two different trends in the SSB since the recovery of the stock, starting about 2005: SAM and SCAA give a bigger SSB than GADGET and approved XSA, the difference being bigger in recent years. XSA (R6) gives a level of SSB in the middle of the two groups, being in recent years more similar to SCAA and SAM (Figure 25). This is probably due to a combination of factors within them:

-Recruitment: SAM, SCAA and XSA (R6) give bigger recruitment than the other two models (Figure 26). It is remarkable the difference in recruitment between models in the strong 2010 cohort (2011 recruitment), which is much bigger for the SAM than for the rest. In the case of the approved XSA, the recruitment is always the lowest estimated, being since 2005 much lower than for the rest of the models. This leads in a lower SSB.

-Numbers at age: Taking a look at the number at age in Figure 28, we can see that the differences come from all ages, but the largest percentage-wise differences are mainly from ages older than 4 years. For ages 1 to 4, the main difference derives from the 2010 cohort, starting in 2011 at age 1. For ages 5 and 6, we have two different levels, one for SAM, SCAA and XSA (R6) and other for GADGET and approved XSA. For ages 7 and 8, the trend is similar to the SSB, having two levels and XSA (R6) in the middle. It is remarkable to note that until age 5, the level of the numbers for all the models are within the historical values, but for ages 6+ SAM, SCAA and XSA (R6) give numbers never seen since the beginning of the series. This is reflective of the substantial differences in age-specific mortality. As the oldest ages are the ones than contribute more to the SSB, the difference in the value of SSB between models comes from the difference in ages older than 4.

With regards to fishing mortality, although the values of F are not available for SAM, we can see that, in general, the F_{bar}(3-5) is bigger for the case of the approved XSA (Figure 27), being this logical as we are estimating lower population abundance with the same catch.

It seems that the cod assessment is quite sensitive to the model chosen.

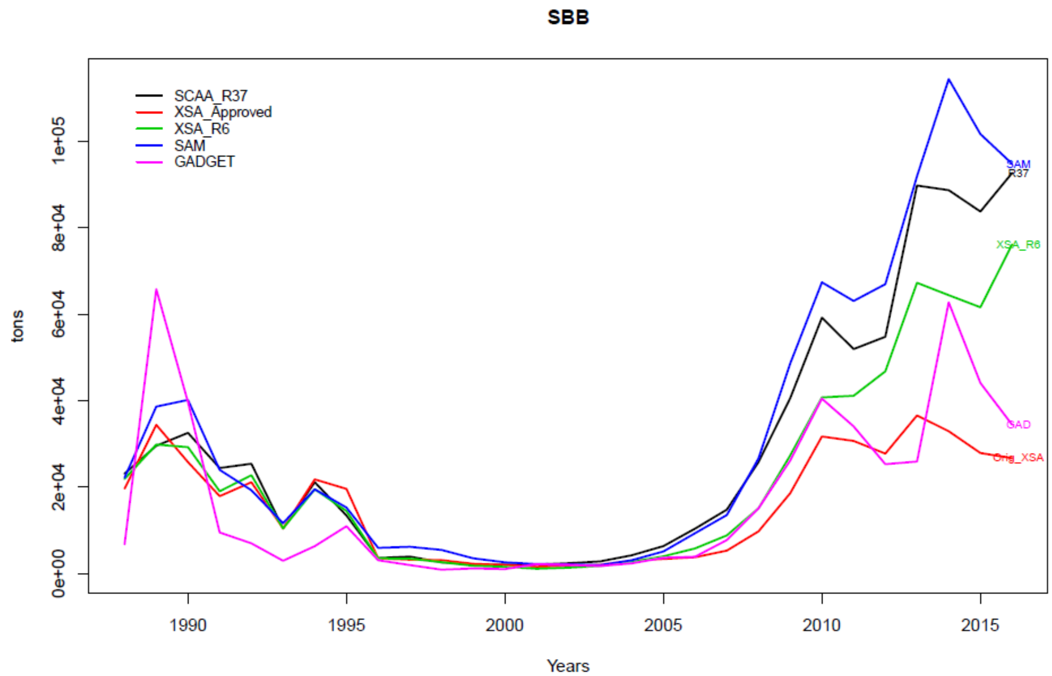


Fig. 25. Comparison between estimates of spawning stock biomass from SCAA (run 37), XSA (run 6), SAM (run 19) and GADGET models, as well as the last approved assessment (XSA).

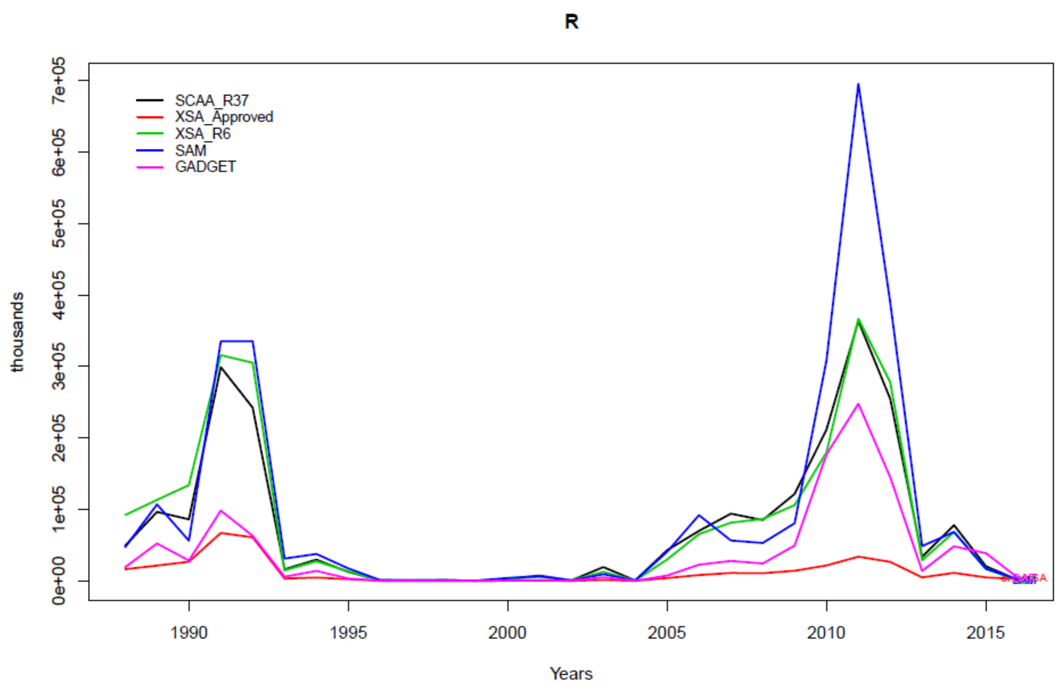


Fig. 26. Comparison between estimates of recruitment from SCAA (run 37), XSA (run 6), SAM (run 19) and GADGET models, as well as the last approved assessment (XSA).

