# ICES WKNSSHREF Report 2018

ICES Advisory Committee

ICES CM 2018/ACOM:45

# Report of the Workshop on the determination of reference points for Norwegian Spring Spawning Herring (WKNSSHREF)

10 – 11 April 2018

ICES HQ, Copenhagen, Denmark

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

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

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

Recommended format for purposes of citation:

ICES. 2018. Report of the Workshop on the determination of reference points for Norwegian Spring Spawning Herring (WKNSSHREF), 10–11 April 2018, ICES Headquarters, Copenhagen, Denmark. ICES CM 2018/ACOM:45. 83 pp. https://doi.org/10.17895/ices.pub.5582

# Contents

## Introduction

The Norwegian spring spawning herring (NSSH) was benchmarked in 2016 (ICES, 2016) and XSAM was accepted as the standard assessment model for this stock. The reference points were to be evaluated during the benchmark. However, due to time constraint only $B_{lim}$ was evaluated. The conclusion was that $B_{lim}$ should remain unchanged at 2.5 million tonnes. After WGWIDE in 2016, it was decided that the completion of the review of the reference points should be done before WGWIDE in 2017. At WGWIDE 2017 meeting a working document on the revision of the reference points was presented (Utne, 2017), but WGWIDE concluded that further work needed to be done. Thus, it was decided that the reference points be reviewed before WGWIDE 2018, using the present assessment model XSAM. Following that decision, the coastal states requested ICES to "*to finish the process of re-evaluation of the reference points for Norwegian spring-spawning (Atlanto-Scandian) herring during the first quarter of 2018. Provided that ICES has completed their work on the reference points, the delegations agreed to meet before 15 May 2018 to discuss a possible revision of the long-term management strategy*."

In order to meet the request, a workshop was set up by ICES and took place in ICES headquarters 10-11th April 2018. The workshop was attended by 12 participants.

XSAM simulations were conducted for the NSSH. These followed the ICES advice technical guidelines as published on 20th of January 2017 (ICES 2017) for the estimation of the precautionary and MSY reference points.

## Terms of reference

2018/2/ACOMXX **The Workshop on the determination of reference points for Norwegian Spring Spawning Herring (WKNSSHREF)**, chaired by Katja Enberg*, Norway, and attended by two invited, external experts Massimiliano Cardinale*, Sweden, and Jason Cope, USA will be established and will meet at ICES, Copenhagen10–11April 2018 to:

a ) Address the request from Iceland on behalf of the Coastal States for a re-evaluation of the reference points for the Norwegian Spring Spawning Herring (her.27.1-24a514a):

   i) Apply ICES reference point guidelines to determine precautionary and MSY reference points for the stock

WKNSSHREF will report by 18 April 2018 for the attention of the ACOM.

### Supporting Information

| | |
|---|---|
| Scientific justification | This workshop is to answer the request received the 21 December 2017 from Iceland on behalf of the Coastal States (see ToR a). |
| Resource requirements | |
| Participants | Experts from WGWIDE and stock assessment experts will be required for the work. |
| Secretariat facilities | Meeting rooms and Webex hosting. |
| Financial | Requested budget outlined in the Special Request Form. |
| Linkages to other committees or groups | WGWIDE |
| Linkages to advisory committees | Advice Drafting Group (April 18, by Webex) and release of advice on 26 April 2018. |

# 1 Reference point analyses

## 1.1 Recruitment

The XSAM model fit provides estimates of recruitment at age 2 and SSB. Given this output from the model, spawning stock-recruitment relationship (i.e. numbers of re-cruits at age 2 versus SSB 2 years before) is considered. The specific stock-recruitment models considered here are segmented regression, Beverton & Holt and Ricker. The models are fitted to the output from the XSAM fit and assumes log-normal error. To explore serial correlation in the recruitment noise, the autocorrelation function and partial autocorrelation function were estimated based on the residuals from these fits

## 1.2 $B_{lim}$ estimation

For estimating $B_{lim}$ a categorization of the stock-recruitment relationship into type is required (ICES 2017). The majority in the group agreed that the recruitment dynamic of this stock correspond to the Type 2 stock-recruitment relationship: Stocks with a wide dynamic range of SSB, with evidence that recruitment is or has been impaired at low SSB levels.

Fitting the segmented regression gives a range of break points centered around 2.5 mill tonnes, which is the current $B_{lim}$ value (Fig.1), even if this range is rather large (WD 01). However, in the absence of better defined candidates, the group proposes keeping the $B_{lim}$ unchanged at 2.5 mill tonnes.

A minority of the group argued that according to the ICES guidelines (ICES 2017), NSSH should be classified as a Type 1 stock (WD 02). This is based on the argument that the stock dynamic is driven by occasional strong year-classes, i.e. NSSH can be classified as a spasmodic stock. It was also argued that the segmented regression method (Type 2 approach) is not appropriate for determining $B_{lim}$ for NSSH, due to the large uncertainty in the estimated break-point (WD 1). WKREF (ICES 2007) observed methodological issues with using the segmented regression technique to establish $B_{lim}$ for NSSH and at that time the method was not considered appropriate until the issues had been resolved.

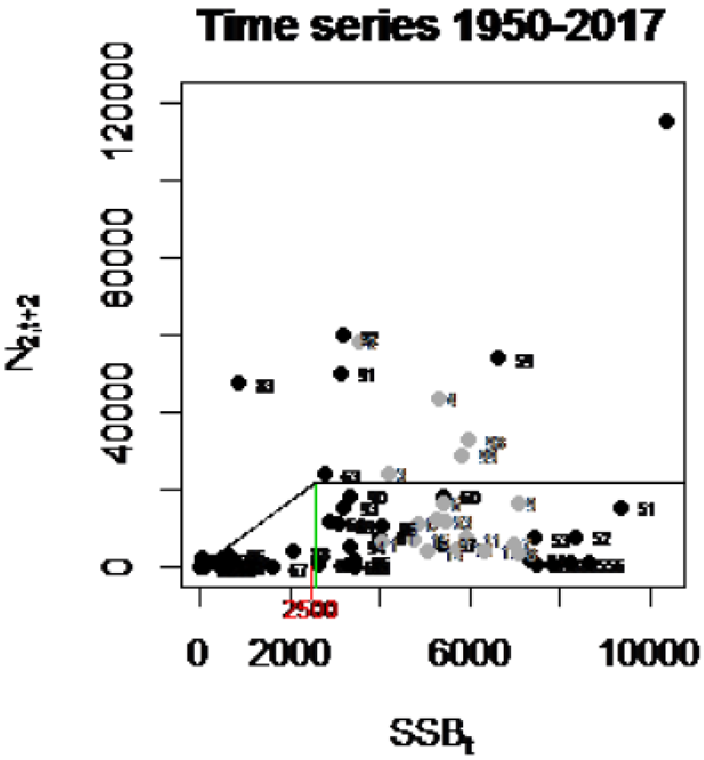


**Figure 1. Segmented regression fit for NSSH.**

## 1.3 XSAM settings

Various estimates of historical time series of abundance at age of NSSH exist, each representing different time- (i.e. 1907-, 1950- and 1988-present) and age- spans and methods (VPA for 1907-1998 (Toresen and Østvedt 2000), Seastar for 1950-2007 (ICES 2007), TASACS 1988-2015 (ICES 2015), and XSAM 1988-2017 (ICES 2017)). The different methods and time-periods used imply that the perception of the stock (and hence stock dynamic) may vary even for overlapping time-periods.

Therefore, XSAM was fitted to a longer time series utilizing catch at age data from 1907—and compared to other historical estimates of the stock as well as the current assessment (see WD 01). Some differences are found, but it is concluded that XSAM estimates represent a reasonable candidate for the historical perception of the stock.

Thus, it was decided to use XSAM and the time series from 1950-2017 for the estimation of the reference points. XSAM was then used in the simulation with the settings described in table 1.

**Table 1. Settings for the XSAM simulation runs for NSSH**

| Data and parameters | Setting | Comments |
|---|---|---|
| SSB-recruitment data | 1950–2017 | Time series from 1907 exists, but with huge deviations in SSB in the early period between models. 1950–2017 was chosen since it represents a large spread in SSB values and less deviations between models. |
| Mean weights and proportion mature | 1988–2017 | |
| Exploitation pattern | Variable according to 1988-2017 | Selectivity is variable according to the XSAM model fit for data 1988–2017 |
| Assessment error in the advisory year. CV of F and SSB | F: 0.26<br>SSB: 0.167 | Based on an average assessment error estimated by retrospective fits and predictions made by XSAM between 2002 and 2017. |
| Assessment error in the assessment year | SSB: 0.147 | Based on the average assessment error estimated by retrospective fits and predictions made by XSAM between 2002 and 2017 |
| Assessment error in last year with catch data | F: 0.152 | |

## 1.4 Weighted F

The WG decided after some discussions to continue to use average $F_{5\text{-}12}$ weighted by stock number as a reference fishing mortality. This approach, which is rarely found within ICES stocks today, has been used annually for NSSH since it was recommended by ACFM in 1995 (ICES 1996). The reasoning for it is the characteristic of the stock's composition and its fishery. When the occasional large year classes are entering the fishable stock, they can be targeted/avoided by the fleets. Using weighed average F in those cases will give different, and more appropriate, measure of fishing level than unweighted average F.

## 1.5 Stock recruitment relation, derived PA points and MSY simulations

Three different recruitment models, segmented regression Beverton & Holt and Ricker were run and applied to the time series, 1950–2016 (Figure 2).

It has been found that point estimates of reference points such as $F_{lim}$ and $F_{MSY}$ may be highly sensitive to the functional relationship for stock recruitment (Simmonds *et al.* 2011), which calls for caution when selecting stock recruitment model. To overcome this problem Simmonds et al. (2011) proposed a method based on model averaging based on AIC. Although this does not appear to be a major issue for NSSH (see WD 01), we used a similar approach; stock recruitment pairs are resampled with replacement, but for each resample, the recruitment model is decided based on AIC. Based on 1000 resamples, the resulting proportion is 0.242 for segmented regression, 0.452 for Beverton & Holt and 0.306 for Ricker.

The simulated recruitment is in line with the observations and fitted model average of the different recruitment models (Figure 3)

During the meeting, it was realized that the results from the simulations in WD01 were presented differently than the approach used in EqSim. In WD01 the statistics for recruitment, spawning stock biomass and catch is shown as the distribution of mean or median of mean or median obtained by each iteration (i.e. representing one time series of simulations) while the approach used elsewhere (e.g. by Eqsim in ICES, 2017) considers the simultaneous distribution across all iterations and time. This affects the perception of variability, as the variability in the latter will be significantly more variable, while the mean and medians are very similar (although not identical). The differences are illustrated in Figure 4, and the group decided to use the simultaneous distribution as in Eqsim as basis for the estimates.

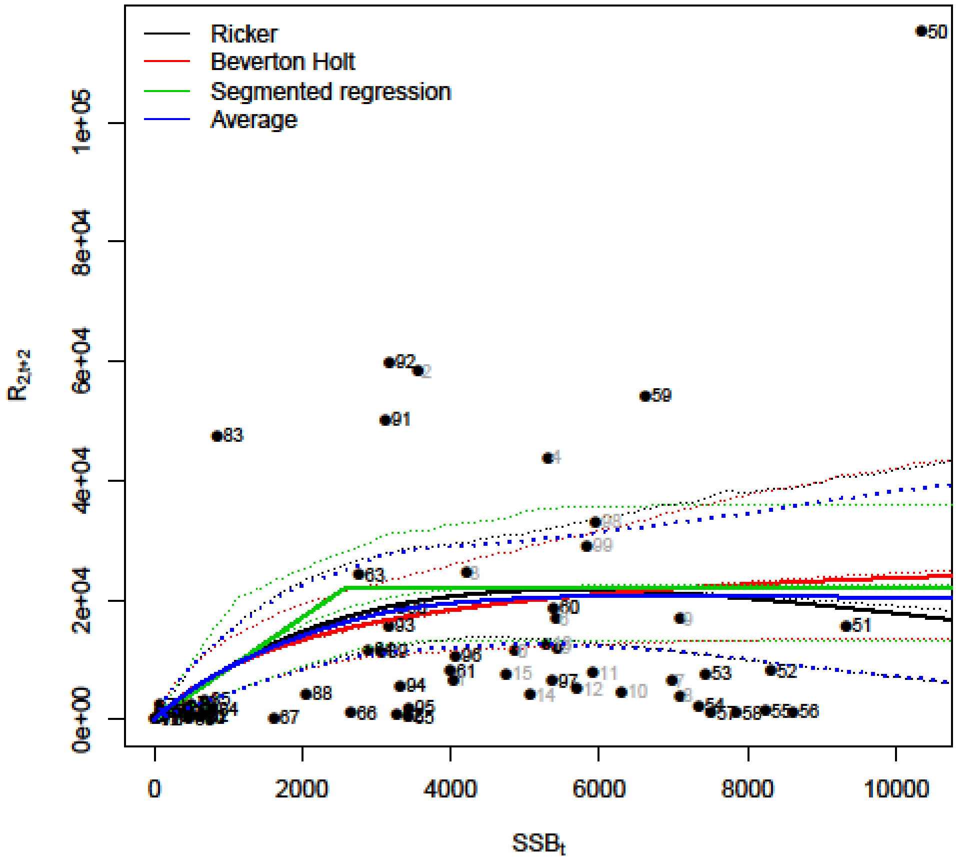


**Figure 2. Recruitment (numbers at age 2) versus SSB (two years before) based on XSAM estimates for 1950–2017. The cohort is indicated alongside the points. The lines are the mean in the fitted recruitment models, segmented regression (green), Beverton & Holt (red) and Ricker (black), and the model average (blue). The model average is based on the AIC-smoothed estimate. The broken lines are 95% confidence intervals of the mean and found by 1000 replicates of pairs of stock recruitment data.**

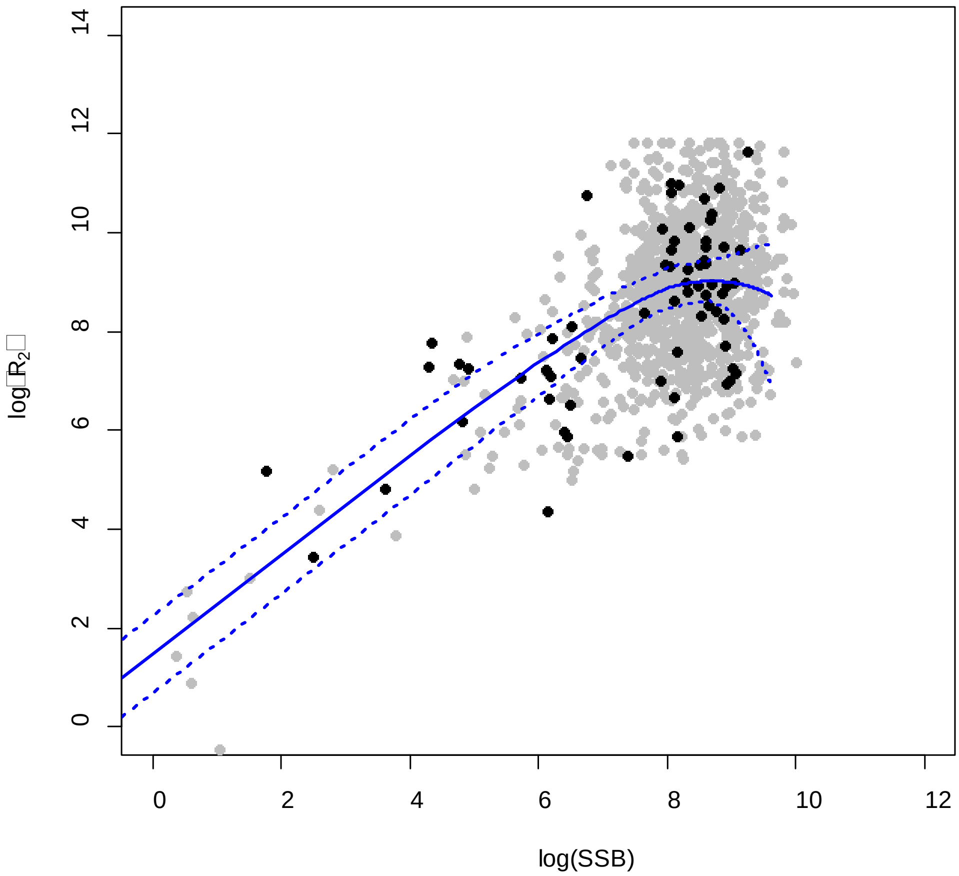


**Figure 3. Observed recruitment versus SSB on log scale (black points). The blue sold line is the fitted mean AIC smoothed recruitment the broken lines are the corresponding 95% confidence intervals. A sample of 1000 simulated values of recruitment when fishing at ~FMSY is included for comparison (gray points).**

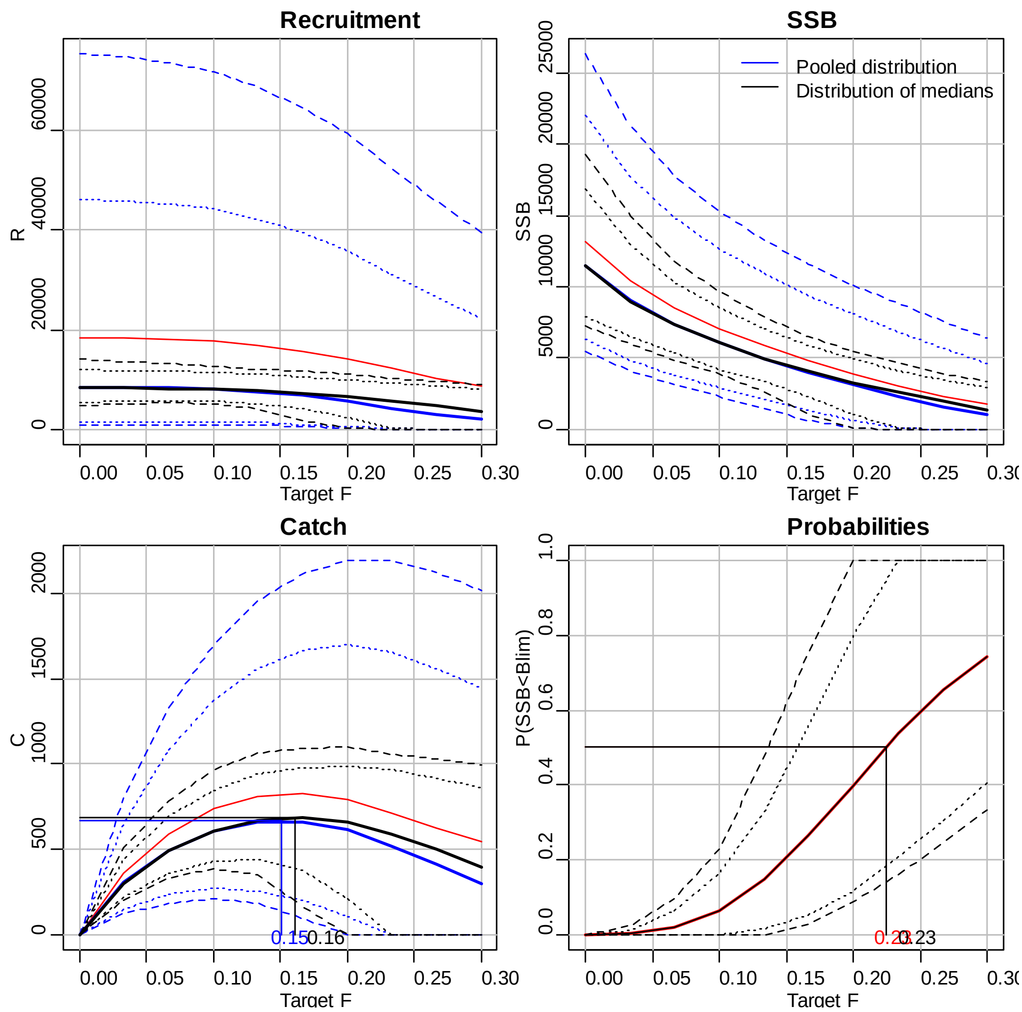


**Figure 4. Comparing the pooled distribution across iterations and time (mean red, median blue, 5% and 95% percentiles dashed blue and 10% and 90% percentiles by dotted blue) by distribution of medians over time by iteration (median of medians by solid black, 5% and 95% percentiles dashed black and 10% and 90% percentiles by dotted black). The mean probabilities of falling below $B_{lim}$ are identical for the two different methods, and can thus not be separated.**

## 1.6 PA reference points

$B_{pa}$ was calculated from $B_{lim}$ (i.e. 2.5 million tonnes) as: $B_{lim}$ * exp(1.645 *σ), where σ is the average SD of ln(SSB) in respective assessment years between 2002 and 2017 - here estimated by XSAM to be 0.147 (Table 1 and WD01). $B_{pa}$ is then estimated at 3.184 million t.