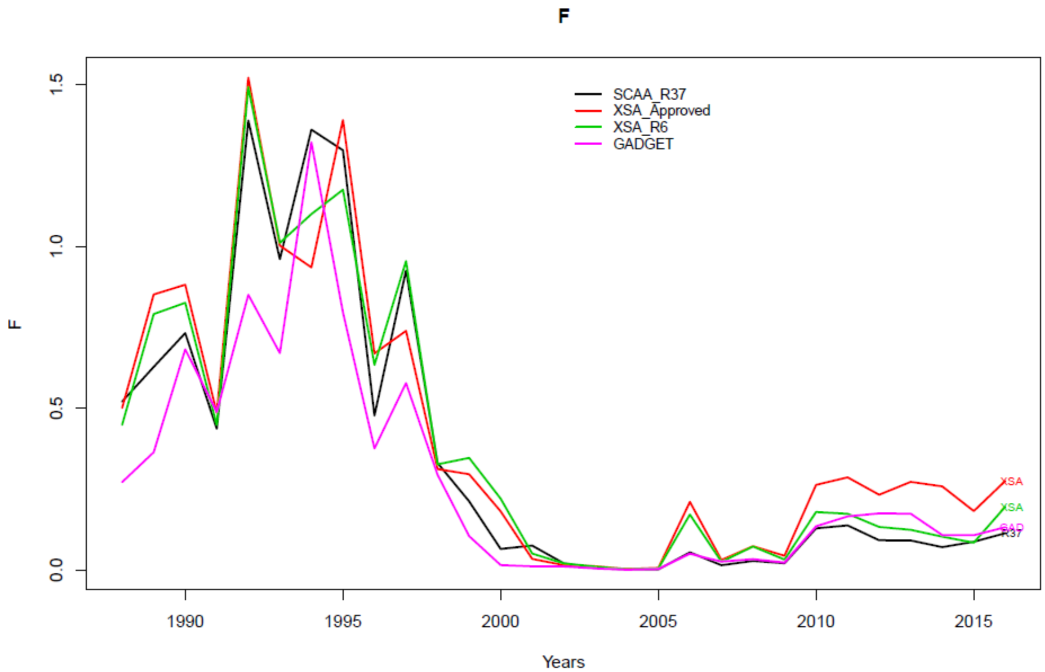
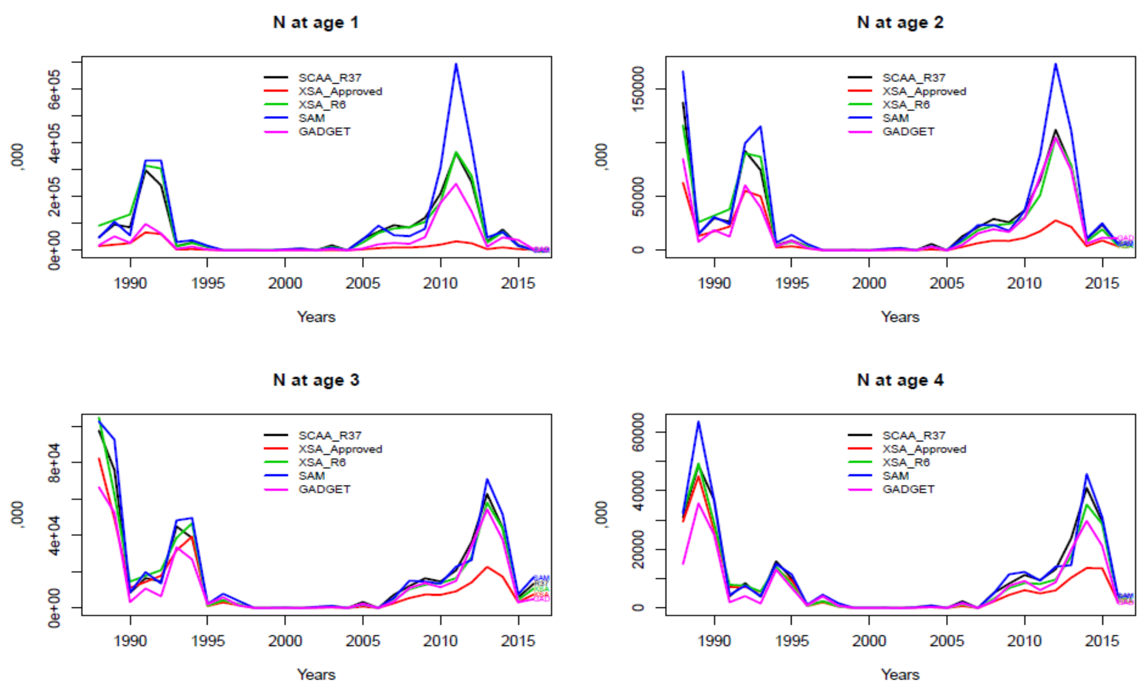


Fig. 27. Comparison between estimates of fishing mortality (F) from SCAA (run 37), XSA (run 6), SAM (run 19) and GADGET models, as well as the last approved assessment (XSA).



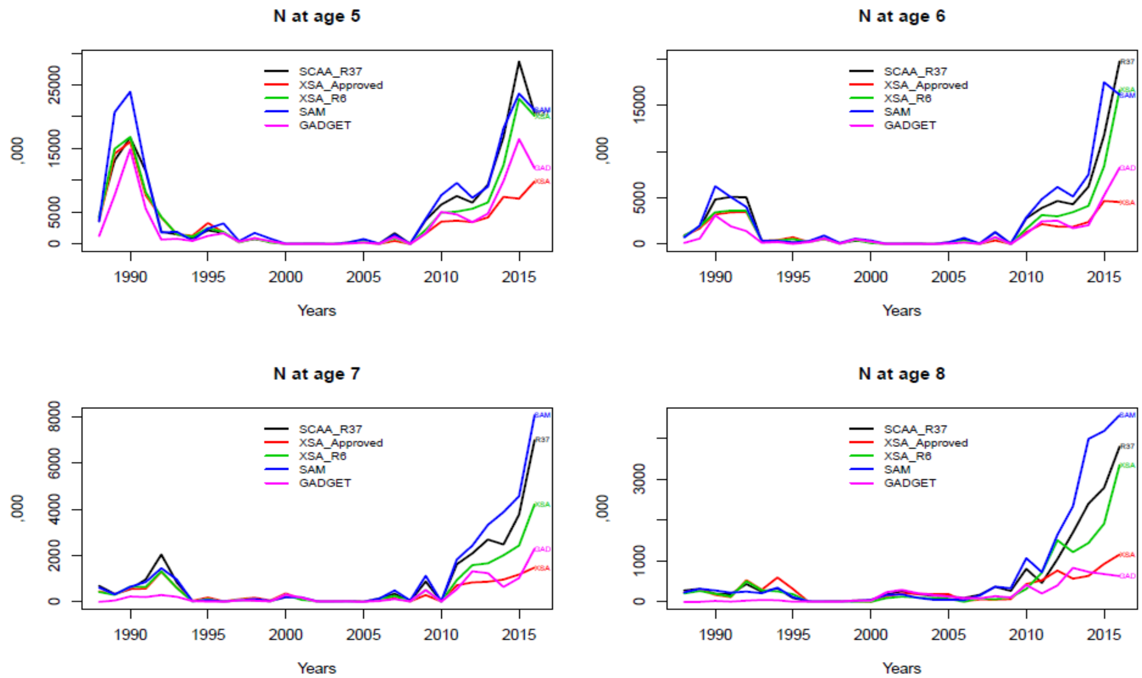


Fig. 28. Comparison between estimates of numbers at age from SCAA (run 37), XSA (run 6), SAM (run 19) and GADGET models, as well as the last approved assessment (XSA).

8. Projections

Due to time constraints it was not possible to consider projections for any of the models during the benchmark meeting: these will be considered during the June SC meeting.

9. Reference points

Two approaches to setting reference points were explored: 30% spawner per recruit from the SCAA and XSA models and the method used by ICES WKMSYREF4 (ICES, 2015) and WKMSYREF5 (ICES, 2017).

Calculation of $F_{30\%SSB/recruit}$

$F_{30\%SSB/recruit}$ was calculated for the Bayesian SCAA run 37, the previous accepted assessment (XSA 2017) and the final XSA run from the benchmark. Th results are presented in table 12

Table 12. Estimates of F_{lim} ($F_{30\%SSB/recruit}$ with 90% CI) for the Bayesian SCAA and two XSA assessments.

F_{lim}	5%	50%	95%
SCAA run37	0.256	0.274	0.293
XSA 2017 run	0.137	0.174	0.215
XSA final run	0.203	0.233	0.263

Effects of changes in Maturity and Stock weights-at-age were considered. Weight and maturity are quite variable with suggestion of density dependence in both biological variables. Applying equilibrium N at age (calculated by applying M=0.19 to an unfished population) to average weight and mat gives SSB per recruit of ca 6kg. However, taking account of variability in weight and maturity at results in large change in time of SSB per recruit. Variability in weight seems to contribute more than variability in maturity. $F_{30\%SSB/recruit}$ was calculated (for average and variable weight and maturity) for M=0.19 and the range of M vectors from the literature considered in the data meeting.

Application of ICES guidelines

Biological reference points for Cod 3M were calculated for this stock based on the ICES recommendations from WKMSYREF4 (ICES, 2015) and WKMSYREF5 (ICES, 2017). First, limit reference points for spawning stock biomass (B_{lim}) were defined based on the stock type identified in ICES guidelines. Then, for each of the stock types, limit fishing mortality (F_{lim}) were defined for different scenarios using EqSIM software.

The stock-recruitment relationship for this stock is shown in figure 29. The dynamic range of SSB goes from 2 to 90 thousand tonnes. The highest biomasses correspond to the last five years of the assessment, on which recruitments were among the lowest observed.

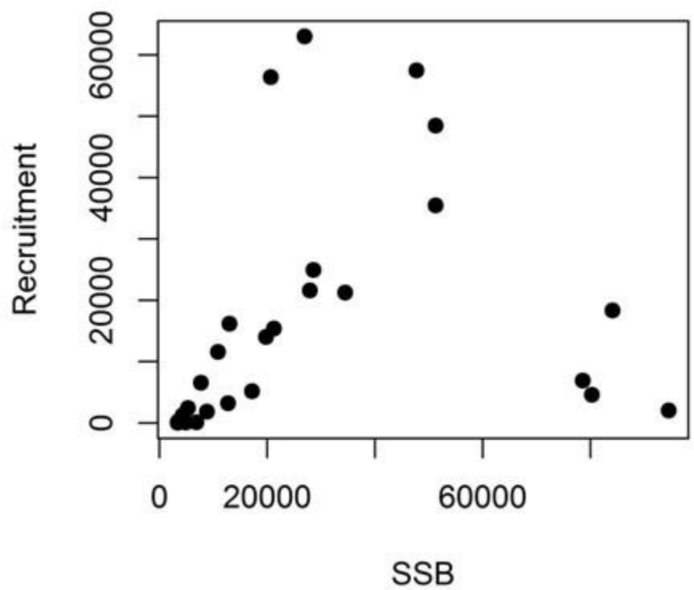


Fig. 29. S-R plot of NAFO COD 3M assessment considering M=0.19.

Following the ICES guidelines, S-R plot presented for NAFO Cod 3M stock would be considered of type 1: *spasmodic stocks – stocks with occasional large year classes* or type 2: *stocks with a wide dynamic range of SSB, and evidence that recruitment is or has been impaired*. Therefore, the calculation of the reference point B_{lim} and F_{lim} was calculated for each stock type in different scenarios.

For each stock type defined, the calculation of B_{lim} was different: as a type 1 stock, B_{lim} = 23 810 t (based on the lowest SSB where large recruitment is observed) and as a type 2 stock B_{lim}= 45 603 t (based on the change point in a segmented regression).

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}. Several scenarios were analyzed for each stock type in the calculation of the B_{lim} and F_{lim} values, such as changes in the number of years used for the biological parameters and the exploitation pattern.

SC Decision:

-A new B_{lim} will be estimated during the June SC meeting based on the results of the assessment approved during this meeting following the NAFO PA guidelines.

-A new F_{lim} will be estimated during the June SC meeting based on the results of the assessment approved during this meeting following the NAFO PA guidelines. In the Benchmark, no information was presented to change the current NAFO SC approved F_{lim} proxy (F30%SPR).

-Discussions about which temporal series of the input data is the most appropriate to calculate F_{lim} were carried out but no agreement was reached. This point is open for the next June SC meeting.

10. Discussion of possible operating model elements

The performance of Harvest Control Rules (HCRs) for this stock will be analyzed under a Management Strategy Evaluation MSE framework. Due to the simplicity of the interpretation of the model free HCRs, a review of all the model free HCRs published were compiled (SCR-Doc 18-02) and presented in order to understand what has been done previously in other species with their pros and cons. The model free HCRs can be based on survey data or on fishery dependent data and for some of them knowledge on life history parameters is required. The performance of each of the HCR is case specific and depends on the quality of the data available and the life history traits of the stock: short, medium or long-lived species. In the case of cod 3M data from the Canadian survey (1978-1985) and the EU survey (1988-2016) are available and it could be used as tuning series. For example, one of the HCRs reviewed was based on a survey index (the HCR 1.4 in the WD) and it was used in the initial Greenland Halibut MSE considering the slope of biomass index as indicator, and the same HCR was analyzed in a previous MSE for cod 3M. However, one of the concerns of a survey-based HCR is that in the case of cod 3M, only one survey is available and therefore if one-year survey data are missing then the HCR based on one survey could not be used.

11. Conclusions

SC considered numerous model formulations of single species population models (Bayesian XSA, Bayesian SCAA and SAM) as well a multispecies model based on GADGET (GADCAP). Model results and diagnostics were explored. Different formulations of the models in some cases gave very different results and often indicated lack of fit to the data. Analyses focused on the Bayesian XSA and SCAA. The final two runs of the Bayesian XSA and SCAA showed better fit to the survey data and results of the two models were similar to each other. The greater flexibility of the SCAA was considered to be an advantage over the XSA, making it a more powerful assessment tool. In addition, more testing was conducted of the SCAA during the benchmark than of the other models. Considering all of these issues SC **recommended** *a Bayesian SCAA with structure similar to run 37 to form the basis of the assessment for this stock in June 2018 pending the sensitivity analyses described below*.

Further sensitivity analyses will be conducted prior to the June SC meeting as follows:

- For the Bayesian XSA: retrospective analyses
- For the SCAA: prior median of recruitment of 45000, CV on prior for recruitment and abundance at age in the first year as 10, including a full set of diagnostics.

F_{lim} and B_{lim} will be updated in June but will have the same basis as at present.

12. Reviewer reports

Reports were provided by the three invited external reviewers and these are presented in Appendix IV.

13. Closing

The meeting Chair thanked the Secretariat for their support prior to and during the Benchmark. In addition, gratitude was expressed to Instituto Português do Mar e da Atmosfera for hosting the meeting and to colleagues who helped facilitate meeting arrangements. The participation and contributions of highly-qualified external reviewers was noted as a strength of this meeting, and the Chair thanked the reviewers for their contributions.

The Chair noted that although the quality of the work conducted using the models was very good, it was impossible to fully explore any of them due to lack of time, as the possibilities are almost endless. This fact was more evident in the case of the SAM as the expert who led this work was unable to attend the meeting in person and contributed via WebEx.

The Chair and all meeting participants expressed sincere gratitude for the considerable volume of work prepared prior to and during the Benchmark, leading to an improved assessment method for Div. 3M cod. The Chair commended all of those involved, with special recognition to Diana González-Troncoso (EU-Spain), who,

as Designated Expert, led a significant portion of this work and produced many re-analyses during the course of the week.

The Chair wished all participants a pleasant return journey and the meeting was closed at 1400h on Friday April 13th.

14. References

Pérez-Rodríguez, A., Howell, D., Casas, J.M., Saborido-Rey, F., and Ávila-de Melo, A. 2017. Dynamic of the Flemish Cap commercial stocks: use of a gadget multispecies model to determine the relevance and synergies between predation, recruitment and fishing. Can. J. Fish. Aquat. Sci. 74: 582-597.

APPENDIX I- PROVISIONAL AGENDA

NAFO Scientific Council Flemish Cap (NAFO Div. 3M) cod stock benchmark assessment meeting
09–13 April 2018
Lisbon, Portugal

Provisional Agenda

1. **Opening - Introductions, Meeting Arrangements**
2. **Appointment of Rapporteur**
3. **Adoption of Agenda**
4. **Introductory Presentations**

4.1 Workshop 2017 results

4.2 Assessment 2017

5. **Assessment Input Data.**

5.1 Consider the variability in the biological parameters (i.e. age at maturity, mean weights, etc.) observed in recent years, agree on an approach to be applied in stock assessment.