$F_{lim}$ was estimated by simulation using the above values of $B_{lim}$ and $B_{pa}$, setting $F_{cv}$, $F_{phi}$ and $SSB_{cv}$ = 0 (no assessment and advice noise) and with no MSY $B_{trigger}$. $F_{50}$ is the median $F_{lim}$, here estimated to be 0.234.

$F_{pa}$ is calculated from $F_{pa} = F_{lim}$ * exp(-1.645 * σ), where σ is SD of ln(F) in 2016 (the last year model estimate) - here estimated by XSAM to be 0.152. $F_{pa}$ is then estimated to be 0.182.

## 1.7 MSY reference points

$F_{MSY}$ is initially estimated as the F that maximize median long-term yield in the simulation under constant F exploitation (Fig. 5) including prediction error for F. Values of cvF = 0.260 and cvSSB = 0.167 were used, which were obtained as an average of the retrospective bias of XSAM fits, including prediction to the data (Table 1). The initial $F_{MSY}$ was estimated at 0.152, which is lower than Fpa (0.182), but resulted in long-term P(SSB < $B_{lim}$) larger than 5%.

MSY $B_{trigger}$ was set as equal to $B_{pa}$, since the median (and mean) distribution of 5% percentiles of SSB when fishing at $F_{MSY}$ was smaller than $B_{pa}$.

The initial $F_{MSY}$ estimate was then checked for precautionarity in simulations using the initial estimate of $F_{MSY}$ in combination with MSY $B_{trigger}$ following the ICES advice rule (Fig. 5). $F_{p05}$, the F that leads to a 5% P(SSB<$B_{lim}$), was estimated at 0.102. The precautionary principle states that if $F_{MSY}$>$F_{p05}$, then $F_{MSY}$ should be reduced to $F_{p05}$. This is the case here and the final $F_{MSY}$ therefore equals to 0.102 (Table 2).

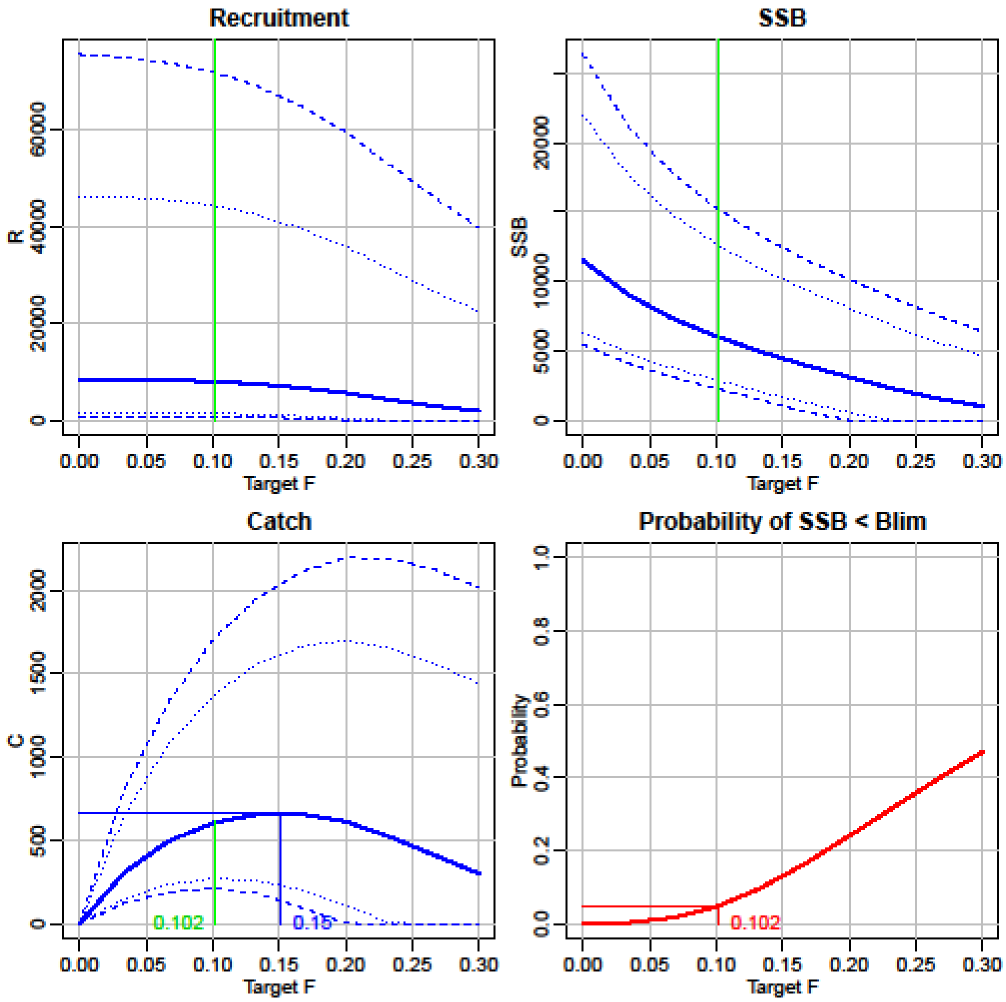


**Figure 5. Median recruitment, SSB and catch when fishing with constant target F without $B_{trigger}$ including prediction error (blue solid lines), and probability of falling below $B_{lim}$ in any year using the MSY approach with $B_{trigger}$ = $B_{pa}$ (red line). Corresponding 5% and 95% percentiles are shown with dashed lines and 10% and 90% percentiles with dotted lines. The $F_{MSY}$ estimate is indicated with the blue line, while the P05 value is indicated with the green lines.**

**Table 2. Final estimated reference points for NSSH. Weights in million t, mean F for ages 5-12.**

| Ref. pt. | MSY $B_{trigger}$ | $B_{pa}$ | $B_{lim}$ | $F_{pa}$ | $F_{lim}$ | $F_{p05}$ | $F_{MSY}$ |
|---|---|---|---|---|---|---|---|
| value | 3.184 | 3.184 | 2.500 | 0.182 | 0.234 | *0.102* | 0.102 |

## 1.8 Sensitivity Runs

A series of simulations based on different configurations of the simulation model were tested, including different stock recruitment models and assumptions on structure in noise for the recruitment process (WD01). Different configurations results in some variability in the estimated reference points, but the points were not significantly different. In addition, another simulation model was also tested based on the same data (WD 3). The results from these runs confirm the range of reference points obtained in WD01 and consequently the estimates obtained by the group.

## 1.9 Exploratory runs

Estimation of $F_{msy}$ and $B_{lim}$ was also done by a model that has been used for HCR evaluation for a number of Icelandic stocks (see WD 03). The work is an update of the work presented at WKPELA-2016. The main emphasis here was to test the sensitivity of estimated reference points to different model settings and data. One of the difference in the settings of the model compared to XSAM is that catch of age 1 was included.

Description of alternative runs.

- Starting years 1907, 1950 or 1975
- Number of selection periods 1-6 or VPA
- Oldest age 12 or 15
- In the estimation phase autocorrelation of recruitment was estimated or set to 0
- In the simulation phase autocorrelation of recruitment was set to 0.35 when not estimated but else the estimated values were used.

Estimated $SSB_{break}$ points using a segmented regression, ranged between 2230 and 3150 thousand tonnes, and $F_{05}$ between 0.1 and 0.15, with most of the $F_{05}$ values lying close to 0.125. The simulations were done with $B_{trigger}$=3000 thousand tonnes.

Results from the simulations show maximum median catch between 0.800 And 1 million tonnes. Those numbers are 15–20% higher than those obtained from the XSAM model and might explain higher values of $F_{05}$. The model does though give comparable value of $F_{05}$ ($\approx$ 0.1) when the simulation period is from 1950 to present. Lower median catches can partly be explained by the fact that age 1 is not included in XSAM model but before the collapse substantial catches of ages 0 and 1 took place.

## 2 Recommendations

Following this reference point estimation, a management strategy evaluation is anticipated. ICES would like to highlight some issues regarding the upcoming evaluation:

1) Current management plan is using spawning stock biomass as the reference unit, i.e., the target F depends on the level of SSB. This means that it is of crucial importance that maturation process is correctly presented in assessment and forecast. However, there is uncertainty associated with this as the maturation process is dependent on year class strength, but at the same time, the maturation ogive for a given year class can only be correctly estimated when all individuals have matured. ICES therefore recommends testing the suitability of using total biomass at ages 4+ or 5+ as the reference biomass in the management plan.
2) In the current management plan, $B_{pa}$ is used as the MSY $B_{trigger}$. ICES recommends testing trigger points in the range from $B_{lim}$ to MSY $B_{trigger}$, in connection with different target F values.
3) Using F as the control variable in the management strategies has long traditions, but it is intuitively not easy to understand the impacts of different F-values in combination with changing selection patters and stock dynamics. ICES therefore recommends testing harvest ratio-strategies, where instead of a given F, a given proportion of the harvestable biomass can be taken. This is an intuitive and easy to understand measure of harvest pressure, and not sensitive to changes in selection patterns. In addition, the discussion about unweighted vs unweighted F would become superfluous if the management strategy would be based on harvesting a given proportion of the harvestable biomass.

# 3 Reviewers' comments

## *3.1* *Review of* **WKNSSHREF - Workshop on the determination of reference points for Norwegian Spring Spawning Herring**

*By Max Cardinale*

**General comments**

The report summarizes comprehensively the work conducted by the WKNSSHREF to estimate the reference points for Norwegian Spring Spawning Herring (NSSH). In general, I agree with the methodologies and the choices made the WG. Moreover, the WG did follow accurately the ICES guidelines for the estimation of the reference points. Therefore, I consider the defined reference points as appropriate for NSSH with the caveats of the choice and use of autocorrelation in R. Moreover, several clarifications are needed as some parts of the report are unclear and it is not easy to fully understand all the assumptions used in the simulations. In this concern, I have also made some language editing which might help to (maybe) to improve readability of the report.

**Specific comments**

It would be important to specify the exact configuration of the latest accepted assessment model of NSSH (in 2017) and if the final model used by WKNSSHREF to derive the reference points is somehow different from the final accepted model. This should be done in the very beginning of the document.

I agree that based on the current analysis the $B_{lim}$ should be left unchanged but given the large uncertainty of the $B_{lim}$ value it would be important to conduct a sensitivity analysis of the effect of the range of $B_{lim}$ on the estimated $F_{p05}$ (see also comments on sensitivity analysis below). Also, I agree that Type-2 stock-recruitment relationship is the most suited for NSSH and that 1950-present should be used in the simulations.

The table (i.e. Table 1) of the settings for the XSAM model of NSSH is incomplete. Mean weights and proportion mature are indicated to be set from 1988 to present but I guess that an average has been used while the model also includes year-specific values for the period 1950-1987. The same applies for the selectivity.

Autocorrelation in recruitment (R) is generally affecting $F_{p05}$. According to literature, pelagic species of fish have an autocorrelation in R around 0.4-0.6. It is not clear from the report if the final model does indeed use autocorrelation in R in the simulations to estimate the reference points and, if this is the case, the autocorrelation in R was set to a specific value or estimated from the stock-recruitment (SR) data. Given the impact of autocorrelation in R and the fact that $F_{p05}$ is limiting $F_{MSY}$, it is important that this particular aspect of the simulation is clearly described and that autocorrelation in R is included if significant.

I appreciated the numerous sensitivity runs conducted by the group but there is a tendency in ICES to spend a lot of work on doing sensitivity but then the results have practically no impact on the final estimates. I would suggest that in the future sensitivity runs results are factually used to build an ensemble of plausible model configurations from which is possible to estimate key parameters as for example $B_{lim}$, MSY $B_{trigger}$,

$F_{p05}$ and $F_{MSY}$. This would allow to properly taking into account the process (model) uncertainty, which is what we generally want to do when conducting sensitivity runs.

I also agree that moving the entire ICES machinery away from F and SSB would greatly simplify the decisions to be taken when estimating reference points (and stock status in general) and will also reduce the impact on the simulations of the choice of several parameters, especially maturity ogives and selection pattern (in reality F at age, XSAM does not include a function for selectivity).

## 3.2 *Review of:* WKNSSHREF - Workshop on the determination of reference points for Norwegian Spring Spawning Herring

*By Jason Cope*

**General observations/comments:**

- Recruitment and $B_{lim}$: It seems the choice of a hockey-stick (e.g., segmented regression) approach was chosen over other possible production relationships as the one to determine $B_{lim}$. Why this was chosen (e.g., historical reasons, expert opinion it is the most biologically realistic production relationship) should be included in the report. I suspect Type 2 is the identifier for this type of S-R relationship. If that is the case, that needs to be more explicit (see my comment in the report), and would then provide the explanation that the expert panel chose that relationship over the other two. Regarding the fit, while the large variability in recruitment given biomass is a normal phenomenon, one worry I have is that the residuals pattern looks strange for the given fit. It seems there are many more points below the lines than above. While the statistical fit may chose the piecewise regression to fit in that manner, as a manger I would be concerned with the possibility that the bigger probability is a low recruitment event. As a thought exercise and possible sensitivity, if you excluded the top 6 most variable recruits above the line, you would likely get a very different break point (my eyeball guess is above 3000). Without being able to see the level of uncertainty in the estimate of recruitment, it is hard to tell how well determined those very large recruitments are. Using this alternative recruitment scenario would provide insight into what a more precautionary $B_{lim}$ could be. If it is not too far from 2500, then $B_{lim}$ is robust to the variability an uncertainty in the recruitment estimates. If it is determined to be significantly higher than 2500, then managers would have the option to consider that sensitivity to uncertainty in any decisions. Just to be clear, this is not altering the candidate S-R relationship, it is just a way to gauge the robustness to the uncertainty and wild variability in the recruitment relationship. I understand the minority reports trepidation over the use of the Type 2- the above suggestion could be a way of offering more investigation into that robustness against uncertainty when choosing the Type 2. I also wonder if it would be useful to provide what $B_{lim}$ would be for the other S-R relationships. Again, this would provide a nice way to present the robustness in this choice and how much it matters.

- XSAM settings: I would agree that having a longer catch time series is more beneficial to set baseline values and capture dynamics. I do not recall the difference between 1907 and 1950 as far as removals go (I would tend toward the longer time series), but 1950 seems reasonable.

- Weighted F: Weighted F is an interesting way to get at fishing intensity. I support the consideration of its use in this assessment.

- Stock-recruitment comparisons: This is a good comparison to include so as to offer insight into different model fits. Given the model averaging exercise has the B-H model the highest weighted and the segmented regression the lowest, should this affect the decision on what model to use for $B_{lim}$? Biologically, the

Ricker and B-H models hypothesize very different reasons for the behaviour of the stock recruit relationship. I would think that enough is known about NSSH to say one model is a better fit to NSSH biology than the other. While statistical fits are good, they usually do not provide everything to determine biological significance. I think it possible that one could rule out either the B-H or Ricker models, then model average with the remaining model and the segmented regression. Something for future consideration.

- PA and MSY reference points: There is no uncertainty being reported for these values (though the MSY figure does show CIs). Given the large measurement and process error involved in assessing NSSH, agreeing on some way to report the uncertainty in the table of these reference points would be good.

- Sensitivity tests: It is mentioned the sensitivity runs were not considered significant, but how this was determined needs to be included.

- Recommendations section: #3 states a harvest ratio could be used as an alternative to using F (which is harder to understand). Just for clarification, this means using exploitation rates (e.g., catch/biomass) instead of instantaneous rates (e.g., F) because the units are easier to comprehend. I agree with all of that—just wanted to make sure I understood the recommendation correctly.

# 4 References


Dingsør, GE. 2018. Blim for Norwegian spring spawning herring. WKNSSHREF WD 2

ICES 1996: Report of the Northern Pelagic and Blue Whiting Fisheries Working Group, ICES CM 1996/Assess:14

ICES. 2007. Report of the Workshop on Limit and Target Reference Points (WKREF), 29 January – 2 February 2007, Gdynia, Poland. ICES CM 2007/ACFM:05. 89 pp.

ICES. 2015. Report of the Joint ICES-MYFISH Workshop to consider the basis for FMSY ranges for all stocks (WKMSYREF3), 17–21 November 2014, Charlottenlund, Denmark. ICES CM 2014/ACOM:64. 156 pp.

ICES 2017. ICES fisheries management reference points for category 1 and 2 stocks. DOI: 10.17895/ices.pub.3036

ICES. 2017. Report of the Workshop to consider FMSY ranges for stocks in ICES categories 1 and 2 in Western Waters (WKMSYREF4), 13–16 October 2015, Brest, France. ICES CM 2015/ACOM:58. 187 pp.

Simmonds, E. J., Campbell, A., Skagen, D., Roel, B. A., and Kelly, C. 2011. Development of a stock-recruit model for simulating stock dynamics for uncertain situations: the example of northeast atlantic mackerel (scomber scombrus). ICES Journal of Marine Science, 68(5):848{859.

Toresen, R., and Østvedt, O. J. 2000. Variations in abundance of Norwegian spring-spawning herring (Clupea harengus, Clupeidae L.) throughout the 20th century and the influence of climatic fluctuations. Fish and Fisheries, 1(3): 231–256.

Utne, K.R. 2017. Evaluation of reference points for Norwegian Spring Spawning Herring. Working document to WGWIDE, 2017.