5.2 Explore alternative values on natural mortality

- Estimation of M from biological models
- Estimation of M via Gadget
- Other (including Bayesian estimates)

5.3 Aging and Age/Length Keys (ALKs): Results of the Nov. 2017 Workshop. Agree on an approach to be applied in stock assessment

6. **Potential assessment models**

6.1 Bayesian XSA

6.2 Bayesian SCAA

6.3 SAM

6.4 Others (GADGET)

7. **Model parameters**

Further exploration of parameters for all potential assessment models, to include (where relevant):

- Bayesian Priors
- the possibility of expanding the current plus group.
- plus group abundance estimation in the stock assessment (VPA or dynamic pool)
- the possibility of first ages catchability depending on abundance.

8. Projections

8.1 Projections inputs.

8.2 Further investigate including variability in the biological parameters (MWS, MWC, MO) used in the projections.

9. Reference points

9.1 Review of F_{lim}

9.2 Review of B_{lim}

10. If time permits, discuss elements of possible operating model variants to be fit, projection specifications, observation models for future generated data and guidance for development of possible HCRs to use in the MSE process.

11. Drafting of Summary conclusions

12. Reviewer reports

APPENDIX II: LIST OF DOCUMENTATION

SCR Documents

Document	Serial Number	Author (s)	Title
SCR Doc. 18-001	N6778	F. Gonzalez-Costas, D. Gonzalez-Troncoso, A. Ávila de Melo and R. Alpoim	3M cod assessment input data
SCR Doc. 18-002	N6779	Marga Andrés, Dorleta Garcia, Agurtzane Urtizberea	Model-free HCR: literature review for NAFO Cod 3M
SCR Doc. 18-003	N6780	F. Gonzalez-Costas and D. Gonzalez-Troncoso	Cod 3M Natural Mortality
SCR Doc. 18-004	N6781	Thomas Brunel	Exploratory assessment of the cod 3M stock using SAM
SCR Doc. 18-029	N6815	Diana González-Troncoso, Carmen Fernández and Fernando González-Costas	Bayesian XSA model for the 3M cod
SCR Doc. 18-030	N6816	Diana González-Troncoso, Carmen Fernández and Fernando González-Costas	Bayesian SCAA model for the 3M cod

SCS Documents

Document	Serial Number	Author (s)	Title
SCS Doc. 18/18	N6849	NAFO	Cod Benchmark Report, 9 – 13 April 2018

APPENDIX III: PARTICIPANT LIST

CHAIR	
Healey, Brian (SC Chair)	Science Branch, Fisheries & Oceans Canada, P.O. Box 5667, 80 East White Hills Road, St. John's, NL A1C 5X1 Tel: +1 709-772-8674 – Email: brian.healey@dfo-mpo.gc.ca
CANADA	
Morgan, Joanne	Science Branch, Fisheries & Oceans Canada, P.O. Box 5667, 80 East White Hills Road, St. John's, NL A1C 5X1 Phone: +709-772-2261 - E-mail: joanne.morgan@dfo-mpo.gc.ca
Regular, Paul	Science Branch, Fisheries & Oceans Canada, P.O. Box 5667, 80 East White Hills Road, St. John's, NL A1C 5X1 Email: paul.regular@dfo-mpo.gc.ca
EUROPEAN UNION (EU)	
Alpoim, Ricardo	Ricardo. Instituto Portugues do Mar e da Atmosfera, Rua Alfredo Magalhães Ramalho, nº6, 1495-006 Lisboa, Portugal Tel: +351 213 02 70 00 – Email: ralpoim@ipma.pt
Avila de Melo, António	Instituto Portugues do Mar e da Atmosfera (I.P.M.A), Rua Alfredo Magalhães Ramalho, nº6, 1495-006 Lisbon, Portugal Tel: +351 213 02 70 00 – Email: amelo@ipma.pt
Brunel, Thomas (via WebEx)	WUR Email: thomas.brunel@wur.nl
De Oliveira, Jose	CEFAS Email: jose.deoliveira@cefas.co.uk
González-Troncoso, Diana	Instituto Español de Oceanografía (IEO), Aptdo 1552, E-36280 Vigo, Spain Tel: +34 986 49 21 11 – Email: diana.gonzalez@ieo.es
González Costas, Fernando	Instituto Español de Oceanografía (IEO), Aptdo 1552, E-36280 Vigo, Spain Tel: +34 986 49 22 39 – Email: fernando.gonzalez@ieo.es
Pérez Rodriguez, Alfonso	WUR Email: alfonso.perezrodriguez@wur.nl
Rodriguez-Alfaro, Sebastian	International Affairs, European Commission, Scientific Advice and Data Collection, Directorate General for Fisheries and Maritime Affairs (DG MARE.B.3), Rue Joseph II, 99, 1000 Brussels, Belgium Email: Sebastian.RODRIGUEZ-ALFARO@ec.europa.eu
Urtizberea, Agurtzane	AZTI. Herrera Kaia, Portualdea z/g – 20110 Pasaia (Gipuzkoa), Spain. Email: aurtizberea@azti.es
INVITED EXTERNAL REVIEWERS	
Fernandez, Carmen	Email: cferllxx@gmail.com
Ianelli, James	NOAA Fisheries External Expert Email: jim.ianelli@noaa.gov
Palmer, Michael	NOAA Fisheries External Expert Email: Michael.palmer@noaa.gov
OBSERVERS	
Quirijns, Floor	ICFA. Sorbier, France. Tel: +337 835 91 476 – Email: fquirijns@pelagicfish.eu
NAFO SECRETARIAT	
Blasdale, Thomas	Scientific Council Coordinator Email: tblasdale@nafo.int
Bell MacCallum, Dayna	Scientific Information Administrator Email: dbell@nafo.int

APPENDIX IV: REVIEWERS' REPORTS

Reviewer's report

NAFO Scientific Council Flemish Cap (NAFO Div. 3M) cod stock benchmark assessment meeting
09–13 April 2018
Lisbon, Portugal

Jim Ianelli, jim.ianelli@noaa.gov
Alaska Fisheries Science Center
National Marine Fisheries Service, NOAA

1. Introductions

The benchmark meeting was conducted from 9am-6pm through the week and was with very active participation and commentary. The meeting was generally informal and run as a workshop with well-coordinated requests to analysts who were responsive in providing updates. Data issues were reviewed, and the group was reminded of decisions from earlier web-based teleconferences and other discussions. Some 37 different models for the Statistical Catch-at-Age (SCAA) model along with a number of GADGET, SAM, and Bayesian XSA, model configurations. The terms of reference for reviewers are provided in Annex 1.

2. Potential assessment models

Specifications of model parameters including priors, general structure, and model predictions of survey data by age were discussed during all of the model presentations. Similarly, projections and reference points that may serve to assist in providing advice were also considered and are presented in section 7 below. Discussions of this latter topic were limited primarily to how the assessment methods can be fed into estimates of limit and other reference points.

The exploration of alternative values on natural mortality (M) consumed a significant amount of time during the week. The history of past values for M were reviewed and the biological basis for age-specific values was provided. These were compared with the GADGET multi-species model results. The GADGET model included three species groups (some with subgroups by size, and/or maturity state and sex): cod, shrimp, and redfish. A number of alternative catch-curve methods were examined to help with selecting values for the age-specific "residual M" (or M1 values in MSVPA parlance). It seems that these values have changed as the project evolved. The latest approach included an evaluation of the relative GADGET fits (comparing "negative log- likelihood scores") and longevity estimates, in addition to the results from the catch-curve analyses. The result was quite high (0.35) and this stimulated much discussion. In the end, the workgroup seemed to reach a consensus (as I perceived it) to use an age-specific vector to account for higher mortality rates for younger ages but use values (as priors) closer to the biologically-based estimates (~0.24 for cod greater than about age 4). The sensitivity to time-varying M, as evaluated in the SAM model below, seemed to be relatively small (results from the "M matrix" and "M vector were similar).

Regarding general data collection issues, the group discussed age-determination workshop results and given the situation, it was recommended that age-length keys (ALKs) be collected for both the survey and for the fisheries. It seems to me that reasonable **recommendation would be to evaluate the impact using ALKs from the fishery for the fishery versus just using the survey ALKs.**

3. Bayesian XSA

These were presented and explained as a way to incorporate uncertainty into the XSA modeling framework, specifically with the specification of priors on survivors in the terminal year and other aspects. This provided a baseline reference case and seems to have resolved some of the issues related to "shrinkage" (a common concern in XSA) and other problems (yet catch-at-age is still assumed to be known exactly).

On Thursday evening the Bayesian XSA (BXSA) runs were configured for direct comparisons with the SCAA and presented on the last day. Results indicated the both the SCAA and BXSA had good retrospective patterns (for the 5-years evaluated). However, the estimated posteriors for the survey catchabilities were quite different (BXSA was almost double the values for SCAA). This pattern was also seen in the relative natural mortalities (generally lower values for the BXSA). These results are problematic because such differences will have repercussions for reference points and the estimated absolute stock size. It seems that the high values of the posteriors on the survey CVs effectively ignores those data, resulting in values of M that are close to the prior (as the data on the survey have less influence) and that the survey catchabilities for the BXSA were much higher to accommodate the lower value of M. The treatment of age 1s and age 7 indices which each had a value of "0" in the series, contributed to the high observation variance estimates for those series.

Should the BXSA be carried forward at the June meeting, my **recommendation is to fix the observation errors for the surveys at values similar to those fixed in the SCAA.** Additionally, the treatment of zeros in the indices should also be done consistently. The remainder of this section of the report deals with evaluations of the different models evaluated.

4. Bayesian SCAA

This model was developed to provide a number of the features available in SAM as well as standard forward projection assessment models. We were provided detailed presentations on the model and given access to the R code and specifically the "JAGS" code. This helped considerably since the code was quite readable and helped understanding the forward projection aspect of this model and how observation equations and prior distributions are specified.

A difficulty (it seems) with this approach is in understanding how different components of the posterior distribution interacted during the MCMC sampling process (components of the posterior apparently are unavailable for examinations). Initial runs with this software were clearly having issues as the observation errors for a number of the age-specific indices were going to very high values (and hence ignoring the data). It was noted that this could be related to the fact that the analysts added a small constant to the observed zero values. It was determined that NA's (i.e. missing values) could be used instead (for sensitivity). The group noted that doing so effectively ignores the fact that a zero value was an actual observation…). This result prompted me to examine the data in a standard stock synthesis-like model which treats the survey data as a bulk biomass and proportions at age in each year rather than as separate time series of age-specific abundances (see section below on "other alternative models" for further details).

The group made a number of suggestions for modifications to help improve how the SCAA performed. This involved two main aspects: fixing the values for the observation-errors rather than trying to estimate them and to try to estimate *M,* with catchabilities for ages 3-8 fixed to a common estimated parameter (to avoid a "dome-shaped" survey selection pattern). For the latter, applying the age-specific vectors from the biological analyses were also considered useful. These changes to the code were completed and run between Tuesday and Wednesday of the meeting. After some other corrections the model seemed to perform reasonably well and in line with expectations ("model 37"). At the close of the meeting, some sensitivities from this model were recommended. I anticipate that these will result in being relatively minor effects and the group's recommendation to use this framework will be warranted.

5. SAM

Tom Brunel presented results from preliminary configurations of a SAM assessment model (Nielsen et al. (2016). The observation and process errors as estimated were quite high and some suggestions for regrouping variances. Specifically, the aggregation of the observation error into two groups: estimates for age 1 and separate for ages 2-8 should be changed to age 1 and 2, and 3-8 since it seemed that the variances for ages 1 and 2 area both high and 3-8 seemed to be relatively lower (and similar).

A number of modifications were suggested and other alternatives were tried. On Thursday the group made a specific request to have a SAM model run that could be compared with other methods. These were accommodated and discussed for comparisons with other models.

My general impression and thoughts on the applicability of the SAM framework for 3m cod stock are as follows. It is clear to me that this model is appropriate when data collection systems are consistent and well conducted over the entire time series. In particular, age-determination approaches should be applied to fishery and survey data appropriately. However, it appears that for the case of 3m cod, the fact that landings have varied considerably over time both in magnitude and by gear types there are serious issues. In particular, the sampling among these fleets and gear types has varied over time and consequently the age composition data have consistency problems and likely result in unreasonably high estimates of both process and observation errors. For these reasons, I find the SAM approach inappropriate for this stock at this time. The estimated observation errors and process errors were difficult to justify as they result in essentially ignoring portions of the data without the ability to structurally account for changes in sampling over time. Furthermore, the process errors (as observed by changes in the population numbers at age) were large and patterned in a way that suggests structural model mis-specification (perhaps due to changes in predation or other emigration factors).

6. Other alternative models

During the week, a "standard" statistical catch-at-age model similar to stock synthesis was constructed with the available data used (see Annex 2 for some details). This was done to help me evaluate the data and patterns from the fishery and issues related to the survey and also evaluate the pattern. This approach had some slight differences than the SCAA presented during the week, namely the following features/assumptions were highlighted:

- Constant maturity and SSB weight-at-age (but variable for survey and fishery catch biomass)
- Fit to catch and survey biomass and proportions-at-age for both gears (avoids issues of "zeros" in the age-specific indices)
- Can be easily extended to split gear types
- Allows sampling errors to be used (and annual aspects of statistical fitting, e.g., if age compositions are poor or different in some years), resolves problem when fishery was closed or catches were low…
- Simple, easy, and fast to implement (and get approximate posterior distributions in less than 30 seconds for 250 parameters…).

Whereas the best software in the world is the one you know how to use…it seems that a number of the issues that arose in dealing with BXSA, SCAA, and SAM can be easily resolved with a simpler more flexible forward projection approach such as that presented in Annex 2. Such models require inputs related to data uncertainty, and this can be a concern. However, the added flexibility of allowing the variance (or CVs) to vary over time would seem to be an important added feature for 3m cod assessments which appears to have data issues that have varied over time. Furthermore, the approach in Annex 2 can treat data as they are available without having to make modifications or massages to the data directly. A disadvantage of the Annex 2 approach is that there are fewer process errors (compared to SAM) that are included and the ones that are included (i.e., recruitment deviations) are not treated as true random effects.

7. Projection and reference point considerations

a) Review of F_{lim} and B_{lim}

The analysts provided an overview of the ICES approach for computing these reference points. It was noted that while formally part of the ICES process (software wise at least), often the scientists involved with deviate from specifications based on expert knowledge and issues apparent in the stock-recruit relationship and reference point calculations. Relative to F_{lim}, the group noted that the current SPR rate of 30% might result in a proxy for Fmsy that is too high (in many settings a more conservative proxy is used for groundfish species). There seemed to be some justification for this SPR rate for cod stocks (noting that it takes into account growth, maturation, and selectivity in the calculations).

Relative to the estimation of F_{lim}, a spreadsheet and other software was provided (by me) to evaluate the sensitivity of the time-varying body mass and maturity-at-age assumed for the stock (see Annex 3 for a slide deck of these evaluations).

8. Management strategy evaluation considerations

The terms of reference noted that if time permits, the group could discuss elements of possible operating model variants to be fit, projection specifications, observation models for future generated data and guidance for development of possible HCRs to use in the MSE process. This occurred and issues on modeling the expected changes in maturity and body mass at age would be challenging. There are some projects apparently underway to examine hypotheses of these changes as a function of water temperatures and other environmental variables. Relative to the variability in body weight-at-age, the model described in Annex 3 might be one approach to account for cohort and year-effect variability for operating model testing.