## Annex 1: List of participants

| NAME | AFFILIATION | COUNTRY | EMAIL |
|---|---|---|---|
| Katja Enberg (Chair) | IMR – Institute of Marine Research | Norway | katja.enberg@imr.no |
| Massimiliano Cardinale (external expert – parts of the meeting by WebEx) | SLU – Institute of Marine Research | Sweden | massimiliano.cardinale@slu.se |
| Jason Cope (external expert) | NOAA - National Oceanic and Atmospheric Administration | USA | jason.cope@noaa.gov |
| David Miller | ICES Secretariat | Denmark | david.miller@ices.dk |
| Alexander Krysov | PINRO -Knipovich Polar Research Institute of Marine Fisheries and Oceanography | Russia | a_krysov@pinro.ru |
| Claus Sparrevohn | Danish Pelagic Producers' Organisation | Denmark | crs@pelagisk.dk |
| Dmitry Vasilyev | VNIRO - Russian Federal Research Institute of Fisheries and Oceanography | Russia | dvasilyev@vniro.ru |
| Erling Kåre Stenevik | IMR – Institute of Marine Research | Norway | erling.kaare.stenevik@imr.no |
| Eydna í Homrum | FMRI - Faroe Marine Research Institute | Faroe Islands | eydnap@hav.fo |
| Gjert E. Dingsør | Norwegian Fishermen's Association | Norway | gjert@fiskebat.no |
| Gudmundur J. Óskarsson | MFRI - Marine and Freshwater Research Institute | Iceland | gudmundur.j.oskarsson@hafogvatn.is |
| Höskuldur Björnsson | MFRI - Marine and Freshwater Research Institute | Iceland | hoskuldur.bjornsson@hafogvatn.is |
| Mimi E. Lam | UiB - University of Bergen | Norway | mimi.lam@uib.no |
| Sondre Aanes | Norwegian computing Center | Norway | sondre.aanes@nr.no |

## Annex 2 – Request

| | |
|---|---|
| Request from | Coastal States |
| | |
| Contact within organisation | Sigurgeir Þorgeirsson (Ministry of Industry and Innovation, Iceland) |
| Content contact person | Sigurgeir Þorgeirsson |
| | |
| Request announced | 21 December 2017 |
| Request received | 21 December 2017 |
| Answer deadline client | 15 May 2018 |
| | |
| Request code (client) | ANR17010027/20.8.1 |
| Request code (ICES) | |
| Request | Re-evaluation of reference points for NSSH (ASH) |
| Background:Following the benchmark for NSSH in 2017, no new reference points were agreed upon. The Coastal States (CS) have underlined the importance of this work as well as the strong wish that it will be completed early enough for the CS being able to revise the long-term management strategy next spring or early summer. The wish of the CS is that the revision may be finalised and evaluated by ICES in time so that ICES can base the advice for 2019 on the new strategy.<br><br>**Request:** *The delegations agreed to request ICES to finish the process of re-evaluation of the reference points for Norwegian spring-spawning (Atlanto-Scandian) herring during the first quarter of 2018. Provided that ICES has completed their work on the reference points, the delegations agreed to meet before 15 May 2018 to discuss a possible revision of the long-term management strategy.*<br><br>**Deadline: 15 May 2018** | |
| | |

| | |
|---|---|
| Planning ICES | The structure and participants required:<br>Work will be conducted remotely by experts familiar with NSSH and MSE work. A 2 day workshop will then be held to review the work done and decide on the final reference points. Experts from WGWIDE and MSE experts will be required for the work.<br>Meetings required: One workshop (WKNSSHREF), 10-11 April 2018<br>Deliverables: Workshop report, advice on appropriate PA and MSY reference points. |
| Request (budget) accepted | |
| ICES contact person | David Miller |
| WG(s) involved | WGWIDE |
| Preparation timing | |
| Review group | Two external reviewers |
| Advice drafting group | ADGNSSHREF, by Webex. 18 April 2018 |
| ACOM Webex | WCNSSHREF, 25 April 2018 |
| Release date | 26 April 2018 |

## Annex 3: Working documents

**List of Working Documents:**

WD01. Aanes, S., Stenevik, E. K. and Enberg, K. 2018. Reference points for Norwegian spring spawning herring. 48 pp.

WD02. Dingsør, G. E. 2018. $B_{lim}$ for Norwegian spring spawning herring. 6 pp

WD03. Bjørnsson, H. 2018. –Norwegian spring spawning herring. Estimation of reference points. 6 pp

# Reference points for Norwegian spring spawning herring

Sondre Aanes[1], Katja Enberg[2], Erling Stenevik[2] and Kjell Utne[2]

[1]Norwegian Computing Center, Oslo, Norway
[2]Institute of Marine Research, Bergen, Norway

## Summary

In WGWIDE XSAM is currently the assessment model adopted for NSS herring where the assessment runs from 1988 which is recognized as a too short time series to provide a basis for estimating reference points. For establishing reference points it is desirable that the basis used for estimation is consistent with the current assessment. Therefore XSAM is fitted to longer time series utilizing catch at age data from 1907—and compared to other historical estimates of the stock as well as the current assessment. Some deviations are found, but it is concluded that XSAM estimates represent a reasonable candidate for historical perception of the stock. Using the XSAM estimated historical stock sizes it is suggested to use the time series 1950-2017 to estimate relevant dynamics enabling specification of the simulation model needed to perform the estimation of effects of harvesting (e.g. reference points). From a range of different configurations concerning specification of recruitment model and assumptions on selection pattern and mean stock-weight- , catch-weight-, and proportion-mature-at-age, the configuration which resembles the history adequately in terms of means and variances is proposed as the final configuration to use for estimation of reference points. The document ends with proposed estimates of the reference points for $B_{lim}$, $B_{pa}$, $F_{lim}$, $F_{pa}$, $F_{MSY}$, $B_{trigger}$,and $F_{P05}$ according to the ICES guidelines. For some of the reference points that have been defined earlier, $B_{lim}$, $F_{pa}$ and $F_{MSY}$, there is little evidence for changing the reference points from the already existing reference point for this stock.

## Introduction

Norwegian spring spawning herring (NSSH) is a commercially important pelagic stock in the northeast Atlantic, with catches in some years exceeding 1.5 million tons (Figure 1). NSSH is not fished in the nursery areas in the Barents Sea, and is recruited to the fishery when maturing at age 4-5 years old. The fishery is taking place during the summer feeding, in the autumn when individuals are returning from the feeding areas, around Christmas on the overwintering areas and in the winter (February-March) at the spawning grounds. The spawning stock biomass has historically shown large fluctuations, mainly due to variable recruitment with occasional very strong year classes. The stock collapsed in the late 1960`s, and did not recover before the very strong 1983-year class recruited to the spawning stock. Several strong year classes in the period from 1991 to 2004 ensured that the spawning stock biomass increased in this period. In the absence of strong year classes recruiting to the stock, the spawning stock biomass has decreased since around 2009, and was estimated to be 4.1 million tonnes in 2017.

ICES reviewed the reference points of NSSH in 2013 in combination with the NEAFC request to evaluate of alternative management plans for this stock (ICES 2013d). ICES then concluded then that $B_{lim}$ should remain unchanged at 2.5 million tonnes. $B_{pa}$ was not revised as it is defined based on $B_{lim}$. $F_{MSY}$ was evaluated and it was considered that it should remain unchanged at $F_{MSY}$ = 0.15. $F_{lim}$ has previously not been defined for this stock.

The NSSH was benchmarked in 2016 (ICES, 2016) where XSAM (Aanes, 2016) was accepted as the standard assessment tool for this stock. The reference points were to be evaluated during the benchmark. However, due to time constraint only $B_{lim}$ was evaluated. The conclusion was that $B_{lim}$ should remain unchanged at 2.5 million tonnes. After WGWIDE in 2016 it was decided that the completion of the review of the reference points should be done before WGWIDE in 2017. At that meeting, a working document on the revision of reference points was presented (Utne, 2017), but it was concluded that further work needed to be done and it was decided that the reference points be reviewed before WGWIDE 2018, using the present assessment model XSAM to perform the simulations needed. Following that decision, the coastal states requested ICES to "to finish the process of re-evaluation of the reference points for Norwegian spring-spawning (Atlanto-Scandian) herring during the first quarter of 2018. Provided that ICES has completed their work on the reference points, the delegations agreed to meet before 15 May 2018 to discuss a possible revision of the long-term management strategy."

# Methods

## Time series of abundance

Ideally, estimation of reference points and MSE should be based on a time series of stock abundances and other population parameters (e.g mortality, size and proportion mature at age) sufficiently long to enable characterizing key features of the processes causing fluctuations in the population dynamics. Narrowing the scope by assuming natural mortality known, the key processes governing fluctuations in numbers are recruitment and fishing mortality.

The perception of the stock used for MSE should be based on a consistent time series which also is consistent with the current method used for management advice. More specifically, this means a time series that is derived using the same method throughout the time series. Furthermore, since MSE is relevant for the current and future management advice process, the basis must be consistent with the current assessment, and it is a major advantage if same modelling framework can be applied for both assessment and MSE. On the other hand, if different methods provide practically the same perception of the stock, it can be argued that a time series can be constructed by combining time series derived using different methods.

The current assessment model adopted by ICES is XSAM (ICES 2017). This model uses catch at age data and survey indices from 1988 and onwards. The start of the time series represents the onset where continuous abundance indices are available up to present time. In the period prior to 1988 the main data source is catch at age. Some tagging data is available in this period but were not used in the assessment from 2006 because of the extreme sensitivity of the assessment to the inclusion of these data, possibly due to low tag recoveries. These data have not been considered here. XSAM for herring was developed and evaluated with the basis having both catch at age and abundance indices at age, although it in principle will work for only catch-at-age data as well. WGWIDE uses the age span 2-12+ as data on ages 0 and 1 and older than 12 is considered too imprecise to contribute with information concerning year-class strength.

Various estimates of historical time series of abundance at age of Norwegian spring-spawning herring (NSSH) do exist, each representing different time- and age- spans and methods (VPA for 1907-1998 (Toresen and Østvedt 2000), Seastar for 1950-2007 (ICES 2007, Tjelmeland 2005), TASACS 1988-2015 (ICES 2015), and XSAM 1988-2017 (ICES 2016, 2017)). Thus, different time periods that are natural to consider are 1907-, 1950- and 1988-. The different methods and time-periods of data implies that the perception of the stock (and hence dynamics) may vary even for overlapping time-periods. As an additional source of information a conventional VPA using data 0-15+ is fitted using updated data (1907-2017).

In all attempts extending the time series backwards, the configuration for XSAM for 1988- are kept as in the latest assessment (ICES 2017).

### Configuration of XSAM for 1907-1987

The quality of the catch data has not been evaluated in this study. For herring it is found that the variance in the input data can be well described by Taylors spatial power law $\sigma^2 = Var(\hat{\mu}) = \alpha\hat{\mu}^{\beta}$, where $\beta$ is close to 1.5 for most available data on sampling errors (ICES 2016, 2017). Therefore, this model for observation error is adopted fixing $\beta$ to 1.5. Estimates of $\alpha$ controlling the observation variance is highly imprecise and sensitive to the time series of catch at age data used (e.g. 1950-1988 or 1907-1988). This illustrates the difficulties separating observational variance from process variance as simultaneously estimating variances for the state process and for the observation model is often difficult due to identifiability problems (see e.g. Newman et al. 2014 and references therein). Therefore, $\alpha$ is fixed such that it corresponds to an RSE of on average ~23% across ages and years which is obtained by an initial fit of the model to data 1907-1987.

XSAM includes a time series model for F. A constant management regime over time implies that this may be a reasonable assumption while a change in management could cause violating this assumption as the process is likely to be changed. The harvesting dynamics for herring has certainly changed, moving the fishing pattern from young fish over to older fish the last 40 years or so, partly due to moving from 20 to 25 cm minimum catch size in the seventies. In addition, the fishery during the collapse period is likely to have caused different fishery dynamics compared to other periods. With this basis, XSAM was configured with a time series model for F also prior to 1988, but specified as a *different* time series process. Two cases were explored: First one process for F in 1907-1950 and an independent process for 1950-1987 compared to one process for 1907-1987. These options gave practically the same results concerning abundances supported by AIC which preferred the simplest setup to have one time series model for F throughout the period 1907-1987 (not shown), thus this configuration is used.

# Recruitment

## Recruitment – Models

The XSAM model fit yields estimates of recruitment at age 2 and SSB. Given this output from the model, spawning stock recruitment relationship (i.e. numbers of recruits at age 2 versus SSB 2 years before) is considered. The specific models considered here are Segmented regression, Beverton Holt and Ricker recruitment models. The models are fitted to the data (i.e. the output from the XSAM fit) assuming log-normal error.  To explore serial correlation in the noise in recruitment, the autocorrelation function and partial autocorrelation function are estimated based on the residuals from these fits.

## Recruitment – Variability in parameters and model averaging

The variability in parameters of the stock recruitment models can be explored by fitting the respective models to resampled pairs of stock recruitment data points with replacement. A sample of the simultaneous distribution of parameters (including correlations) and hence an estimate of the distribution is thus easily available. This procedure ignores potential serial correlation in the data and can therefore not reflect the variability in estimates of serial correlation in the residuals. However, subject to the assumption that potential serial correlation in the residuals not will influence the inference of the parameters in the stock recruitment relationship, the estimates may still be valid to reflect the variability in the stock recruitment relationship. It is found that point estimates of reference points such as $F_{lim}$ and $F_{MSY}$ may be highly sensitive to the functional relationship for stock recruitment (Simmonds et al. 2011) which calls for caution selecting stock recruitment model. To overcome this problem Simmonds et al. 2011 and Rindorf et al. 2017 propose a method based on model averaging aided by AIC to objectively find probability of stock recruitment models. We use a related approach by extending the resampling procedure described above, but for each resample, the recruitment model is decided based on AIC, and thus obtain a model average. This is similar to the method in Rindorf et al. 2017: First pairs of stock recruitment data is resampled with replacement, then the model is selected based on AIC for each resample for a pre-determined set of stock-recruitment models and finally the population is simulated forward for a number of years ($T_{tot}$) for a prescribed HCR where the first $T_{init}$ values before the system has stabilized are discarded leaving $T = T_{tot} - T_{init}$ for each resample. By a large number of resamples $n$, the distribution of an AIC smoothed stock recruitment relationship and hence population trajectory, is obtained. In other words, this approach requires $n \times T_{tot}$ simulations to obtain the distribution of required stock statistics. Note that when comparing different HCRs, for each set of parameters, the HCR should be simulated under the same environmental conditions. This can be accomplished by setting the seed in the random number generator to the same value for each set of parameters/models across HCRs to be compared.

## Recruitment – Number of replicates

Due to runtime considerations it is highly desirable to keep $n$ and $T_{tot}$ as low as possible yet sufficiently large to provide adequately robust estimates of the distribution. To establish sufficient sizes of $n$, $T_{tot}$ and $T_{init}$ the following approach is made. $T_{init}$ is set by a visual inspection of generated time series, and $T_{init} = 250$ appear sufficient for the time series to have reached stationarity, e.g. stable fluctuations around equilibrium. To determine $n$ and $T$ they are first both set to 1000. Then the moments (mean and variance) and percentiles are extracted by selecting a large number of random subsets varying size for both $n$ and $T$. In this way we can numerically inspect measures of stability in the estimate as a function of varying $n$ and $T$. The results is then verified by running two or more independent simulations with specified $n$ and $T$ for the same settings, and the difference in the results should match the first resampling procedure for corresponding $n$ and $T$.

### Serial correlation

For NSS herring there is statistically strong evidence that the noise in the stock-recruitment process is serially correlated (see Results this document). This was also found in Lillegård et al. (2005) who demonstrated that the serial correlation in noise can be caused by external environmental influence as they showed that the noise was statistically significantly correlated with mean annual sea temperature in the "Kola section", which again follows a AR(1) process, and thus induces serial correlation in the noise. Here, we make no attempt to link the noise to a mechanism, but simply conclude that the noise is serially correlated. To account for serial correlation in addition to variability in parameters the following approach is taken to match the above described method for variability in parameters. First consider the AR(1) model for the residuals

$$\varepsilon_t = \alpha\varepsilon_{t-1} + W_t$$

where $E(W_t) = 0$ and $Var(W_t) = \sigma_W^2$ . The variance in the distribution of the residuals is $Var(\varepsilon_t) = \sigma_\varepsilon^2$ is given by $\sigma_\varepsilon^2 = \alpha^2\sigma_\varepsilon^2 + \sigma_W^2$ such that $\sigma_W^2 = (1-\alpha^2)\sigma_\varepsilon^2$. Then for a given $\alpha$, $\sigma_W^2$ can be set by $(1-\alpha^2)\sigma_\varepsilon^2$ such that the variance in the residuals is maintained at the same time as a time series structure is imposed. To be very specific, the method described above is extended by for each resample, an estimate $\sigma_\varepsilon^{2*}$ is obtained. Then the corresponding variance in the AR(1) process is $\sigma_W^{2*} = (1-\alpha^2)\sigma_\varepsilon^{2*}$, and thus the total variability in the noise is preserved. We use the value of $\alpha$ as estimated from the original data. A similar approach can be used for time dependencies at higher time lags. Note that this approach ignores variability in the $\alpha$ parameter, but is believed to be of minor importance as long as the variability in $\sigma_\varepsilon^2$ is appropriately accounted for.

## Simulation model

A fundamental part of estimating effects of harvesting and establishing reference points is both to use a population dynamical model which contains key features such that the projections contains some realism as well as quantifying realistic levels of random variability which occurs due to stochastic processes (such as recruitment) and due to uncertain parameter estimates. Since we here assume that natural mortality is known and constant, the stock recruitment model becomes crucial in estimating and simulating realistic dynamics.

Using XSAM as basis for the simulations, this model includes a time series model for fishing mortality including a model for selectivity. Thus, it is natural to consider simulating time varying selectivity according to this model. To reach a prescribed level of F is still possible by scaling the selectivity to meet the F, yet maintaining the variability in selection. A more standard approach is to set fishing pattern constant and use an F multiplier to achieve a prescribed fishing mortality (usually as an average over some ages).

The most commonly used approaches for MSE also assumes natural mortality known and constant and other biological parameters such as mean stock-weight-, catch-weight, and proportion-mature-at-age. Little information on variability in M is available for this stock (but see Bjørkvoll et al. 2012) and will be set known and constant as in WGWIDE. There is variability in all of mean stock-weight-, catch-weight, and proportion-mature-at-age where it is likely to believe that there may be density dependent effects (Engelhard and Heino 2004, Lillegård et al. 2005 modelled proportion mature at age as a function of density) although it is not well documented for NSS herring. To follow common practice we will first explore several cases for establishing constant values of mean stock-weight-, catch-weight, and proportion-mature-at-age to use in the simulations such as long term mean, long term weighted mean using stock size as weights, average of last year and last three years. If time permits we will explore possibilities and consequences of using density dependent, and thus time invariant values of mean stock-weight-, catch-weight, and proportion-mature-at-age. To aid in selecting the appropriate configuration of the simulations, results will be compared with historical estimates of key estimates, both means and variability across time when simulating fishing as it has been observed in the most recent years.

## Establishing reference points

Once the configuration of the simulation model is established, the reference points will be estimated by long term simulations accounting for variability in parameters and possibly in functional forms.

From ICES guidelines (ICES 2017) for ICES fisheries management reference points the definitions of reference points are

### $B_{lim}$

$B_{lim}$ is a deterministic biomass limit below which a stock is considered to have reduced reproductive capacity. The basis is the biomass below which recruitment reduces with spawning-stock biomass (SSB), e.g. the change point of a segmented regression.

### $F_{lim}$

$F_{lim}$ is the exploitation rate which leads SSB to $B_{lim}$. The basis is the fishing mortality rate (F) that in stochastic equilibrium will result in median(SSB) = $B_{lim}$ (i.e. 50% probability of SSB being above or below $B_{lim}$).