The presentation reviewing some 10 different HCR systems was useful, but quite complex and (I suspect) had some overlapping characteristics. I think advice going forward with respect to MSE work is to be aware of the elements and principles that I think are worth remembering:

- Multiple objectives, more than simply fishing at F_{MSY}.
- Uncertainty must be characterized, e.g., are future data included in a feedback form for setting TACs? If future data collection system are not taken into account, then evaluation of risks may be less well determined.
- Stakeholder input is required, objectives elicited iteratively (i.e., a scientific publication of a simulation analysis in and of itself is not a full MSE
- Tradeoffs are an explicit part of the process (i.e., socio economic considerations versus maintaining spawning biomass above lower limit.

9. Concluding remarks

The process for considering aspects of the modeling approach included healthy discussions of data and related issues. Some models considered resulted in variance estimation that effectively down-weighted the available data which, if the model was structured perfectly (it most certainly is not), would be acceptable. As such, I found the application of SAM and other models in which the variance estimates of the data (observation errors) unacceptable since they effectively ignored data. Nonetheless, a surprising number of models and configurations provided quite similar results indicating that the high levels of catch in the early part of the series likely governs much of the stock's population scale (especially when natural mortality was pre-specified and survey catchabilities were constrained).

The working group interactions were excellent, and I was grateful for the exceptional support provided by the NAFO Secretariat. I think their contributions were important for facilitating the smooth running of the meeting.

3M Cod Benchmark April 9-13, 2018. Report from external reviewer Carmen Fernández.

Comments on process:

I felt the benchmark process for the 3M cod stock worked well. The working atmosphere was very good, it was clear that all people involved were working to provide a positive contribution to the process. An enormous amount of work was conducted by the designated stock assessor, the meeting was well chaired and the venue facilities were also good. As a reviewer, I was given plenty of opportunity to interact and exchange views with the rest of the benchmark participants. I felt the benchmark process achieved good progress in the assessment of this cod stock and was successful as a whole, and I endorse the outcomes of the meeting.

The reviewers were invited to be part of the process already during the preparatory scientific work for the meeting; in particular, we were invited to participate in preparatory webexes. I feel this early involvement facilitates the reviewer's task, as it helps to gain background and understanding in advance of the main benchmark meeting.

I think it would be helpful if more of the technical material could be made available to the reviewers in advance of the meeting. For most main pieces of work that the benchmark dealt with, there was no scientific document that the reviewers could study before the start of the meeting. I completely understand the time pressures scientists are under, but I think that having the main pieces of technical work available in advance of the meeting would allow for increased quality of the reviews.

The benchmark meeting was strongly focused on assessing the stock (i.e. assessing the historic period). There was not much time to review reference points or projection settings in depth. This is unsurprising. In my experience from other benchmark processes, which comes mostly from ICES, stock assessments always raise substantial issues that end up taking most of the meeting's time. However, it is also important to keep in mind that the catch advice for stocks depends strongly on reference points and projection settings, not just on the stock assessments. Therefore, I believe that careful consideration and review of appropriate settings for reference points and projections is necessary before catch advice is next provided for this cod stock.

It was good to hear of the willingness of the fishing industry to cooperate with the scientific process. I think it is important to keep in mind their offer to provide data that could be used to improve the assessment of this stock and to try and find ways to achieve a useful cooperation.

Comments on technical aspects:

<u>Trends in biological parameters:</u>

The available biological information indicates there have been changes in cod weights and maturity over the 30-year assessment period. An increasing trend in the weights-at-age and maturity-at-age can be observed from the mid or late 1990s to the mid or late 2000s, followed by a decreasing trend in the last decade. In view of these trends, careful consideration should be given to the period of years to be used in the calculation of reference points and stock projections.

Exploratory work trying to model the observed trends in stock biology, and aiming to include the uncertainty arising from these changes into the projections, was attempted during the cod workshop held in Vigo in 2017, but the outcomes were not considered satisfactory. Additional analysis (e.g. fitting biological models by cohort rather than by year, or including density-dependence or environmental variables in the model) could be attempted in the context of the planned MSE for this stock, but I have no specific modelling suggestions in this regard.

In addition to the trends over time, substantial interannual variability (between consecutive years) can also be seen in the current estimates of biological parameters. The estimation of most biological parameters has been done separately from year to year. This may introduce excessive interannual variability in the results due to sampling variability and it is worth considering whether there could be better options to estimate these parameters (e.g. by smoothing over time or within cohorts).

<u>Appropriate specification of natural mortality</u>

The issue of how to treat M in this assessment was extensively discussed at the meeting. The discussion involved whether M should be treated as a single value for all years and ages, as an M-at-age vector constant over time, or as an M(y,a) matrix. It was also discussed whether M should be treated as a fixed input to the stock assessment model or be estimated within the stock assessment model. Relevant work on potential specification of an M(y,a) matrix was presented from a multispecies model in the Flemish Cap (GadCap), and a separate proposal to estimate a time-varying M within the stock assessment model was also presented and discussed.

After considerable discussion and examination of results during the meeting, the benchmark decided to include a vector of M-at-age, constant over time, in the stock assessment. The M-at-age vector was assigned a prior distribution centered on a vector of M-at-age derived from results from various life-history methods and the GadCap multispecies model, and a small CV (15%). The reason for choosing a small prior CV was so that the stock assessment model could only have limited flexibility to revise the value of M-at-age away from the prior; as estimation of M-at-age in the stock assessment model will likely be confounded with aspects such as commercial fleet or survey selectivity-at-age, this seems like an appropriate approach at this time.

Further consideration of issues around M could be considered during the upcoming MSE.

<u>Appropriate assessment model: Bayesian XSA, Bayesian SCAA, SAM; multispecies GadCap</u>

The three options considered for the assessment of this stock were a Bayesian XSA model similar to that used to assess this stock since 2008, a Bayesian SCAA (statistical catch-at-age) model specifically developed for this stock, and a SAM model. The multispecies GadCap model seems very useful in the context of progressing towards a multispecies and ecosystem approach, and may be used in the upcoming MSE, but was not considered as an option for the stock assessment of 3M cod at this point.

Most of the meeting's time was spent exploring and discussing the Bayesian SCAA model. The expert working with the SAM model was not at the meeting (although available remotely frequently during the meeting), and this made holding in-depth discussions on the SAM model more difficult. A particular feature of SAM is that it includes process error. From the runs conducted, the magnitude of the estimated process error appears to be large for this cod stock and, if SAM was used as the main stock assessment model, consideration should be given to the best way for dealing with it in short-term forecast and risk computation.

Concerning the Bayesian XSA model, substantial sensitivity issues on the results of the assessment for recent years were detected by the NAFO Scientific Council in their June 2017 meeting. It was clear at the benchmark that if the Bayesian XSA were to continue to be the assessment model for this stock, an approach to deal with these issues would need to be agreed.

Two advantages of the SCAA over the XSA model are that it does not need to follow the observed catch numbers-at-age exactly and that it uses a forwards-projecting population dynamics model, consistent with that used in short-term projections or with a potential Operating Model for the MSE.

The Bayesian SCAA model has many different parameters that can interact with each other in ways that are not easy to uncover. Setting appropriate prior distributions on the large amount of parameters in the Bayesian SCAA model is a difficult exercise. Many alternative model settings had been explored by the stock assessor in advance of the benchmark. The results of that work indicated that assessment results for recent years are sensitive to the specification of model settings; particularly the estimated abundance of fish of ages 6, 7 and 8+ in recent years can differ substantially depending on model settings. When the model was allowed to estimate the CVs of the observation equations for the (log-Normal) catch numbers-at-age and survey abundance indices-at-age, it estimated large CVs in all cases and some really large CVs for some ages. This essentially meant the model fit was effectively dismissing part of the input data, which was not considered appropriate by the workshop.