### $B_{pa}$

$B_{pa}$ is defined as a stock status reference point above which the stock is considered to have full reproductive capacity, having accounted for estimation uncertainty. In the guide lines it may be interpreted as two different basis for this is given

1. The value of the estimated SSB, which ensures that the true SSB has less than 5% probability of being below Blim, i.e. the 95th percentile of the distribution of the estimated SSB if the true SSB equals Blim.
2. Bpa = Blim × exp(1.645 × σ) where σ is the standard deviation of ln(SSB) at the start of the year following the terminal year of the assessment. If σ is unknown 1.4 can be used as default for “exp(1.645 × σ)”, equivalent to σ = 0.20.

However, since 1.645 is the 95% quantile in a normal distribution, it is interpreted that if it is assumed that the estimated SSB follows a lognormal distribution, then the first point is achieved by the second point as it is not clear how to derive a Bpa from the first point alone.

### $F_{pa}$

$F_{pa}$ is an exploitation rate reference point below which exploitation is considered to be sustainable, having accounted for estimation uncertainty. The basis is the value of the estimated F, which ensures that the true F has less than 5% probability of being above $F_{lim}$, i.e. the 5th percentile of the distribution of the estimated F if the true F is equal to $F_{lim}$. $F_{pa} = F_{lim} \times \exp(-1.645 \times \sigma)$ where σ is the standard deviation of ln(F) in the terminal year of the assessment. If σ is unknown $1.4^{-1}$ can be used as default for "exp(-1.645 × σ)", equivalent to σ = 0.20.

### $F_{MSY}$

$F_{MSY}$ is the F expected to give maximum sustainable yield in the long term. The basis is the F that provides maximum yield given the current assessment/advice error and biology and fishery parameters, constrained so that the long-term probability of SSB < $B_{lim}$ is ≤ 5% when applying the ICES MSY advice rule (AR):

$F = F_{MSY}$ (if SSB ≥ MSY $B_{trigger}$)
$F = F_{MSY} \times SSB/MSY\ B_{trigger}$ (if SSB < MSY $B_{trigger}$)

### MSY Btrigger

MSY Btrigger is a lower bound to the SSB when the stock is fished at FMSY. The point at which F is reduced when applying the ICES MSY advice rule (AR). The basis for MSY Btrigger is MSY Btrigger = maximum(Bpa, the 5th percentile of the distribution of SSB when fishing at FMSY), modified according to the scheme for determining MSY Btrigger (described in section on MSY reference points).

### $F_{P.05}$

$F_{P.05}$ is the value of F that provides an upper F limit that is considered precautionary for management plans and MSY rules. The basis is the value of F, including modification with biomass criteria that, if applied as target in the advice rule would lead to SSB ≥ $B_{lim}$ with a 95% probability. The derivation of $F_{P.05}$ should include expected stochastic variability in biology and fishery, as well as advice error.

In the following estimation process the ICES guideline will be followed to obtain estimates of these reference points.

## Assessment/prediction error

In practice there are always errors in assessment and predictions and are thus transferred into the advisory process it is. It is therefore important to perform simulations with realistic errors.

XSAM is a statistical model that can do stochastic projections. Thus, the model estimated values of both point estimates and their errors of F and SSB are available both in the assessment year and in the quota year. These values are conditioned on the model being able to correctly estimate the variability. A retrospective run of the assessment model will then provide time series of model estimates of errors both in assessment year and in quota year and thus inform values of assessment error to use.

A second alternative is to compare the historical assessment values of estimates and predictions, but this approach will depend on the actual model that was in play in the respective years, and the validity of using these numbers as basis for establishing the assessment error may therefore be questioned since they may not be representative of the current assessment model.

A third alternative is to do a retrospective run using the current assessment model XSAM. Then, for each run, the point estimate of predicted SSB and F for the advice year is extracted. The F in the assessment year correspond to the annual historic TACs and not the advised TACs provided by ICES. Assessment can then be calculated by comparing the output from the retrospective runs to the most recent assessment of the stock.

In this document the first and third option is considered.

# Results

## Time series of abundance

Figures 1-5 summarizes the XSAM estimates and compares with other estimates of recruits, fishing mortality, SSB and stock biomasses. Note that since XSAM starts at age 2, the recruits refers to number at age 2 and stock biomass to biomass of age 2+.

It is noted that XSAM produces lower estimates of abundances compared to SEASTAR and VPA. XSAM appears more comparable to TASACS in that respect for overlapping time series. The current XSAM configuration uses data from 2-12+. To examine whether the exclusion of ages 0-1 in the estimation resulted in lower estimates due to changes in initial values (i.e. the recruits) the model was run including ages 0-1 for the time periods 1950-2017 and 1988-2017. For these periods the inclusion of ages 0-1 made no practical difference concerning levels of biomasses or fishing mortalities for ages 2+ (not shown). With this basis it is concluded that the inclusion of ages 0-1 does not contribute with information on dynamics of herring and is omitted from the analysis.

To further examine potential causes for the discrepancies in levels of abundance prior to the collapse the predicted catches by the various methods are compared in Figure 6 and 7. Surprisingly, it is found that the reported VPA estimates by Toresen and Østvedt (2000) of F and N do not match the reported catch as the predictions are higher than what was reported, particularly in periods where the discrepancies in abundance (between VPA and XSAM) are largest. To further examine the discrepancies between particularly the VPA by Toresen and Østvedt (2000) and XSAM, a VPA was fitted to the entire time series of catch data. The discrepancies between this VPA and XSAM in the earliest part of the time series then disappear but remains in the period 1940-1950.

## SSB/R

Spawning stock-recruit relationship based on SSB and recruits at age 2 is not very clear (Figure 8 and 9), although not visually worse than previously considered relationships using recruits at younger age (previously age 0 has been used). The relationships are shown for the three different time periods with onset in 1907, 1950 and 1988 respectively. The time period starting in 1988 is too narrow as there is no information of recruitment for low abundances of SSB. Comparing the recruitment models segmented regression, Beverton-Holt and Ricker, there is no evidence in the data to make a conclusion on which one to use based on statistical significance (Figure 8 and 9 and Table 1). Overall, it is confirmed that the time window 1988-2017 is too short to establish any relationship such that the time window must be increased. By increasing the time window (1907-2017 or 1950-2017) the recruitment variability explained by the model is increased (r squared ~0.4 on log scale), but still indicate a high degree of environmental stochasticity. Note that the point estimate for the breakpoint in the segmented regression is ~2560 using data 1950-2017 while ~2670 using the entire time series 1907-2017 (Table 1) which is very close to the current $B_{lim}$ value 2500.

Figure 10 shows the spawning stock recruitment for the time series 1907-2017 and 1950-2017 along with the fitted segmented regression model. With segmented regression as basis, the time series dynamics of the residuals clearly depict dependence in time. More specifically, this suggest that the residuals follows an autoregressive (AR) model including dependency at time lag 1, an AR(1) model, possibly with a seasonal component at lag 13 (Figure 11), say a $\mathrm{SAR}(1)_{13}$ model, due to the significant correlation at this time lag. Note that this match the average distance between the peaks in recruitment seen in Figure 1. There is no immediate biological hypothesis for the mechanism explaining the dependency at lag 13 but a candidate hypothesis is that lags at higher lags may be caused by the age structure. Fitting the models 1) white noise $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, 2) AR(1) model and 3) AR(1) with seasonal dependence at lag 13 ($\mathrm{SAR}(1)_{13}$) it is noted that the $\mathrm{SAR}(1)_{13}$ has the lowest AIC (Table 2) while the ACF of the residuals for the AR(1) model still suggest dependency at lag 13 (not shown). The time series structures in residuals for the other models are qualitatively similar as those based on segmented regression. Despite the statistical significance: Due to the lack of biological mechanism supporting the lag 13 dependency, which can be generated by the underlying age structure, we will only consider the noise processes white noise and AR(1) models for the noise in recruitment.

To summarize, the period 1907-1949 does not appear to improve the stock recruitment relationship compared to using 1950-2017, mostly since it contains no further information on dynamics at low stock sizes. Moreover the validity of using the period 1907-1949 to describe the current dynamics can be questioned. Therefore, the rest of the results will be based on the dynamics as perceived in the time period 1950-2017. And for now we narrow the noise processes down to white noise and AR(1) models.

The uncertainties in the parameters are considerable (Table 3 and Figure 12). The ranking of models based on frequency of each model type selected (Table 4) is the same as the ranking obtained by AIC and $r^2$ and the model average ("AIC smoothed") is shown in Figure 12.

Table 1. Fitted recruitment models segmented regression, Beverton Holt and Ricker, for recruitment at age 2 as function of SSB two years before for different time periods. Estimates of recruitment and SSB are taken from XSAM estimates fitted to the corresponding time period. The models are fitted assuming log normal error and $r^2$ refer to the fit on the log-scale.

| Time period | Model | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Segmented regression | | | | Beverton Holt | | | | Ricker | | | |
| | Par | Estimate | AIC | $r^2$ | Par | Estimate | AIC | $r^2$ | Par | Estimate | AIC | $r^2$ |
| | $\alpha_{SR}$ | 3.588 | | | $\alpha_{BH}$ | $0.345 \times 10^9$ | | | $\alpha_R$ | 11.433 | | |
| 1988-2017 | $\beta_{SR}$ | 3359.580 | 76.618 | 0.034 | $\beta_{BH}$ | $29.105 \times 10^3$ | 77.59 | 0+ | $\beta_R$ | 0.0003 | 76.504 | 0.038 |
| | $\sigma_{SR}$ | 0.854 | | | $\sigma_{BH}$ | 0.869 | | | $\sigma_R$ | 0.852 | | |
| | $\alpha_{SR}$ | 3.322 | | | $\alpha_{BH}$ | 4.821 | | | $\alpha_R$ | 4.124 | | |
| 1950-2017 | $\beta_{SR}$ | 2559.733 | 235.417 | 0.384 | $\beta_{BH}$ | 0.0004 | 234.434 | 0.393 | $\beta_R$ | 0.0002 | 234.527 | 0.392 |
| | $\sigma_{SR}$ | 1.376 | | | $\sigma_{BH}$ | 1.366 | | | $\sigma_R$ | 1.367 | | |
| | $\alpha_{SR}$ | 3.705 | | | $\alpha_{BH}$ | 4.744 | | | $\alpha_R$ | 4.171 | | |
| 1907-2017 | $\beta_{SR}$ | 2668.794 | 356.978 | 0.424 | $\beta_{BH}$ | 0.0003 | 356.708 | 0.425 | $\beta_R$ | 0.0001 | 356.713 | 0.425 |
| | $\sigma_{SR}$ | 1.211 | | | $\sigma_{BH}$ | 1.209 | | | $\sigma_R$ | 1.209 | | |

Table 2. Fitted models to the residuals of fitted segmented regression model for recruits at age 2 versus spawning stock for the time period 1950-2017.

| Model | Estimates | AIC |
|---|---|---|
| White noise:<br>$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ | $\hat{\sigma}_\varepsilon^2 = 1.922$ | 231.42 |
| AR(1):<br>$\varepsilon_t = \alpha\varepsilon_{t-1} + \omega_t$<br>$\omega_t \sim N(0, \sigma_\omega^2)$ | $\hat{\alpha} = 0.3750\ (0.12)$<br>$\hat{\sigma}_\omega^2 = 1.633$ | 223.83 |
| SAR(1)$_{13}$:<br>$\varepsilon_t = \alpha\varepsilon_{t-1} + \beta\varepsilon_{t-13} + \omega_t$<br>$\omega_t \sim N(0, \sigma_\omega^2)$ | $\hat{\alpha} = 0.3872\ (0.12)$<br>$\hat{\beta} = 0.3151\ (0.12)$<br>$\hat{\sigma}_\omega^2 = 1.461$ | 219.84 |

Table 3. Summary of parameter estimates fitting recruitment models segmented regression, Beverton Holt and Ricker, for recruitment at age 2 as function of SSB two years before for the time period 1950-2017. Standard deviation and confidence intervals are estimating resampling pairs of stock recruitment data at random with replacement. The estimates are based on 1000 replicates.

| Model | Parameter | Estimate | SD | 95% CI | AIC | $r^2$ |
|---|---|---|---|---|---|---|
| Ricker | $\alpha_R$ | 4.124 | 1.396 | (2.197,7.534) | 234.527 | 0.392 |
| | $\beta_R$ | 0.0002 | 0.0001 | (0+,0.0003) | | |
| | $\sigma_R$ | 1.367 | 0.282 | (1.275,2.388) | | |
| Beverton Holt | $\alpha_{BH}$ | 4.821 | 2.201 | (2.288,10.699) | 234.434 | 0.393 |
| | $\beta_{BH}$ | 0.0004 | 0.0003 | (0.0001,0.0014) | | |
| | $\sigma_{BH}$ | 1.366 | 0.272 | (1.295,2.377) | | |
| Segmented regression | $\alpha_{SR}$ | 3.322 | 1.468 | (1.795,7.054) | 235.417 | 0.384 |
| | $\beta_{SR}$ | 2559.733 | 1414.322 | (990.998,5452.761) | | |
| | $\sigma_{SR}$ | 1.376 | 0.283 | (1.293,2.407) | | |

Table 4. Proportion of 1000 resamples of stock-recruitment pairs with lowest AIC values for models.

| Segmented Regression | Beverton Holt | Ricker |
|---|---|---|
| 0.242 | 0.452 | 0.306 |

## Evaluation of simulation model

To enable evaluation of the validity of simulations the first objective is to simulate population trajectories that mimic the history. First, fishing mortalities are generated according to the fitted XSAM model. In the configuration of the model, the model for F was divided into 1950-1987 and 1988-2017. We choose the model for F that corresponds to the most recent period. Stock and catch weights at age and proportion mature at age varies over time (Figures 13-15) and several selection procedures can be made (last value, average over selected years, randomized etc). We consider three options:

1) Long term mean
2) Weighted long term mean using stock numbers as weights
3) Average of last three years

The respective values are given in Tables 5-7 below.

Table 5. Mean weight at age in stock in 2017, average 2015-2017, average 1950-2017, average 1988-2017, weighted average 1950-2017 and weighted average 1988-2017. Weights in weighted averages are stock numbers at year and age.

| Age | 2017 | 2015-2017 | 1950-2017 | 1988-2017 | w1950-2017 | w1988-2017 |
|---|---|---|---|---|---|---|
| 2 | 0.054 | 0.051 | 0.053 | 0.040 | 0.044 | 0.036 |
| 3 | 0.115 | 0.123 | 0.116 | 0.107 | 0.093 | 0.091 |
| 4 | 0.190 | 0.188 | 0.191 | 0.170 | 0.172 | 0.159 |
| 5 | 0.247 | 0.246 | 0.249 | 0.231 | 0.217 | 0.213 |
| 6 | 0.282 | 0.291 | 0.288 | 0.274 | 0.256 | 0.256 |
| 7 | 0.322 | 0.323 | 0.316 | 0.307 | 0.284 | 0.291 |
| 8 | 0.338 | 0.330 | 0.339 | 0.328 | 0.307 | 0.314 |
| 9 | 0.351 | 0.350 | 0.356 | 0.347 | 0.326 | 0.339 |
| 10 | 0.359 | 0.353 | 0.374 | 0.358 | 0.337 | 0.351 |
| 11 | 0.361 | 0.356 | 0.388 | 0.371 | 0.347 | 0.368 |
| 12+ | 0.374 | 0.378 | 0.400 | 0.395 | 0.361 | 0.389 |

Table 6. Mean weight at age in catch in 2016, average 2014-2016, average 1950-2016, average 1988-2016, weighted average 1950-2016 and weighted average 1988-2016. Weights in weighted averages are stock numbers at year and age.

| Age | 2017 | 2015-2017 | 1950-2017 | 1988-2017 | w1950-2017 | w1988-2017 |
|---|---|---|---|---|---|---|
| 2 | 0.111 | 0.145 | 0.103 | 0.116 | 0.057 | 0.113 |
| 3 | 0.212 | 0.216 | 0.163 | 0.179 | 0.105 | 0.175 |
| 4 | 0.255 | 0.266 | 0.228 | 0.229 | 0.175 | 0.216 |
| 5 | 0.290 | 0.302 | 0.271 | 0.263 | 0.235 | 0.243 |
| 6 | 0.333 | 0.328 | 0.304 | 0.294 | 0.273 | 0.276 |
| 7 | 0.339 | 0.347 | 0.325 | 0.316 | 0.301 | 0.306 |
| 8 | 0.361 | 0.361 | 0.352 | 0.338 | 0.324 | 0.326 |
| 9 | 0.367 | 0.369 | 0.376 | 0.353 | 0.342 | 0.344 |
| 10 | 0.370 | 0.373 | 0.385 | 0.368 | 0.352 | 0.358 |
| 11 | 0.381 | 0.379 | 0.396 | 0.370 | 0.365 | 0.369 |
| 12+ | 0.380 | 0.381 | 0.396 | 0.387 | 0.381 | 0.388 |

Table 7. Proportion mature at age in 2017, average 2015-2017, average 1950-2017, average 1988-2017, weighted average 1950-2017 and weighted average 1988-2017. Weights in weighted averages are stock numbers at year and age.

| Age | 2017 | 2015-2017 | 1950-2017 | 1988-2017 | w1950-2017 | w1988-2017 |
|---|---|---|---|---|---|---|
| 2 | 0.0 | 0.0 | 0.000 | 0.000 | 0.000 | 0.000 |
| 3 | 0.0 | 0.0 | 0.029 | 0.010 | 0.004 | 0.002 |
| 4 | 0.4 | 0.4 | 0.387 | 0.307 | 0.177 | 0.174 |
| 5 | 0.8 | 0.8 | 0.747 | 0.753 | 0.606 | 0.653 |
| 6 | 1.0 | 1.0 | 0.947 | 0.957 | 0.882 | 0.927 |
| 7 | 1.0 | 1.0 | 0.985 | 0.997 | 0.950 | 0.992 |
| 8 | 1.0 | 1.0 | 1.000 | 1.000 | 1.000 | 1.000 |
| 9 | 1.0 | 1.0 | 1.000 | 1.000 | 1.000 | 1.000 |
| 10 | 1.0 | 1.0 | 1.000 | 1.000 | 1.000 | 1.000 |
| 11 | 1.0 | 1.0 | 1.000 | 1.000 | 1.000 | 1.000 |
| 12+ | 1.0 | 1.0 | 1.000 | 1.000 | 1.000 | 1.000 |

Since it is difficult to conclude on functional form of stock recruitment we consider all three considered models (Beverton-Holt, Ricker and segmented Regression). Finally, we compare the time varying selectivity according to XSAM with average fishing mortality (i.e. implies constant fishing pattern) over the time period.

Initial trials shows that all recruitment models occasionally generate recruits higher than observed (Figure 16) and particularly the segmented regression model. This indicates that the corresponding log-normal distribution for variation in the number of recruits has a too heavy tail to be realistic. Consequently, both mean and variability in recruitment becomes too high, particularly for the segmented regression model. It is therefore necessary to restrict the range of possible recruitment in the simulations to maintain realistic values (see also Simmonds et al. 2011 and Lillegård et al. 2005). Here we consider restricting maximum recruitment to the highest observed and 20% above highest observed (see Figure 17 for illustration).

For combinations of the various options we compare mean and standard deviation of fishing mortality at age, recruitment, total stock numbers, SSB and total catches (Table 8). All considered configurations are similar to the observations, but evident that combinations that includes:

- Weighted long term means of stock and catch weights at age and proportion mature at age
- Serial correlation in noise structure in stock recruitment
- Ricker or Beverton Holt recruitment models

resembles the historical estimates closest. Note that all configurations somewhat over estimates SSB and variability in SSB and consequently increases recruitment variability and variability in stock sizes and catches. Using weighted long term means reduces the overestimation and is an indication that these parameters may vary with density. Note that based on this simulation experiment it is indicated that Ricker or Beverton Holt recruitment models may be preferred. This is in line with the weak statistical evidence when fitting the models to data (Tables 1 and 3). Furthermore, it supports that serial correlation in the noise of the recruitment process should be included. Note that the effect of time varying versus constant selectivity appears small.

In summary: as basis for simulations to estimate reference point the following configuration is used throughout this document

- Stock recruitment models fitted to data 1950-2017
- Weighted long term means of stock and catch weights at age and proportion mature at age
- Serial correlation in noise structure in stock recruitment as AR(1)
- Time varying selectivity in F according to the fitted model in XSAM
- Due to the weak evidence for choosing stock recruitment model, all three will be considered as well as AIC-smoothed estimates.

Table 8. Comparing moments (average and standard deviation) of estimates (1950-2017 and 1988-2017) with simulated values for recruitment ($R_2$), total stock size ($N$), spawning stock biomass (SSB) and total annual catch ($C$) using Recruitment models (Rec model) Ricker, Beverton Holt (BH) and segmented regression (HS) with different assumptions on residual dynamics (Res type) independent identical distributed (iid), following an AR(1) model, different values of stock size-, catch size- and proportion mature-at age weighted long term mean (wltm weighted with stock numbers), average last 3 years (m 3 y) and long term mean (ltm), restricting recruitment (maxR) to maximum observed m and 20% higher than maximum observed 1.2m, and variable F corresponding to the process estimated by XSAM (Y) and a constant F corresponding to long term mean at age (N). Moments of simulated values are based on 10 000 time steps

| Data | Rec model | Res type | Par Type | maxR | Variable F | $\bar{R}_2$ | $SD(R_2)$ | $\bar{N}$ | $SD(N)$ | $\overline{SSB}$ | $SD(SSB)$ | $\bar{C}$ | $SD(C)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1950-2017* | – | – | – | – | – | 12482.36 | 19596.53 | 35773.41 | 29755.59 | 3771.59 | 2798.00 | 710.28 | 584.36 |
| 1988-2017* | – | – | – | – | – | 16213.15 | 16840.83 | 45999.04 | 22579.84 | 4590.46 | 1541.57 | 777.30 | 492.30 |
| Sim | Ricker | Iid | w ltm | 1.2m | Y | 16462.25 | 25132.74 | 48339.76 | 31010.94 | 5180.92 | 2864.68 | 708.36 | 513.05 |
| Sim | Ricker | AR(1) | w ltm | 1.2m | Y | 15123.99 | 24031.56 | 44668.39 | 34918.29 | 4846.14 | 3502.92 | 639.57 | 561.91 |
| Sim | Ricker | $SAR(1)_{13}$ | w ltm | 1.2m | Y | 15014.43 | 24938.51 | 44314.50 | 38703.61 | 4802.74 | 3944.68 | 635.12 | 617.85 |
| Sim | Ricker | Iid | m 3 y | 1.2m | Y | 16663.00 | 25393.62 | 48899.29 | 30632.79 | 6355.92 | 3369.38 | 829.91 | 592.54 |
| Sim | Ricker | AR(1) | m 3 y | 1.2m | Y | 15661.52 | 24717.03 | 46193.07 | 34924.92 | 6056.85 | 4182.39 | 767.82 | 659.34 |
| Sim | Ricker | $SAR(1)_{13}$ | m 3 y | 1.2m | Y | 15615.31 | 25343.13 | 46074.06 | 37655.70 | 6046.18 | 4581.80 | 764.56 | 704.96 |
| Sim | Ricker | Iid | ltm | 1.2m | Y | 16694.89 | 25415.51 | 48991.94 | 30675.68 | 6371.08 | 3317.09 | 784.01 | 559.80 |
| Sim | Ricker | AR(1) | ltm | 1.2m | Y | 15701.23 | 24739.07 | 46307.97 | 34983.12 | 6078.52 | 4126.22 | 725.69 | 623.37 |
| Sim | Ricker | $SAR(1)_{13}$ | ltm | 1.2m | Y | 15656.29 | 25348.34 | 46192.02 | 37651.74 | 6068.70 | 4518.67 | 722.57 | 665.07 |
| Sim | Ricker | Iid | w ltm | m | Y | 16126.94 | 23311.57 | 47354.88 | 28846.34 | 5075.25 | 2686.77 | 693.97 | 486.69 |
| Sim | Ricker | AR(1) | w ltm | m | Y | 14826.38 | 22313.05 | 43777.39 | 32683.99 | 4747.06 | 3291.84 | 627.35 | 535.60 |
| Sim | Ricker | $SAR(1)_{13}$ | w ltm | m | Y | 14682.78 | 23110.17 | 43333.48 | 36331.09 | 4695.80 | 3729.50 | 621.22 | 587.50 |
| Sim | Ricker | Iid | m 3 y | m | Y | 16346.87 | 23555.49 | 47971.36 | 28461.18 | 6235.20 | 3148.81 | 814.20 | 562.56 |
| Sim | Ricker | AR(1) | m 3 y | m | Y | 15347.19 | 22862.73 | 45252.22 | 32516.39 | 5930.61 | 3906.66 | 753.05 | 627.54 |
| Sim | Ricker | $SAR(1)_{13}$ | m 3 y | m | Y | 15282.12 | 23452.54 | 45085.62 | 35240.66 | 5915.26 | 4315.81 | 748.50 | 669.50 |
| Sim | Ricker | Iid | ltm | m | Y | 16377.36 | 23572.45 | 48059.92 | 28495.43 | 6249.73 | 3102.83 | 769.13 | 531.03 |
| Sim | Ricker | AR(1) | ltm | m | Y | 15386.99 | 22882.32 | 45367.89 | 32571.88 | 5952.18 | 3857.62 | 711.74 | 592.74 |
| Sim | Ricker | $SAR(1)_{13}$ | ltm | m | Y | 15325.20 | 23468.66 | 45210.82 | 35257.74 | 5938.72 | 4263.39 | 707.51 | 631.50 |
| Sim | BH | Iid | w ltm | 1.2m | Y | 16109.71 | 24737.23 | 47364.29 | 31790.59 | 5087.58 | 3036.33 | 691.63 | 517.03 |
| Sim | BH | AR(1) | w ltm | 1.2m | Y | 14987.14 | 23762.16 | 44358.20 | 36260.59 | 4831.38 | 3741.56 | 630.97 | 574.69 |
| Sim | BH | $SAR(1)_{13}$ | w ltm | 1.2m | Y | 15434.16 | 25526.16 | 45646.01 | 42544.00 | 4965.81 | 4478.62 | 649.94 | 669.30 |
| Sim | BH | Iid | m 3 y | 1.2m | Y | 17520.68 | 26087.47 | 51491.25 | 33068.52 | 6707.38 | 3787.49 | 869.70 | 630.39 |
| Sim | BH | AR(1) | m 3 y | 1.2m | Y | 16537.48 | 25269.60 | 48901.42 | 37985.25 | 6437.97 | 4697.20 | 805.88 | 706.21 |
| Sim | BH | $SAR(1)_{13}$ | m 3 y | 1.2m | Y | 16946.38 | 26754.31 | 50107.10 | 43713.54 | 6596.80 | 5524.02 | 825.33 | 805.77 |
| Sim | BH | Iid | ltm | 1.2m | Y | 17546.66 | 26109.86 | 51567.73 | 33096.15 | 6721.33 | 3750.44 | 821.50 | 595.18 |
| Sim | BH | AR(1) | ltm | 1.2m | Y | 16580.71 | 25312.99 | 49030.17 | 38048.70 | 6464.03 | 4662.32 | 761.87 | 667.61 |
| Sim | BH | $SAR(1)_{13}$ | ltm | 1.2m | Y | 16992.76 | 26792.54 | 50245.41 | 43765.76 | 6624.77 | 5496.88 | 780.18 | 760.13 |
| Sim | BH | Iid | w ltm | m | Y | 15635.26 | 22833.57 | 45965.34 | 29426.63 | 4936.49 | 2830.60 | 671.40 | 487.05 |
| Sim | BH | AR(1) | w ltm | m | Y | 14571.72 | 21976.36 | 43113.39 | 33712.34 | 4692.76 | 3486.97 | 613.96 | 542.52 |
| Sim | BH | $SAR(1)_{13}$ | w ltm | m | Y | 14923.60 | 23454.03 | 44135.29 | 39445.89 | 4801.18 | 4172.28 | 628.51 | 627.21 |
| Sim | BH | Iid | m 3 y | m | Y | 16990.41 | 24034.35 | 49929.81 | 30537.94 | 6503.36 | 3518.62 | 843.51 | 593.48 |
| Sim | BH | AR(1) | m 3 y | m | Y | 16063.98 | 23335.26 | 47484.12 | 35262.71 | 6247.83 | 4370.29 | 783.54 | 667.39 |
| Sim | BH | $SAR(1)_{13}$ | m 3 y | m | Y | 16374.90 | 24524.37 | 48413.94 | 40412.00 | 6373.07 | 5129.93 | 797.68 | 754.06 |
| Sim | BH | Iid | ltm | m | Y | 17013.49 | 24051.86 | 49998.03 | 30560.77 | 6516.15 | 3487.24 | 796.64 | 559.92 |
| Sim | BH | AR(1) | ltm | m | Y | 16103.18 | 23369.16 | 47600.83 | 35313.94 | 6271.82 | 4338.89 | 740.61 | 630.43 |
| Sim | BH | $SAR(1)_{13}$ | ltm | m | Y | 16419.01 | 24560.42 | 48545.70 | 40457.87 | 6399.81 | 5106.83 | 754.02 | 711.14 |
| Sim | HS | Iid | w ltm | 1.2m | Y | 18108.30 | 26771.56 | 53175.28 | 32902.09 | 5699.89 | 3055.46 | 779.02 | 550.65 |
| Sim | HS | AR(1) | w ltm | 1.2m | Y | 16969.98 | 26096.76 | 50116.47 | 38683.52 | 5436.21 | 3913.08 | 717.77 | 628.21 |
| Sim | HS | $SAR(1)_{13}$ | w ltm | 1.2m | Y | 16757.58 | 27060.13 | 49453.66 | 44038.93 | 5358.49 | 4582.92 | 708.84 | 702.81 |
| Sim | HS | Iid | m 3 y | 1.2m | Y | 18484.12 | 27074.05 | 54269.08 | 32925.44 | 7059.22 | 3673.84 | 919.49 | 640.21 |
| Sim | HS | AR(1) | m 3 y | 1.2m | Y | 17783.99 | 26682.24 | 52474.64 | 38768.71 | 6884.86 | 4708.87 | 871.36 | 738.33 |
| Sim | HS | $SAR(1)_{13}$ | m 3 y | 1.2m | Y | 17876.87 | 27748.36 | 52756.76 | 43779.54 | 6924.66 | 5466.23 | 874.88 | 817.76 |
| Sim | HS | Iid | ltm | 1.2m | Y | 18490.25 | 27076.21 | 54286.31 | 32928.30 | 7065.23 | 3622.05 | 867.30 | 603.45 |
| Sim | HS | AR(1) | ltm | 1.2m | Y | 17806.52 | 26701.89 | 52540.13 | 38796.83 | 6901.49 | 4653.99 | 822.55 | 696.72 |
| Sim | HS | $SAR(1)_{13}$ | ltm | 1.2m | Y | 17906.37 | 27754.85 | 52843.40 | 43766.09 | 6944.82 | 5414.83 | 826.00 | 770.41 |
| Sim | HS | Iid | w ltm | m | Y | 17608.02 | 24698.13 | 51706.38 | 30421.41 | 5542.51 | 2849.29 | 757.49 | 519.31 |
| Sim | HS | AR(1) | w ltm | m | Y | 16477.40 | 23996.08 | 48650.08 | 35844.73 | 5274.73 | 3644.24 | 697.35 | 592.08 |
| Sim | HS | $SAR(1)_{13}$ | w ltm | m | Y | 16217.96 | 24843.51 | 47857.45 | 40889.55 | 5184.60 | 4281.03 | 686.19 | 660.31 |
| Sim | HS | Iid | m 3 y | m | Y | 17972.53 | 24965.89 | 52765.83 | 30396.06 | 6863.44 | 3412.98 | 894.10 | 603.81 |
| Sim | HS | AR(1) | m 3 y | m | Y | 17266.78 | 24516.12 | 50936.12 | 35840.10 | 6680.35 | 4370.66 | 846.58 | 695.86 |
| Sim | HS | $SAR(1)_{13}$ | m 3 y | m | Y | 17312.67 | 25473.34 | 51083.55 | 40565.40 | 6703.18 | 5091.18 | 847.67 | 768.32 |
| Sim | HS | Iid | ltm | m | Y | 17979.01 | 24970.57 | 52784.21 | 30400.59 | 6869.50 | 3368.05 | 843.36 | 568.82 |
| Sim | HS | AR(1) | ltm | m | Y | 17288.20 | 24533.00 | 50998.34 | 35864.89 | 6696.15 | 4322.09 | 799.15 | 656.33 |
| Sim | HS | $SAR(1)_{13}$ | ltm | m | Y | 17342.34 | 25480.99 | 51170.77 | 40552.61 | 6723.03 | 5045.93 | 800.37 | 723.73 |
| Sim | Ricker | Iid | w ltm | 1.2m | N | 16528.63 | 25070.62 | 46775.20 | 30450.74 | 4629.89 | 2431.87 | 781.42 | 371.27 |
| Sim | Ricker | AR(1) | w ltm | 1.2m | N | 14956.82 | 23673.86 | 42325.78 | 33922.91 | 4189.43 | 2977.92 | 707.13 | 462.21 |
| Sim | Ricker | $SAR(1)_{13}$ | w ltm | 1.2m | N | 14782.95 | 24878.96 | 41833.35 | 38564.02 | 4139.68 | 3481.49 | 698.68 | 548.48 |
| Sim | BH | Iid | m 3 y | 1.2m | N | 15492.46 | 23996.61 | 43842.53 | 29966.20 | 4339.52 | 2429.04 | 732.41 | 376.80 |
| Sim | BH | AR(1) | m 3 y | 1.2m | N | 14173.58 | 22784.03 | 40112.90 | 33682.10 | 3970.86 | 2981.90 | 670.25 | 469.47 |
| Sim | BH | $SAR(1)_{13}$ | m 3 y | 1.2m | N | 14632.91 | 24774.03 | 41410.45 | 40094.69 | 4098.16 | 3651.78 | 691.68 | 584.84 |
| Sim | HS | Iid | ltm | 1.2m | N | 18262.75 | 26818.73 | 51682.92 | 32175.09 | 5115.69 | 2548.17 | 863.41 | 386.96 |
| Sim | HS | AR(1) | ltm | 1.2m | N | 16841.78 | 25810.79 | 47660.68 | 37325.53 | 4717.37 | 3287.88 | 796.24 | 511.24 |
| Sim | HS | $SAR(1)_{13}$ | ltm | 1.2m | N | 16166.87 | 26839.52 | 45744.88 | 43078.36 | 4526.11 | 3925.85 | 763.89 | 623.69 |

*Estimates

## Stochastic simulations and estimation of reference points

Based on the configuration model above we proceed with stochastic simulations accounting for variability in parameter estimates as well as the AIC-smoothed approach for stock recruitment relationship similar to Simmons et al. 2011.

### Number of iterations

For simulations with fixed stock recruitment function and parameters it is only necessary to run the simulation one time for many time steps to display the mean responses. In this case it is found that running the model for 10000 time steps, discarding the first 250, is sufficient to have stabilizes the respsonses.

The necessary number of replicates for stochastic simulations is determined for the model averaging case, as this case will represent more sources of variability than cases where the functional from of stock recruitment is fixed (although parameters varies). For this case, using the variable selection pattern, setting other biological parameters equal to the long term weighted mean and fishing with constant F we find that the mean and median values of Recruitment, SSB, catch and probability of falling below 2500 tons displays less than 2% difference from the mean of the simulations when $n$ and $T_{tot}$ is larger than ~500 (Figure 18). Note that this figure shows results for F=0.2 which is a value of F in the proximity of the F that maximizes the yield when no advice error is present. A larger number is required for the percentiles in the distributions, but note that the measure for the lower percentiles may be misleading when boundaries such as 0 or 1 (e.g. for proportions) are met; for example small values close to 0 may show small differences on absolute scale but appear large on relative scale (see e.g. performance statistics for lower 5% percentile). The robustness of the estimates is confirmed by comparing the statistic as a function of F's for one simulation based on $n = T_{tot} = 1000$ and one based on $n = T_{tot} = 500$, i.e. the mean and median values are very similar while larger differences are detected for the percentiles in the distributions (in all cases $T_{tot} = 250$) (Figure 19). This means that setting $n = T_{tot} = 500$ is sufficient to obtain a relatively robust estimate of mean and median values of Recruitment, SSB, catch and risk values, but estimates of percentiles must be treated with more caution. The difference in runtime for the code used here for the two selections of sample sizes is ~1 hour for the large samples size and ~15 mins for the small sample size when each is run over 10 different values of F. For the rest of the analysis presented here $n = T_{tot} = 500$ will be used unless otherwise stated.

## Basis for estimation of reference points

First a simulation is performed with the same criteria as for identifying Flim: The simulation is conducted based on a fixed F (i.e. without inclusion of a Btrigger) using one of the described models without inclusion of assessment/advice errors.  In all simulations uncertainty in parameters are accounted for and for the AIC-smoothed estimate the functional form is also selected for each resample.

It is noted that estimates of both mean recruitment, SSB, annual yield and risk are highly imprecise as percentiles in the distributions are wide and overlapping comparing different functional forms for stock recruitment. Secondly, it is noted that point estimates of mean recruitment and SSB are somewhat different depending on functional form of stock recruitment, particularly for low Fs (Figure 20), and the AIC smoothed estimates provides an "average" of the respective functional forms as expected. Values of F that maximizes mean annual yield or F that gives P(SSB<Blim) appear

less dependent on functional form as they can hardly be visually separated. It is further noted that P(SSB<2500)>0 even for no fishing reflecting the highly variable and imprecise estimates of recruitment variability.

If $B_{lim}$ is set to 2500, this implies that $F_{lim}$ is slightly larger than 0.2.

For all stock recruitment models except for the AIC-smoothed version, Flim may be estimated ignoring uncertainty in the parameters. It is noted that the corresponding result is very similar to the mean values shown in Figure 20, and thus not included for comparison.

In summary, the AIC smoothed version of Stock recruitment is taken forward to estimate reference points.

### Sensitivity to assumption on serial correlation

The effect of accounting for serial correlation versus assuming no serial correlation in the noise process for recruitment is compared in Figure 21. The estimates are similar, but note that if there is no serial dependence risk is reduced.