Given the difficulty in finding an appropriate configuration for the Bayesian SCAA model only from examination of model fits (residuals and Deviance Information Criterion), the meeting decided to let the model settings be largely guided by biological or fishery understanding. This led to proposing the following model settings:

- Survey q, make the following age groups: 1, 2, 3, 4+
- Power model for catchability of EU survey: exponent only at age 1
- Estimate M-at-age, constant over time, with prior median derived from life-history and GadCap and prior CV=15%

- Fix CVs for observation equations of catch-at-age and survey abundance indices-at-age [at 20% and 30%, respectively]
- Fix CVs for interannual changes in the fishery selectivity-at-age [at 20%]
- Remove age 1 from the catch-at-age matrix and set F of age 1 equal to 0
- Replace the 0s in the observed catch-at-age or survey abundance indices-at-age matrices by missing values.
- At a later stage (at the end of the meeting), it was also considered appropriate to increase the median of the prior distribution of annual recruitment to 45000 and to increase the CV of the prior distribution on annual recruitment and population abundance-at-age in the first assessment year to 10.

These choices appear to me as reasonable given the issues encountered with the stock assessment.

A proposal for similar model settings for the SAM model was made as follows:

- q of the survey grouped for ages 1, 2, 3, 4+
- Power model for catchability of EU survey: exponent only at age 1
- Vector of M(a), constant over time. The value used is very close to the posterior median from the Bayesian SCAA run just described.
- CV of observation equation for EU survey, grouped by ages: 1-2 and 3+.
- CV of observation equation for catch numbers-at-age: 1-2, 3-6, 7-8+
- Independence between observations
- CV of interannual changes in F(y,a), grouped by ages: 1 and 2+; with independence across ages
- CV of process error, grouped by ages: 1, 2+
- Observed catch-at-age and survey abundance indices-at-age matrices: no changes to these matrices (SAM already treats 0 as missing values)

Similar model settings were also examined for the Bayesian XSA model:

- Survey q, make the following age groups: 1, 2, 3, 4+
- Power model for catchability of EU survey: exponent only at age 1
- Estimate M-at-age, constant over time, with prior median derived from life-history and GadCap results and prior CV=15%
- Replace 0s in survey abundance indices-at-age by missing values.
- Prior on survivors: base the prior median on a recruitment value of 45000 and set the prior CV to 10.

With stock assessment runs configured as noted above and focusing attention on the Bayesian XSA and SCAA models, the stock assessment results were found to be similar with both models. In these circumstances, the meeting participants preferred to use the SCAA model given its greater flexibility over the XSA model, which allows it to be more in line with the way forward projections are performed. I consider the meeting's decision appropriate.

Reference points

As noted above, there was not sufficient time to explore this in depth at the benchmark and some aspects (in particular, the time period(s) to be used to calculate reference points and projections) need to be carefully considered before next providing catch advice for the stock.

MSE

An extensive presentation of potential HCR types that could be considered in the upcoming MSE was provided. Aspects of this cod stock that raised substantial discussion during the benchmark, such as how best to treat M, the estimation of biological parameters and the impact of selectivity, could be further considered within the MSE process.

External Expert Report:

Michael Palmer

Northeast Fisheries Science Center

National Marine Fisheries Service

National Oceanic and Atmospheric Administration

michael.palmer@noaa.gov

Overview

The Scientific Council (SC) of the Northwest Atlantic Fisheries Organization (NAFO) conducted a benchmark assessment of the Flemish Cap (NAFO Div. 3M) cod stock. Several webinar meetings were conducted prior to the actual benchmark meeting to discuss assessment model data inputs and refine the scope of the ensuing benchmark. The benchmark meeting took place in Lisbon, Portugal from April 9-13, 2018, with daily meeting times generally running from 9 AM to 6 PM, local time. The terms of reference for the benchmark process and its review are listed in Appendices 1 and 2, respectively. The meeting format was more similar to an assessment workshop meeting rather than a formal assessment review in that discussions were informal, and model configurations were revised extensively throughout the meeting.

Process

Benchmark processes are useful for advancing fisheries stock assessments. Among other things benchmarks allow for the injection of new ideas, further data exploration, incorporation of new data, and exploration of new modelling approaches. The newly implemented NAFO benchmark process is positive step towards advancing the fisheries science used to manage fishery resources in the NAFO governance areas.

The benchmark meeting of the Flemish Cap (NAFO Div. 3M) cod stock was a combination working group and benchmark review process. Many, but not all, of the outstanding data issues had been addressed during the webinars leading up to the benchmark meeting (handling of plus groups, years to include in the assessment model). Considerable time was spent during the benchmark meeting discussing remaining data and/or model configuration questions such as whether domed shaped selectivity was reasonable for the European Union (EU) survey or whether natural mortality (M) could be reliably estimated within the assessment model. Given the confounding nature of some of these issues, combined with the lack of data to definitely evaluate, these types of questions would be best explored further outside the actual benchmark meeting.

The Northeast United States benchmark assessment process (https://www.nefsc.noaa.gov/saw/) utilizes a working group to prepare the assessment, which is then reviewed by external experts through a formal peer review process. Data evaluation and major questions around model configuration and selection are addressed by the working group through a series of working group meetings that are held in advance of the actual benchmark review. The benchmark review is primarily focused on whether the working groups preferred assessment approach is scientifically defensible and constitutes the 'best available science' with which to craft management advice. This type of formalized peer review may not work well for the NAFO assessment process, but additional time spent with pre-meeting working group discussions may allow for a more focused benchmark meeting. One additional option would be to scale back on the terms of reference for the benchmark meeting.

An area of concern was the amount of model changes and number of model sensitivities that were explored during the meeting week. While the assessment leads should be commended for their willingness to rerun the many model configurations that were requested, I'm apprehensive about model selection decisions based on model changes which are hastily performed without the proper time to evaluate whether the reconfigurations had been properly structured, and without adequate review of the ensuing model diagnostics.

Other suggestions or comments on the process include:

- It would have been helpful for the externals to have been provided with additional background documentation on survey design and protocols and commercial sampling methods in advance (e.g., list of appropriate background SCRs for the uninformed external experts).

- Investigate a better file sharing mechanism (compared to SharePoint or Dropbox) that would allow working group members to more easily share data and files during the meeting
- It was difficult to keep track of the various model sensitivities and the parameter changes between sensitivity runs. Having a master list of model variations and a summary of the major configuration, diagnostic, and result differences would have been helpful.
- The data input file that was loaded to the SharePoint site pre-meeting was not the same data inputs used in the models presented during the meeting. Efforts to ensure that the data inputs provided to the externals were the most recent versions would be helpful.
- Industry engagement in the meeting was encouraging. It would also have been helpful to have had direct fleet knowledge (captain/skipper perspective) available at the meeting (or at a pre-meeting webinar) to inform the working groups interpretation of catch data.
- The work ethic of all participants, but in particular the assessment leads, was tremendous. Additionally, I found the group to be exceptionally welcoming and the Portuguese hosts were extremely gracious and provided a good working facility for the meeting.

Data inputs

There were initial concerns about the variability and trends in the weights-at-age and maturity-at-age data (i.e., were these patterns a product of sampling noise rather than actual signal?). However, based on the information provided by local experts on sampling density, and the confirmatory work done on maturity using commercial data, it seems more likely that there are strong signals in the data. There were however some odd values early in the maturity ogive where full maturity was never reached, even in the plus group. These patterns were never fully explained during the benchmark meeting that perhaps could be further explored, or at least, a more full explanation provided.

Regardless of the sampling intensity of the biological samples, it may be worth exploring approaches which use model-type smoothers (GLM, GAMs) to fit the weight-at-age and maturity data where year and cohort effects can be incorporated and possibly better account for density dependence and/or environmental effects. This could be particularly useful when developing reference points and projections as evidenced by the large influence of the weights-at-age on the fishing mortality reference point (Ianelli, AFSC/NOAA).