### Sensitivity to assumptions on biological parameters

The initial simulations for evaluating the simulation model suggested that the long term weighted average results in mean and variance values of key parameters that were closest to the actual data when fishing according to the model as both unweighted long term mean and last year average produces on average too high mean values and variability. To explore the effect on the results of the simulations is compared replacing the biological parameters with the average of the last 3 years (Figure 22). This average represents larger weights at age and 'faster' maturation at age. As indicated by the initial simulations, this will lead to a, on average, larger and more variable stock than using the weighted mean. Consequently, recruitment and SSB will be higher, giving higher catches and lower risks, and thus significantly changes the estimate of the reference point.

### Sensitivity to weighted vs unweighted F

All results have been shown using F as the unweighted F over ages 5-11. If F is replaced by the weighted F (Figure 23) note that stock sizes are reduced and risk increased.

### Assessment/prediction error

The XSAM model conditioned estimates of assessment/prediction error (ref 'option 1') is shown in Table 9 and Figure 24. This is based on a retrospective run from 2002 and onwards. The average relative standard error of the SSB and F for the quota year (i.e. the values entering the HCR) is 0.167 and 0.260, respectively. Using the deviation from point estimates (ref 'option 3') from the retrospective run gives somewhat lower values; 0.101 and 0.152, respectively. Note that the former accounts for the full uncertainty defined by the model, while the latter is based on point estimates only, and it should thus not be surprising that these are lower. Provided the assumption that point estimates alone can inform the assessment error, it is possible to examine trends and time dependencies in the residuals (Figure 25). Here it is clear that, on average, the predictions overestimate SSB, while F is underestimated although the means are not significantly different from 0. The residuals are negatively correlated and there may be some positive serial correlation in the error for the predicted values of F. However, none of the correlations are significant. The residuals show no dependency to mean SSB or F (not shown).

In summary, the prediction error for SSB correspond to a RSE ~0.1-0.17 while for F RSE ~0.15-0.26 depending on which approach to rely on. Note that these magnitudes of error are in line with the deviations between historical estimates and predicted values seen in Figure 24 and are thus candidates to use directly as iid errors for the assessment/prediction error.

If it is decided to base prediction error on the point estimates (i.e. the third option), it is possible to account both for the possible dependency between the error in log F and log SSB as well as potential serial correlation for log F by first fitting the model

$$\varepsilon_{Ft} = \rho\varepsilon_{Ft-1} + \omega_t$$

to the residuals for log F where $\omega_t \sim N(0, \sigma_\omega^2)$, and then fit

$$\varepsilon_{SSBt} = \gamma\varepsilon_{Ft} + \varphi_t$$

where $\varphi_t \sim N(0, \sigma_\varphi^2)$ to the residuals for log SSB. This approach will induce some serial correlation in $\varepsilon_{SSBt}$, but less than for $\varepsilon_{Ft}$ (as evident comparing the ACF's in Figure 25), and is thus likely to mimic the empirical data. In the simulations, $\varepsilon_{Ft}$ must be generated first, and then $\varepsilon_{SSBt}$ can be generated with $\varepsilon_{Ft}$ as input.

Estimates of the parameters are $\hat{\rho} = 0.431$ and $\hat{\sigma}_\omega = 0.188$ for the residuals for log F, and $\hat{\gamma} = 0.3569$ and $\hat{\sigma}_\varphi = 0.08869$.

Table 9. The historical estimates of SSB and F from the 2017 assessment and the corresponding estimates in the assessment year (subscript AY) and quota year (subscript QY). The corresponding estimates of relative standard error (also called CV but specific to estimates of precision) is denoted with rse obtained by a retrospective run including predictions of XSAM from 2002-2017.

| year | SSB | $SSB_{AY}$ | $SSB_{QY}$ | F | $F_{AY}$ | $F_{QY}$ | rseSSB | $rseSSB_{AY}$ | $rseSSB_{QY}$ | rseFW | $rseFA_{AY}$ | $rseF_{QY}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2002 | 3501.576 | 3896.717 | NA | 0.226 | 0.200 | --- | 0.057 | 0.238 | --- | 0.110 | 0.321 | --- |
| 2003 | 4128.285 | 4345.190 | 4734.770 | 0.143 | 0.125 | 0.117 | 0.054 | 0.211 | 0.261 | 0.111 | 0.243 | 0.403 |
| 2004 | 5213.377 | 5868.047 | 5539.040 | 0.105 | 0.113 | 0.125 | 0.053 | 0.187 | 0.225 | 0.120 | 0.270 | 0.296 |
| 2005 | 5338.671 | 6431.493 | 6520.471 | 0.150 | 0.135 | 0.125 | 0.055 | 0.152 | 0.207 | 0.123 | 0.236 | 0.347 |
| 2006 | 5307.069 | 5993.947 | 5978.549 | 0.161 | 0.145 | 0.125 | 0.055 | 0.131 | 0.168 | 0.120 | 0.194 | 0.274 |
| 2007 | 6823.548 | 8370.166 | 7140.166 | 0.154 | 0.127 | 0.125 | 0.052 | 0.122 | 0.149 | 0.104 | 0.151 | 0.266 |
| 2008 | 6907.899 | 8127.713 | 8603.169 | 0.201 | 0.161 | 0.125 | 0.055 | 0.117 | 0.137 | 0.105 | 0.145 | 0.165 |
| 2009 | 6864.865 | 7962.193 | 8222.629 | 0.207 | 0.178 | 0.125 | 0.060 | 0.127 | 0.133 | 0.099 | 0.139 | 0.152 |
| 2010 | 6047.976 | 6311.243 | 7364.610 | 0.208 | 0.208 | 0.125 | 0.069 | 0.153 | 0.148 | 0.112 | 0.162 | 0.177 |
| 2011 | 5637.568 | 6141.125 | 5776.339 | 0.157 | 0.147 | 0.125 | 0.077 | 0.138 | 0.178 | 0.118 | 0.181 | 0.209 |
| 2012 | 5416.111 | 4592.594 | 5642.560 | 0.145 | 0.180 | 0.125 | 0.084 | 0.184 | 0.154 | 0.127 | 0.199 | 0.226 |
| 2013 | 5018.936 | 4552.735 | 4471.812 | 0.131 | 0.168 | 0.109 | 0.088 | 0.141 | 0.205 | 0.129 | 0.173 | 0.296 |
| 2014 | 4819.419 | 4175.632 | 4214.796 | 0.091 | 0.102 | 0.101 | 0.093 | 0.133 | 0.169 | 0.139 | 0.190 | 0.290 |
| 2015 | 4526.983 | 4773.173 | 3857.999 | 0.071 | 0.069 | 0.091 | 0.096 | 0.114 | 0.146 | 0.158 | 0.156 | 0.239 |
| 2016 | 4265.679 | 4438.162 | 4608.355 | 0.084 | 0.078 | 0.113 | 0.098 | 0.104 | 0.119 | 0.157 | 0.228 | 0.278 |
| 2017 | 4130.827 | 4130.827 | 4316.601 | 0.191 | 0.191 | 0.104 | 0.100 | 0.100 | 0.111 | 0.222 | 0.222 | 0.278 |
| Aver. | 5246.799 | 5631.935 | 5799.458 | 0.152 | 0.145 | 0.117 | 0.072 | 0.147 | 0.167 | 0.128 | 0.201 | 0.260 |

## Estimation of reference points

Except for Fpa and Bpa, the other reference points are found based on statistics as a function for F which is achieved by running the simulations over a range of specific values for F. Consequently optimum or intersecting values of the parameters may have their solution for Fs in between actual values and thus cause numerical instabilities in estimates of reference points or risk. To reduce this problem, values (mean and percentiles) for each parameter are smoothed using the R-function loess.

The original estimates are compared with the smoothed value in Figure 26, and since the two versions match very well for the specific values for F used in the simulations, the smoothed values will be used as basis for estimating optimums and intersecting points of the curves.

The mean of the simulated long-term yield can have undesirable properties when yield distributions are highly skewed (with a high proportion of values in the tails of the distribution) and may occasionally contain very large values. The ICES guidelines therefore advice to use median of the distribution at each F is often considered to be more robust to these issues. This approach is followed here.

In the stochastic simulations time series are simulated for each set of parameters. Each time series gives a mean or median of performance (i.e. recruitment, SSB, annual yield and), and hence the mean and quantiles of the distributions is based on the distribution of these statistics. Then, as a point estimate one can for example choose the median or mean (e.g. mean of medians or median of medians), which both can be justified. It is not always clear which one to choose, as they will differ if the distribution is skewed. Therefore both candidates are presented in the following tables.

### $B_{lim}$

The results from the segmented regression (Table 1 and Table 3) give a value close 2500, which is the existing $B_{lim}$, as a candidate for $B_{lim}$ although the estimate is imprecise. Therefore, with no other obvious options, this value will be used as $B_{lim}$ throughout this document.

### $F_{lim}$

Doing a simulation with no assessment/prediction error and no Btrigger point establishes the $F_{lim}$ value. From Figure 27 we see that the F that estimates of the proportion of SSB being below 2500 in a specific year is very variable, but the mean and median values are 0.216 and 0.234. The average of the two is rounded to 0.22 and we set $F_{lim}$=0.22 in the rest of this document

### Bpa and Fpa

Using $B_{pa} = B_{lim} \times \exp(1.645 \times \sigma)$ where σ is the standard deviation of ln(SSB) at the start of the year following the terminal year of the assessment, then Table 9 and Figure 25 gives two optional values for σ (since for a log-normal distribution we have that $RSE = \sqrt{exp(\sigma^2) - 1} \approx \sigma$, where $\sigma^2$ is the variance on the log-scale), namely 0.101 or 0.167 (the average over the years in the retrospective run) depending on the basis. Consequently, if $B_{lim}$=2500, then $B_{pa}$ ranges from 2952 to 3290. The average of the two is ~3121 and a somewhat pragmatic choice of $B_{pa}$ is 3100.

Similarly, using $F_{pa} = F_{lim} \times \exp(-1.645 \times \sigma)$ where σ is the standard deviation of ln(F) in the terminal year of the assessment, we have that σ ranges from 0.152 to 0.201. Using $F_{lim}$=0.22 as basis, $F_{pa}$ ranges from 0.158 to 0.171 with an average of 0.164. This is close to the current $F_{pa}$=0.15, and a pragmatic choice is to keep the existing.

### $F_{MSY}$ and $B_{trigger}$

According to the guidelines $B_{trigger}$ is set as the maximum value of $B_{pa}$ and the 5% percentile of SSB when fishing at the F that maximizes annual yield in presence of assessment/prediction error.

With 2 optional values for error in realized F two simulations were run with each respective error (Table 10).

Table 10. Median values of $F_{MSY}$ for different assessment error. The table reports two figures for each: Mean of medians, and median of medians. The corresponding annual catch is given in parenthesis.

| Assessment error cv F | $F_{msy}$ mean | $F_{msy}$ median |
|---|---|---|
| 0 | 0.145 (591) | 0.167 (607) |
| 0.15 | 0.136 (591) | 0.152 (607) |
| 0.26 | 0.133 (578) | 0.151 (574) |

Including assessment error result in a decrease in point estimates of $F_{msy}$ (and corresponding catch) but the reduction is rather marginal. Despite the reduction due to assessment error, the estimates are still close to, and not significantly different from, 0.15 which is lower than $F_{pa}$, and it is therefore suggested to keep $F_{msy}$=0.15.

When simulating with a constant F=0.15 without assessment error, $B_{trigger}$ should be set as the maximum of $B_{pa}$ and the achieved 5th percentile of the distribution of SSB. The distribution of the corresponding 5th percentile of the distribution of SSB is shown in Figure 28 and both the mean and median value is lower than $B_{pa}$=3100 which means $B_{trigger}$=$B_{pa}$=3100.

### Fp05

The value of F, including modification with biomass criteria that, if applied as target in the MSY advice rule would lead to SSB ≥ $B_{lim}$ with a 95% probability including advice error is shown in Table 11 for the different candidates of advice error and Figure 29 for CV F=0.26 and CV SSB=0.16.

Table 11. Mean and median values of $F_{p05}$ for different values of error in F and SSB.

| CV F | CV SSB | Mean $F_{p05}$ | Median $F_{p05}$ |
|---|---|---|---|
| 0 | 0 | 0.084 | 0.101 |
| 0.15 | 0.1 | 0.078 | 0.094 |
| 0.26 | 0.1 | 0.077 | 0.094 |
| 0.15 | 0.16 | 0.078 | 0.095 |
| 0.26 | 0.16 | 0.078 | 0.091 |

The estimates of $F_{p05}$ implies estimating a statistic which has low probability ($\mathrm{P}(\mathrm{SSB} \leq \mathrm{B_{lim}}) = 0.05$). This means that the estimates presented here will suffer from monte carlo error to a larger degree than the other reference points considered in this document, and in order to increase the precision in the estimates, it the sample sizes and length of each time series should be considered increased. However, for the levels of advice error considered here the estimated Fp05 is stable and decreased compared to having no advice error. The low values are not surprising since even without fishing, there is a small but positive probability of SSB falling below 2500.

## Summary and conclusions

In WGWIDE XSAM is configured for the time series 1988- and ages 2-12+. This was decided in connection with the last benchmark WKPELA in 2016 with basis in the onset of survey series used in the assessment. XSAM can in principle be fitted to catch at age data only, but this has until now been beyond the scope. Since this model is a forward running model, the current state depends on the initial states. In other words, the exact starting year may influence the estimates of the current state of the stock. Here XSAM is fitted to longer time series utilizing catch at age data from 1907—and compared to other historical estimates of the stock as well as the current assessment. It is found that XSAM estimates lower abundances prior to 1955 and after mid-eighties compared to the VPA by Toresen and Østvedt (2000) and SEASTAR, but matches very well with TASACS from 1988 onwards. Furthermore it is found that this VPA and SEASTAR predicts higher catches than reported in some time periods (particularly in late 30'ties and just prior to the collapse 67-68) which can explain the deviation in levels for stock abundances. Due to the discrepancies it was decided to run a VPA on the entire time series of catch data (1907-2017).

Based on the VPA by Toresen and Østvedt (2000), it is puzzling that this VPA prediction of catch deviates from the reported catch since VPA is designed to match the catches exactly. However, it is noted that the VPA from Toresen and Østvedt (2000) is based on Popes approximation in addition to definition of reference ages for which total mortality is assumed known from external sources (see Toresen and Østvedt 2000 for further details). In summary these are candidates for explaining the main differences. In the new VPA these problems are much less apparent as this match XSAM very well in the period 1920-1940 in contrast to the VPA by Toresen and Østvedt (2000), but the discrepancies in 1940-1950 remains. In any case, the fact that VPA (and SEASTAR) predicts higher catches than the reported catches will estimate higher abundances as increased catches will add to the population sizes. Whether this is sufficient for the discrepancies (VPA/SEASTAR vs TASACS/XSAM) is not examined in detail. It is also worth noting that XSAM uses a plus group of 12 whereas some of the other methods do not use a plus group (VPA) and uses higher maximum age 15 (VPA and TASACS) or 16 (SEASTAR). It is well known that estimates in cohort models are sensitive to the choice of plus group. In summary it is concluded that XSAM estimates represent a reasonable candidate for historical perception of the stock.

Using the XSAM time series as basis for determining the recruitment process it is found that the results are comparable to previous studies for this stock and suggest a Blim close to 2500 thousand tons with a time structure similar to what has earlier been reported.

The key for establishing realistic simulations is the model for recruitment including both functional form as well as the structure in the noise. A common finding, confirmed in this study, appear to be that standard statistical methods cannot be used alone to conclude on the most likely functional form for stock recruitment model (Simmonds et al. 2011). Simmonds et al. 2011 propose to reflect the uncertainty in both model and parameters by model averaging based on AIC values of model fits. This is of course of particular importance when estimates of reference points are sensitive to the choice of functional form, and hence the methods in Simmonds et al. (2011) offer a method how to objectively overcome that problem by using model averaging. Here we take a similar approach and find that the estimates of reference points are not particularly sensitive to the choice of functional form, but we still cannot conclude on the most appropriate, and may thus argue to use the model average approach to be used for estimation of reference points.

In the data there is evidence that stock weight at age decreases with density, hence ignoring this effect will artificially create too high SSB at higher stock sizes and too low SSB at low stock sizes.

Estimates of reference points are inevitable imprecise, particularly due to the poorly known recruitment dynamics. For some of the reference points that have been defined earlier $B_{lim}$, $F_{pa}$ and

$F_{MSY}$, we find little evidence for changing the reference points from the already existing reference point for this stock based on this updated analysis. For $B_{pa}$ there is some evidence that it could be decreased from the current $B_{pa}$=5000 to a value somewhat above 3000.

## References

To be completed

# Figures

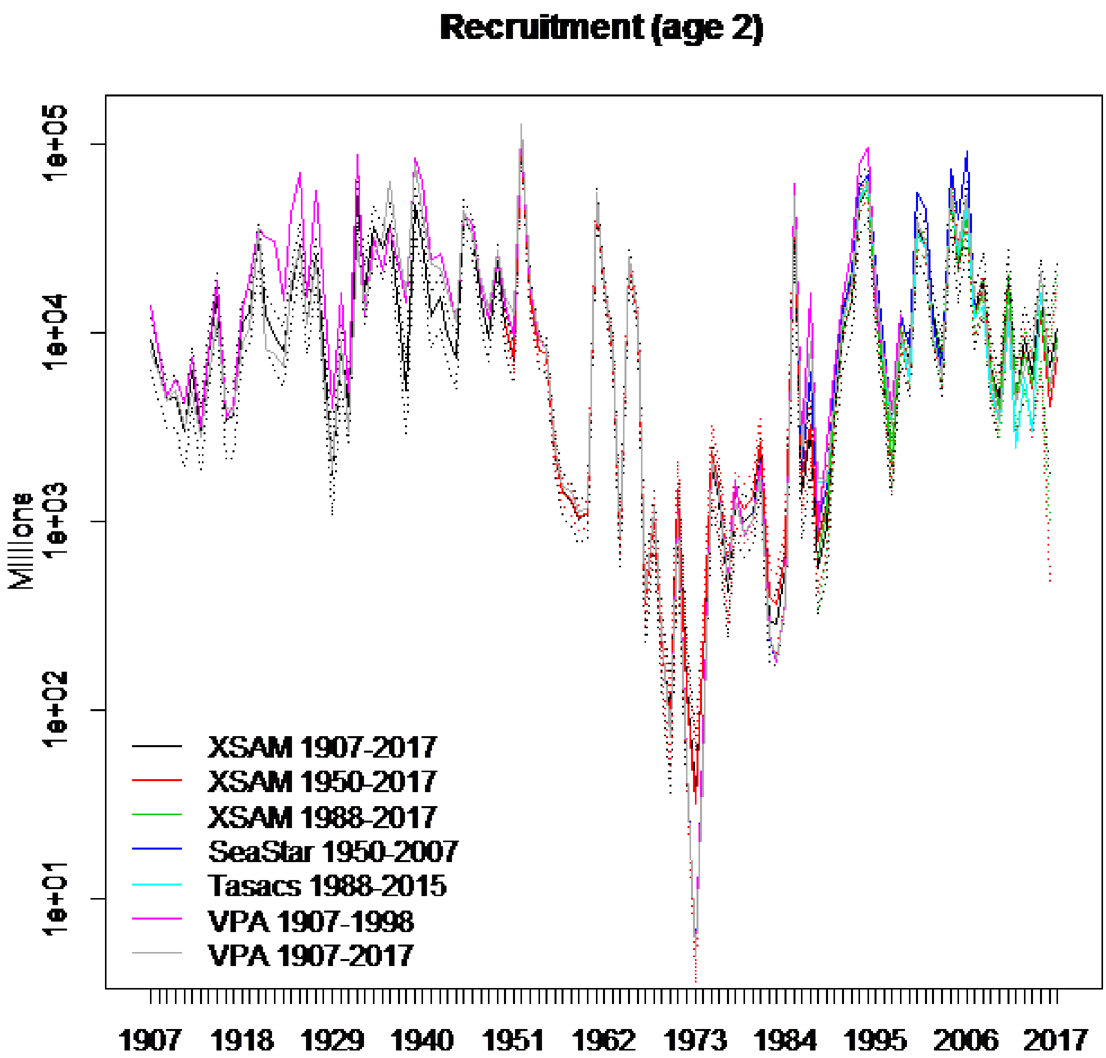


Figure 1. Estimated number of age 2 by different methods and time series of data indicated by the legend. The XSAM model fits includes approximate 95%confidence intervals shown by the broken lines. Note that the scale of the y-axis is logarithmic.

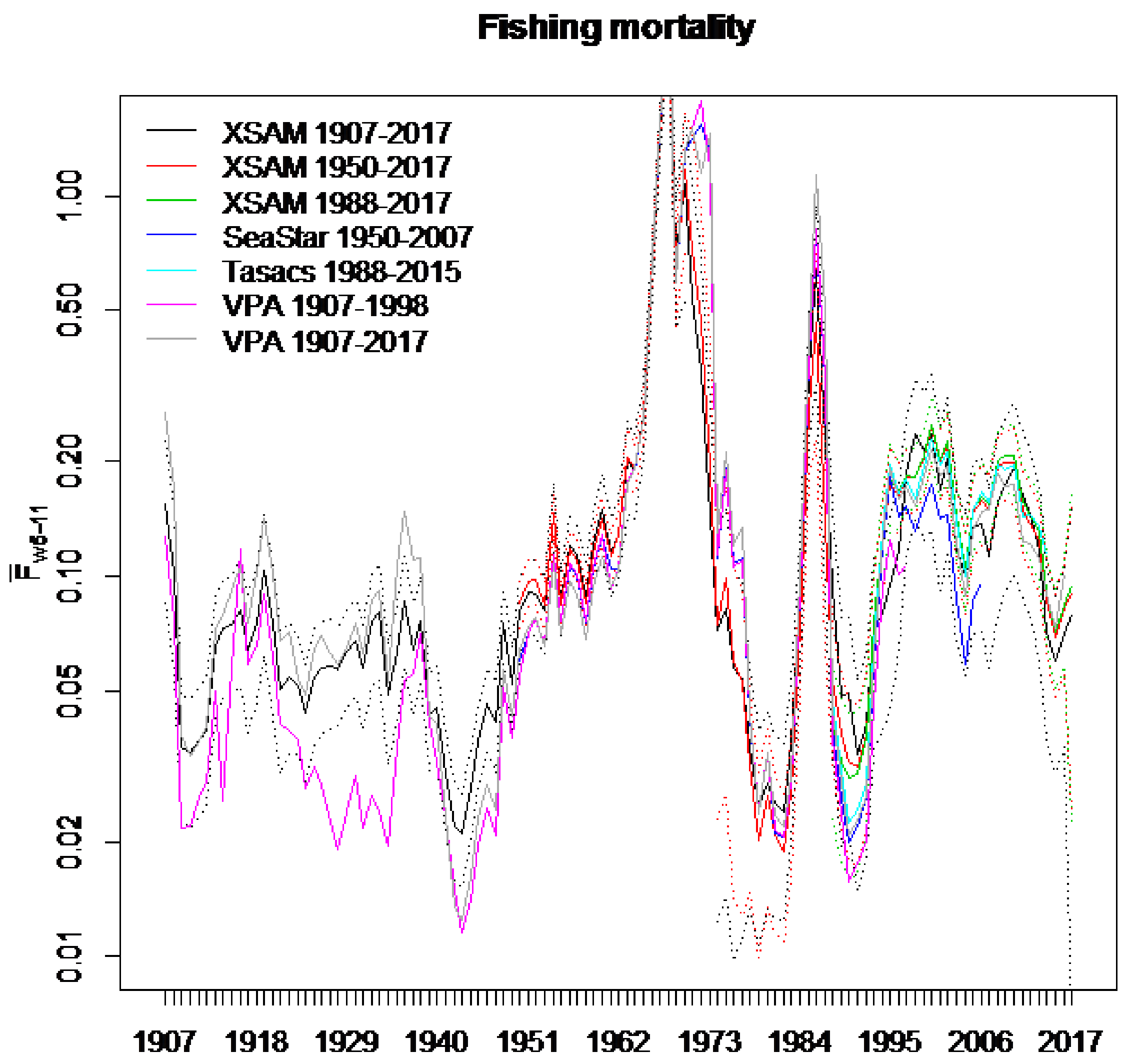


Figure 2. Estimated weighted average fishing mortality ages 5-11 by different methods and time series of data indicated by the legend. The weights are the stock numbers. The XSAM model fits includes approximate 95%confidence intervals shown by the broken lines. Note that the scale of the y-axis is logarithmic.

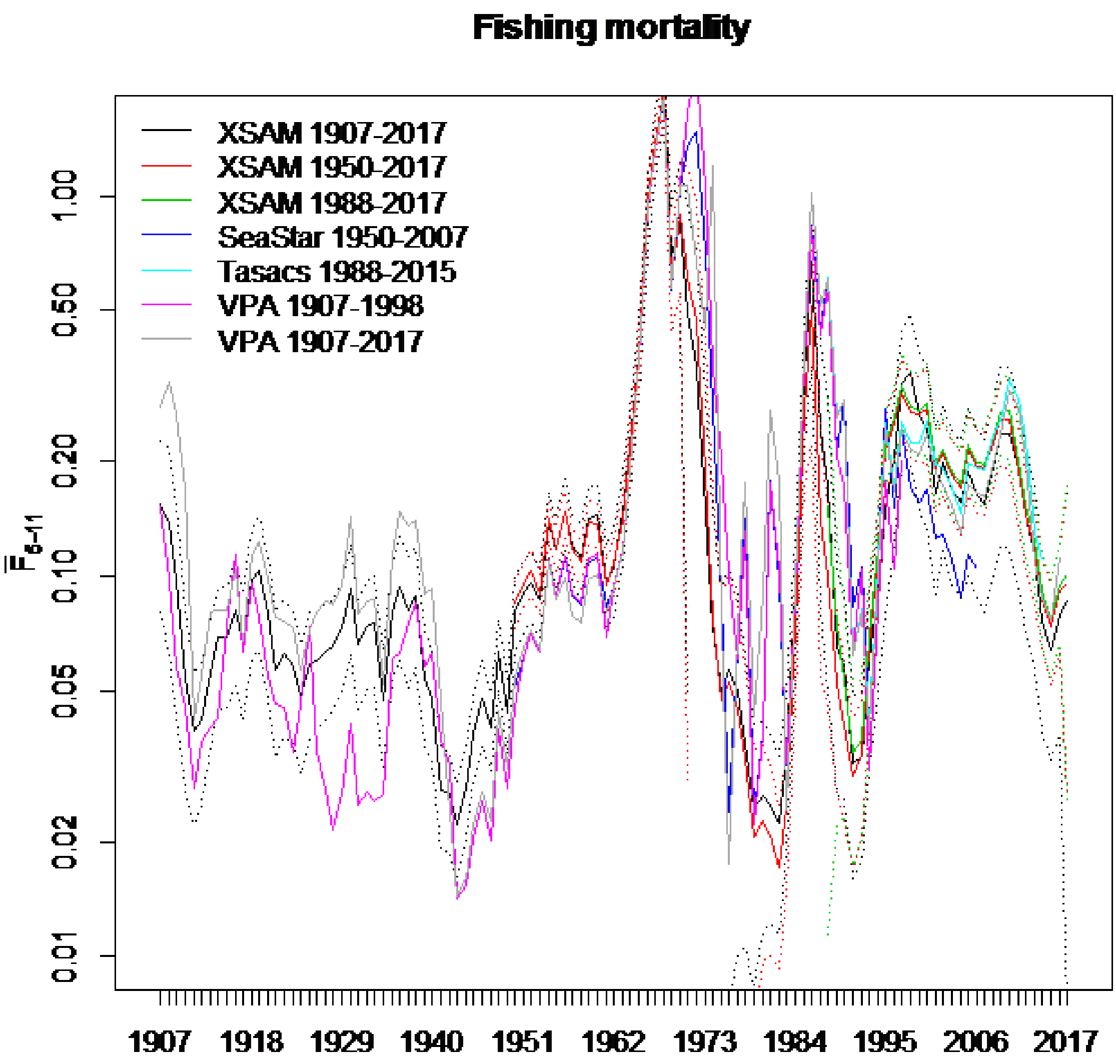


Figure 3. Estimated average fishing mortality ages 5-11 by different methods and time series of data indicated by the legend. The XSAM model fits includes approximate 95%confidence intervals shown by the broken lines. Note that the scale of the y-axis is logarithmic.

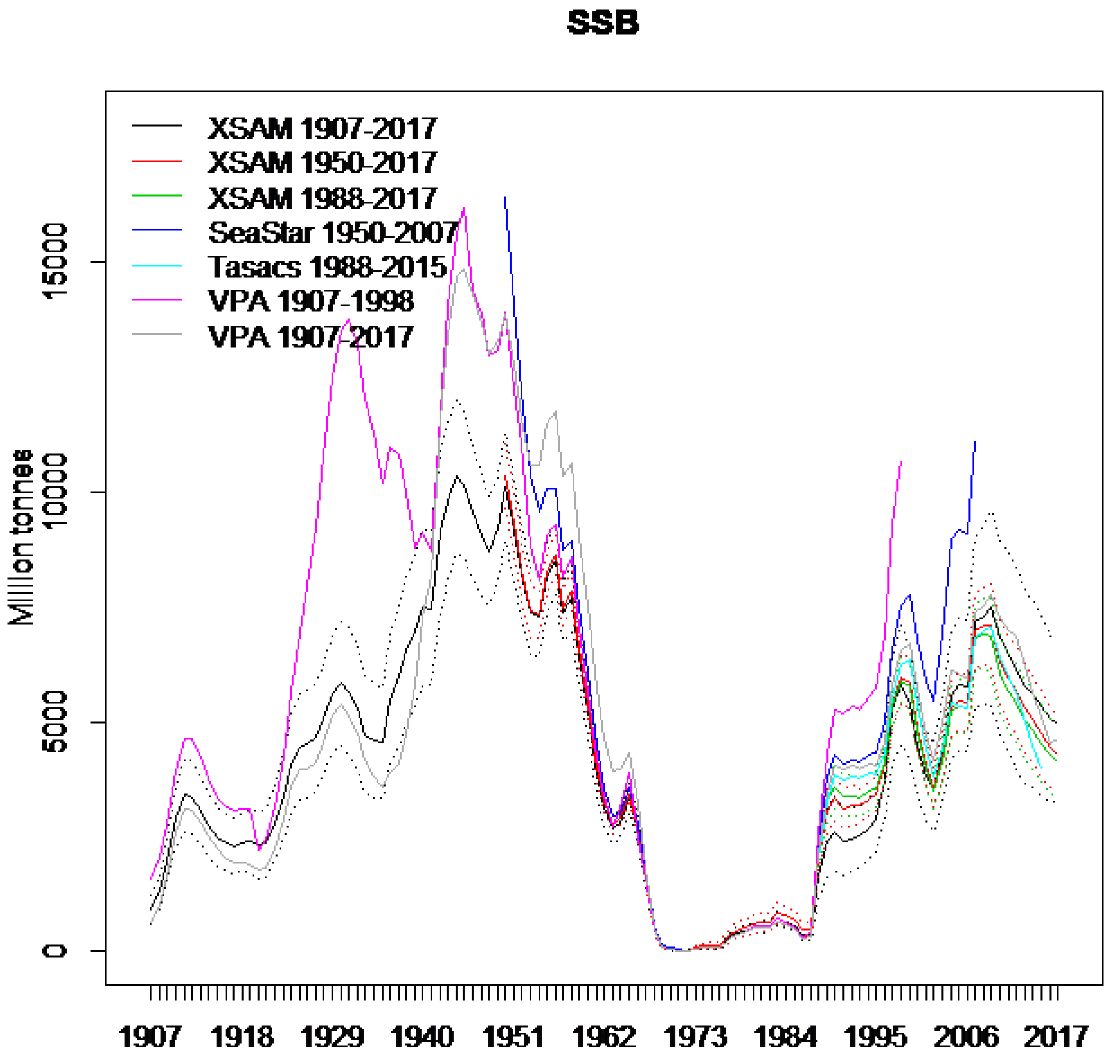


Figure 4. Estimated spawning stock biomass by different methods and time series of data indicated by the legend. The XSAM model fits includes approximate 95%confidence intervals shown by the broken lines.

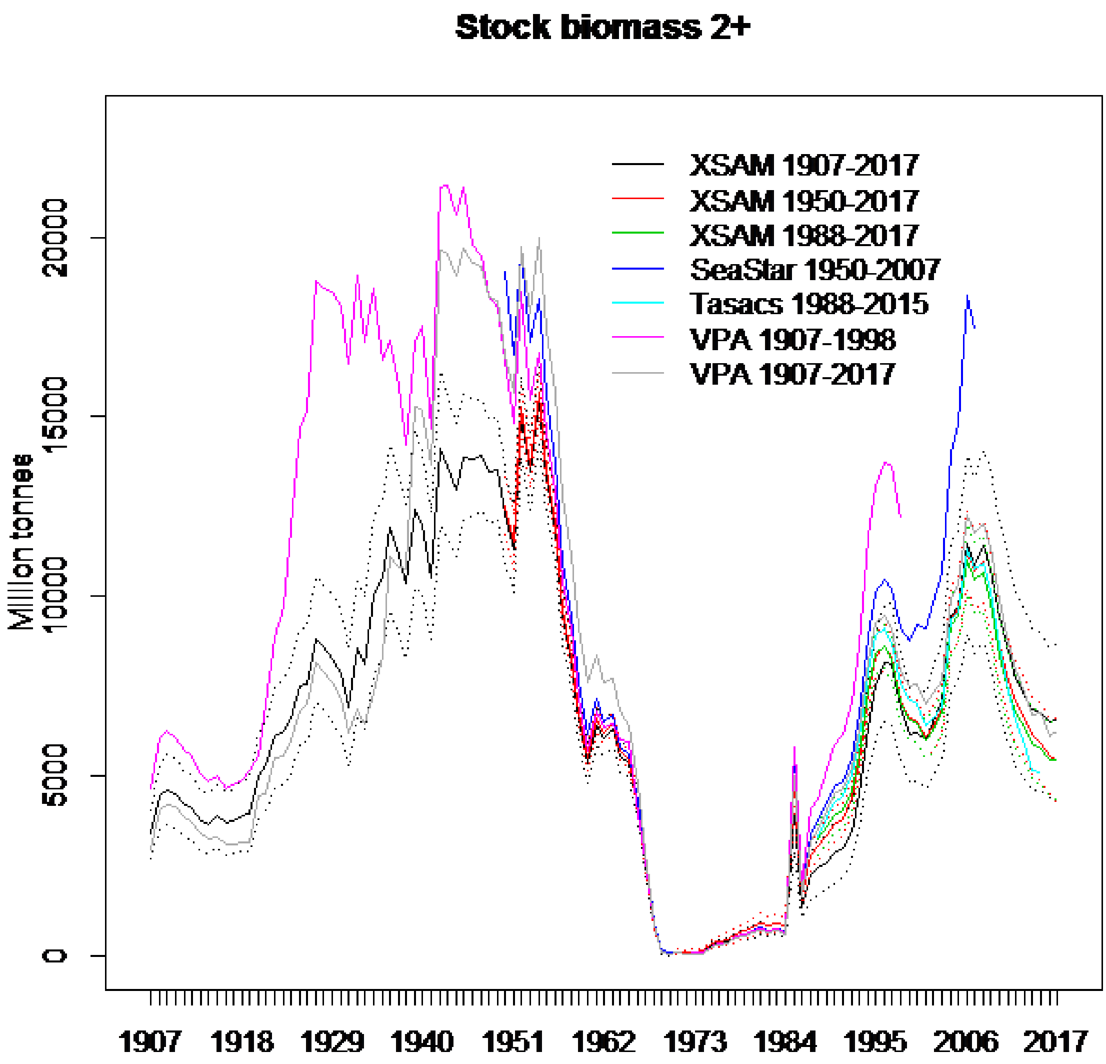


Figure 5. Estimated stock biomass ages 2+ by different methods and time series of data indicated by the legend. The XSAM model fits includes approximate 95%confidence intervals shown by the broken lines.

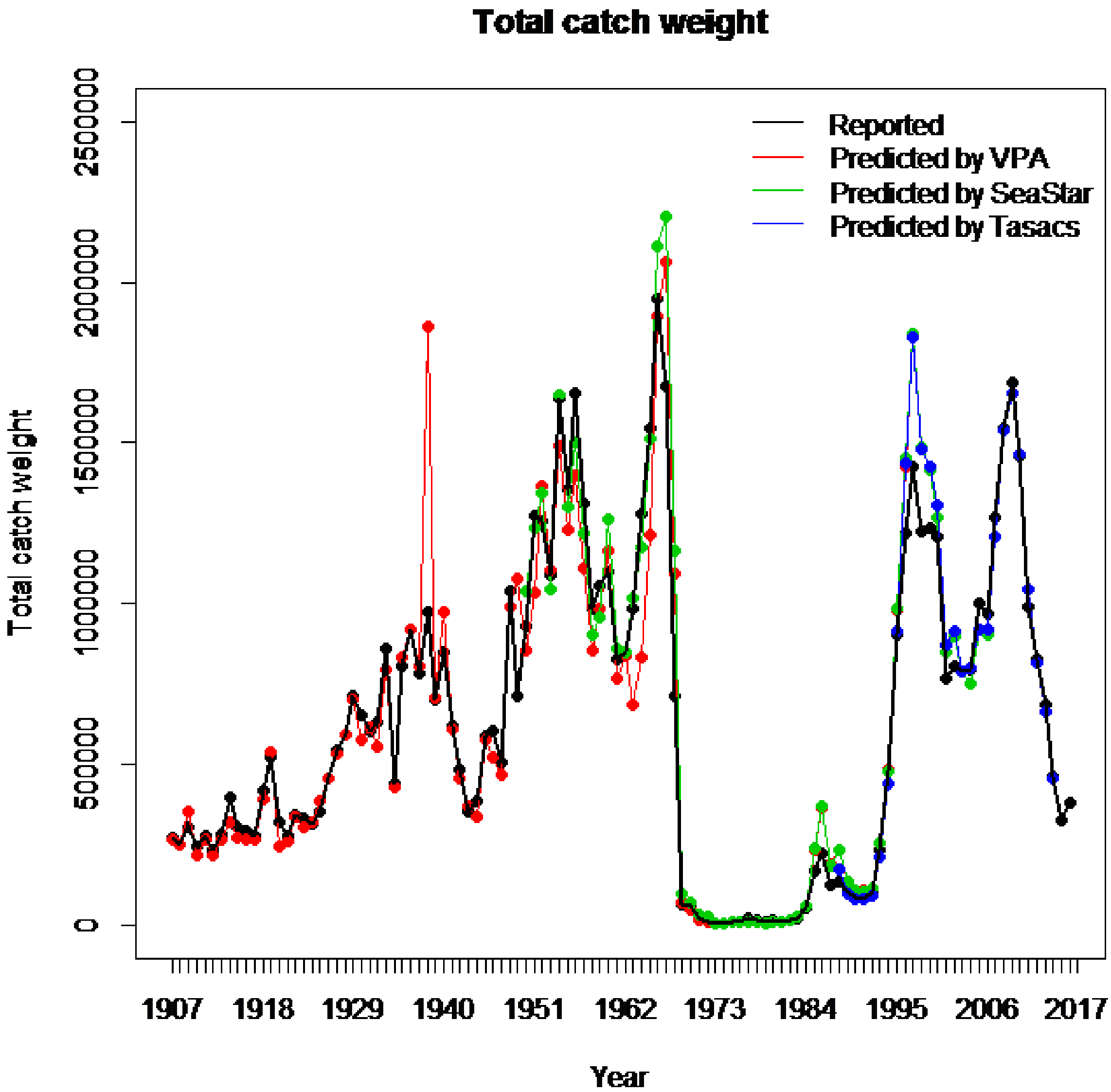


Figure 6. Total catch weights as reported and predicted by the different estimates indicated in the legend. Note that the figure does not include predicted catches by XSAM as it starts on age 2 while the reported catches includes catches for ages 0 and 1.