There was considerable discussion during the benchmark meeting about the absolute scale of the population coming from the various modelling approaches. In some of the earlier runs of the Bayesian XSA model, the survey catchability values were well above 1.0 which brought the question of whether this was reasonable? Based on the information provided by the local experts it did not seem that were had been any external corroboration of EU survey catchability by length or age to help bound what catchability values are reasonable. Given the scale concerns for this stock, field research to address this topic would be useful.

Assessment models

Four different modelling approaches were explored during the benchmark meeting: 1. Bayesian XSA (*status quo*); 2. Bayesian SCAA, 3. SAM; and 4. GADGET. The first three are single species assessment models and the last was a multispecies model adapted for the Flemish Cap ecosystem. Model specifications were discussed and diagnostics were evaluated for all models.

While both the SAM model and the GADGET model hold promise and should continue to be developed, there were issues with both models that precluded their further consideration during the benchmark meeting as acceptable operational models for this stock. While it was difficult for me to thoroughly evaluate the sufficiency of the SAM model given my lack of experience with state-space models, the large process errors that were being estimated by the model need to be better understood. The fact that the assessment lead for the SAM model was not present at the benchmark meeting, and only available via webinar, made it difficult to fully evaluate this modelling approach. The GADGET model was primarily used to inform natural mortality (M) estimates. GADGET results suggested higher mortality at younger ages, which was primarily attributed to cannibalism. This is a departure from the status quo approach which assumed that M was constant at age. The group reached consensus that the relative M patterns at age derived from the GADGET model were more biologically realistic than the status quo approach. After considerable discussion the group acknowledged that there was insufficient information for any of the models to reliably estimate the actual scale of natural mortality. This led the group to explore external estimates of natural mortality including catch curve analyses and biologically based meta analyses (proportion-at-age, length-weight relationships, etc.). The group consensus was to apply a constant vector of age-specific M values ranging from 1.26 at Age1 to 0.24 at ages 7+. It should be noted that the implied M from this approach is considerably higher than the status quo estimates ~0.2 – at least in terms of the numbers of Age8+ survivors (Figure 1).

The group reached consensus to move from the existing Bayesian XSA approach to a Bayesian SCAA. Given the various data issues and uncertainty in the catch data and the inability of the XSA to deal with zeros in survey indices-at-age without replacement with some artificial proxy, this seems warranted. There were some large differences between the earlier runs of the XSA model and SCAA model results in terms of scale, though these were mitigated in some of the XSA runs explored late in the week. The SCAA results appeared to be relatively robust to alternative configurations lending additional confidence to the results. The R37 run of the SCAA model was the final variant explored during the benchmark meeting.

In an effort to better evaluate the patterns in the input data, and to understand the impact of various model configurations, I found it helpful to place the input data into the statistical catch-at-age model framework that I am more familiar with. For this exercise I used the ASAP model (Legault and Restrepo 1998, Legault 2012). This framework is different than that employed in the SCAA model evaluated during the benchmark meeting, most notably:

- Fishery selectivity was time invariant, though allowed to vary by age.
- Can incorporate time varying sampling error.
- Models catchability-at-age using multinomial distribution of proportions-at-age and fits to the aggregate survey index.

The data evaluation, diagnostics and results of an example model run are provided in Annex 1. In general, I found that given the available data the ASAP model was relatively robust to different model configurations (e.g., survey selectivity assumptions) with only minor retrospective patterning. Perhaps most importantly, the ASAP results were similar to those of the SCAA Run 37 evaluated during the benchmark meeting.

Projection and reference points

There was inadequate time during the benchmark meeting to discussion reference points and catch projections in depth. One issue that did come up which should be further explored as time allows is the adequacy of the current limit fishing mortality reference point (F_{lim}) of F30%. The central question is whether 30% Spawner Per Recruit relationship is appropriate for this cod stock.

Research recommendations

During the course of the benchmark meeting, several areas of uncertainty were highlighted. Some of these uncertainties could be partially addressed through the research recommendations listed below:

- Age-length keys – attempt resolve the discrepancies between age readers and apply separate commercial and survey age-length keys (ALK) when possible.
- Disaggregate historical ALK data to allow you to explore expansion of the plus group. I'm not sure how useful this would be in an operational model since there isn't a lot of information at the older ages early on in the time series; however as we saw with the survey selectivities and natural mortality estimates there are some dynamics going on in the plus group that are not fully understood (i.e., why are the older fish disappear from either the population or survey at a faster rate than expected) and the ability to look at the dynamics beyond age 8+ in the early period may help better explore this.
- Consider additional targeted field work to collect data that can externally inform natural mortality estimates (e.g., dedicated tagging study).
- Given the concerns over population scale and the believability of survey catchability estimates, field work to externally estimate survey catchability would be informative – is their herding between the doors, wings, etc.? What is catchability? How does itvary by size/age?

References

Legault, CM, Restrepo VR. 1998. A flexible forward age-structured assessment program. ICCAT. Col. Vol. Sci. Pap. 49:246-253.

Legault CM. 2012. Technical Documentation for ASAP Version 3.0 NOAA Fisheries Toolbox (http://nft.nefsc.noaa.gov/).

Figures

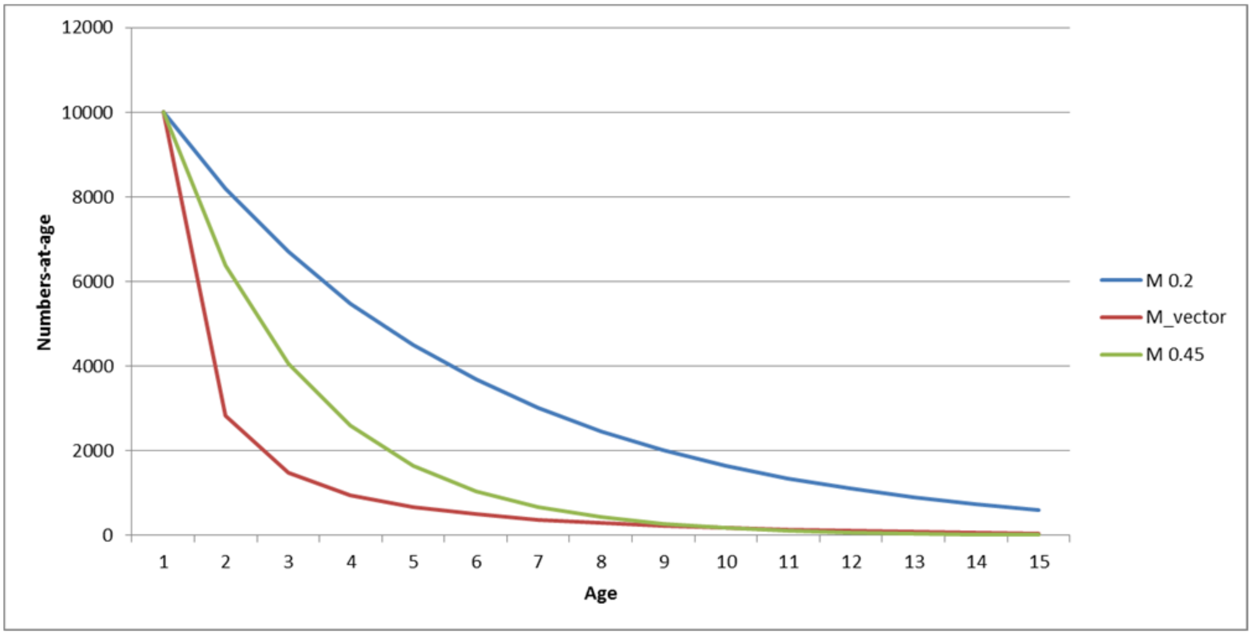


Fig. 1. Comparison of survivors-at-age from an initial population of 1000 fish under three different natural mortality (M) assumptions: M=0.2 (age invariant); M variable at age (Ages1-8+: 1.26,0.65,0.44,0.35,0.30,0.27,0.24,0.24); and M=0.45 (age invariant).