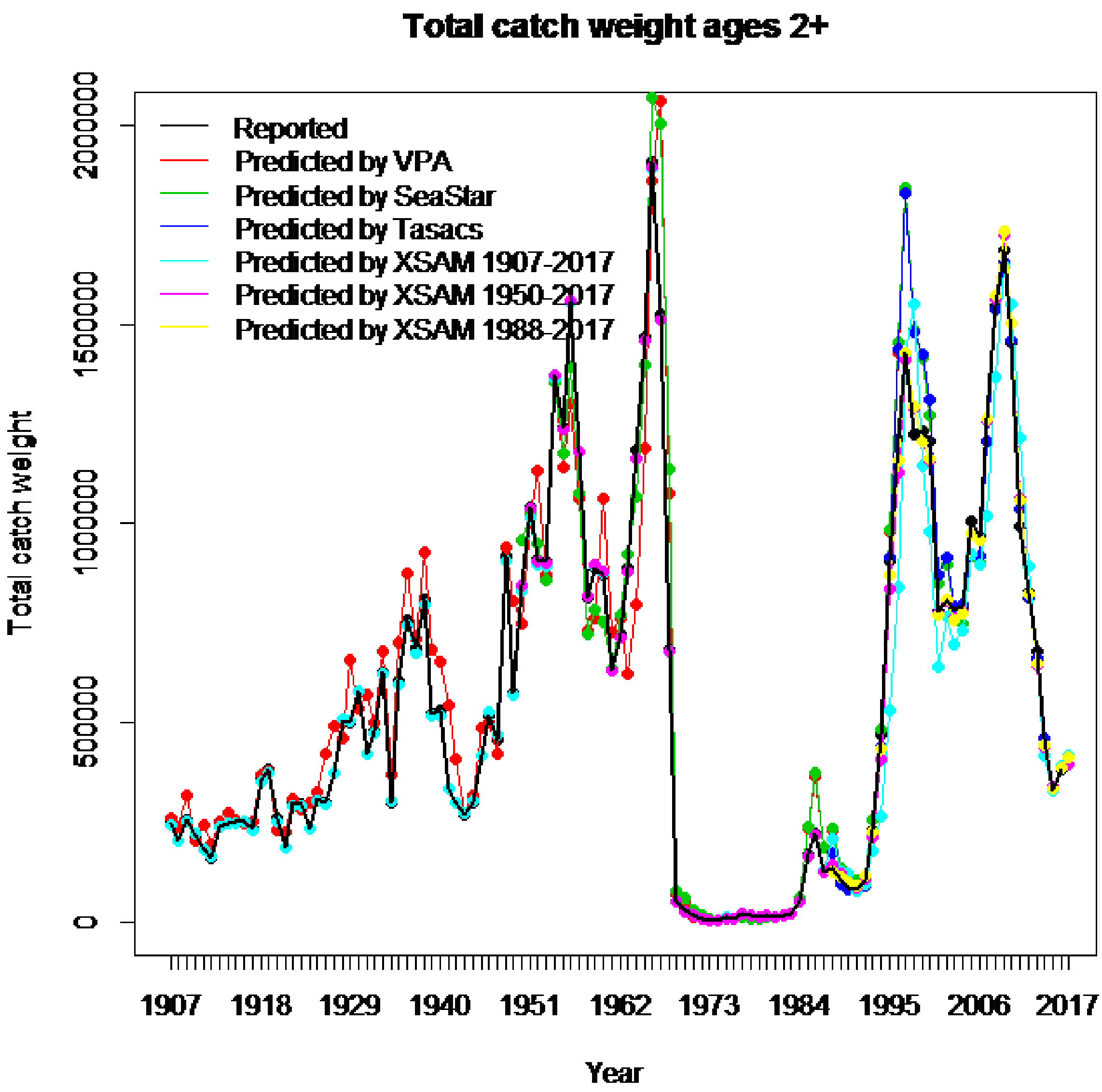


Figure 7. Total catch weights for ages 2+as reported and predicted by the different estimates indicated in the legend.

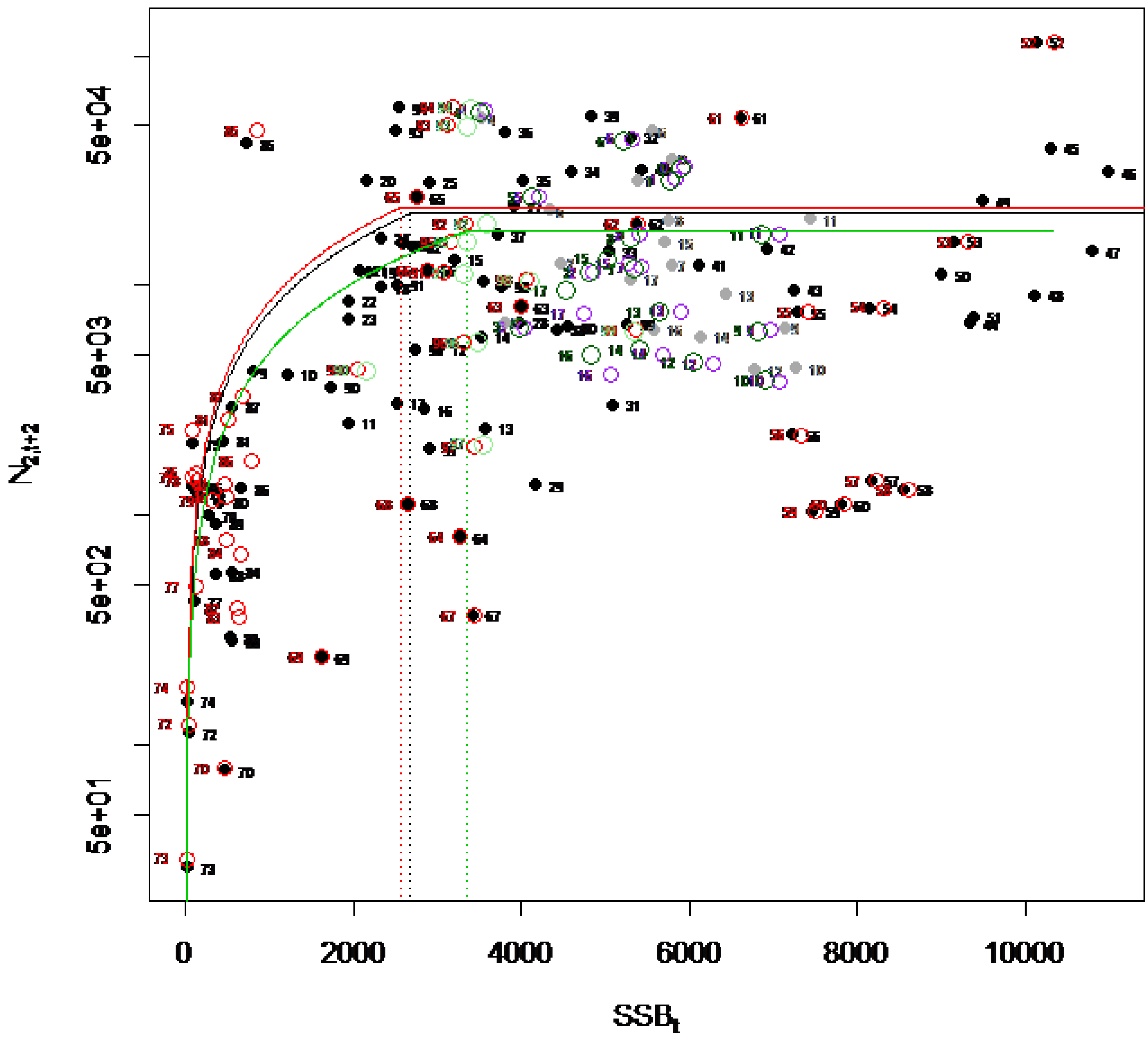


Figure 8. Recruitment (numbers at age 2) versus SSB (two years before) based on data XSAM estimates for 1907-2017 (1907-1999 black and 2000-2017 gray), 1950-2017 (1950-1999 red and 2000-2017 purple) and 1988-2017 (1988-1999 light green and 2000-2017 dark green). The observation year is indicated alongside the points. The lines show the respective fits of the segmented regression model. Note that the scale of the y-axis is logarithmic (and thus the fit appear nonlinear up to the breakpoint)

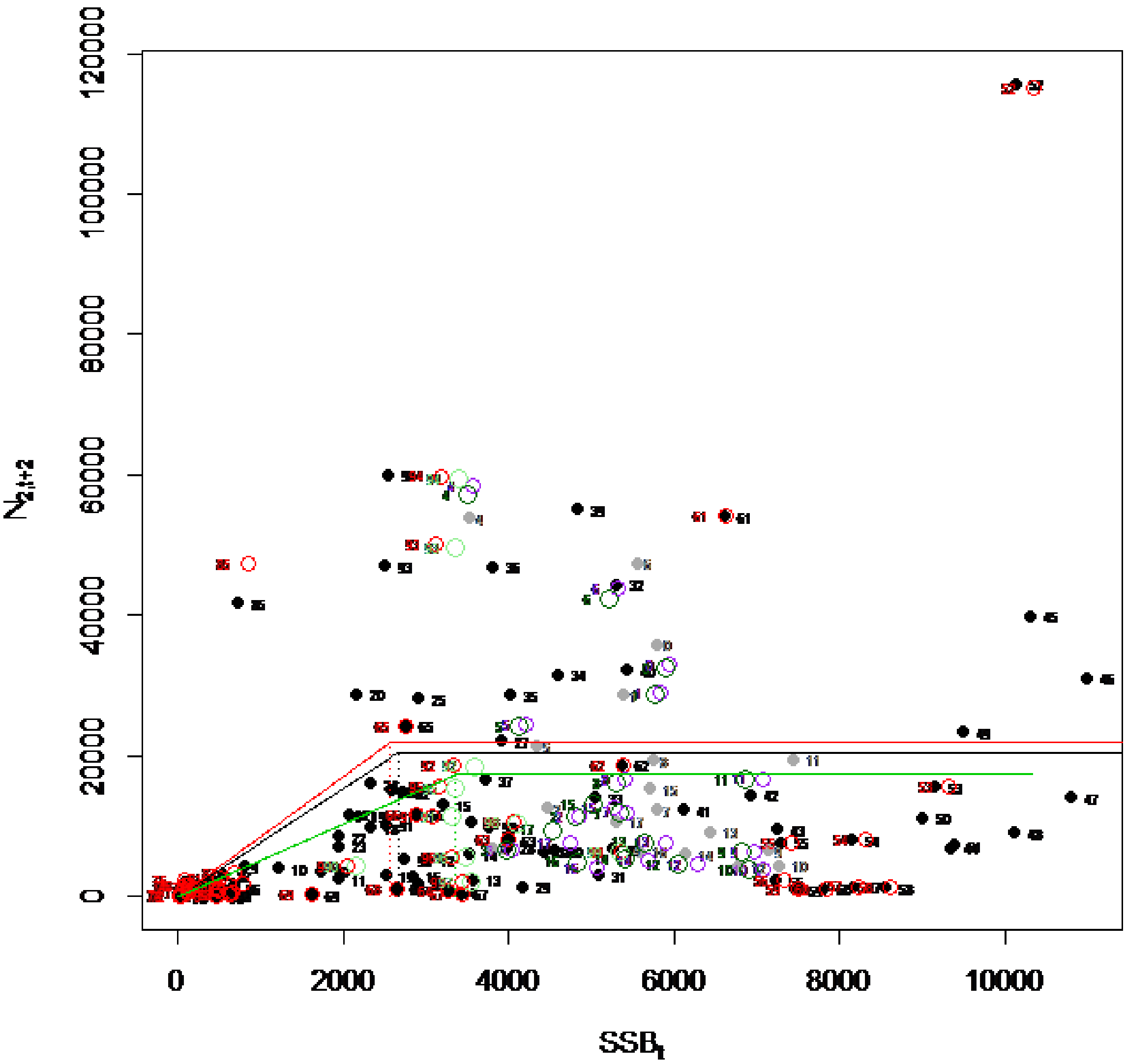


Figure 9. Recruitment (numbers at age 2) versus SSB (two years before) based on data XSAM estimates for 1907-2017 (1907-1999 black and 2000-2017 grey), 1950-2017 (1950-1999 red and 2000-2017 purple) and 1988-2017 (1988-1999 light green and 2000-2017 dark green). The observation year is indicated alongside the points. The lines show the respective fits of the segmented regression model.

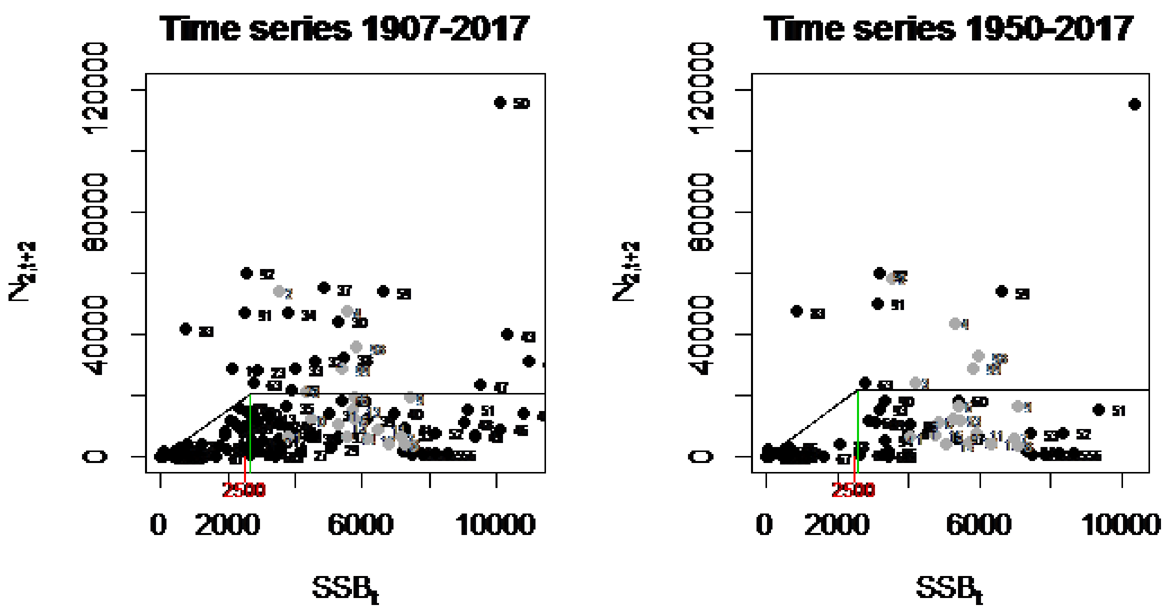


Figure 10. Recruitment (numbers at age 2) versus SSB (two years before) based on data XSAM estimates for 1907-2017 (left panel; 1907-1999 black and 2000-2017 grey), 1950-2017 (right panel; 1950-1999 black and 2000-2017 gray). The cohort is indicated alongside the points. The black lines show the respective fits of the segmented regression model, the green line the corresponding breakpoint, and the value 2500 indicated in red.

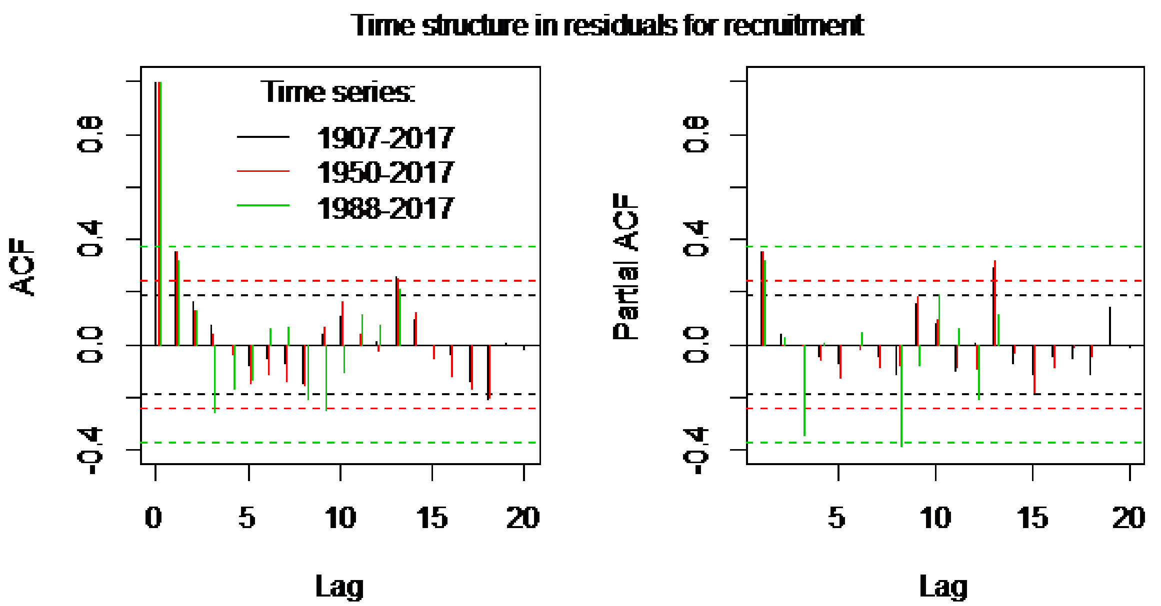


Figure 11. Auto correlation (left) and partial autocorrelation (right) for the residuals after fitting the segmented regression models for spawning stock recruitment (at age 2) for the time series 1907-2017 (black), 1950-2017 (red) and 1988-2017 (green). Corresponding confidence limits are shown with the broken lines.

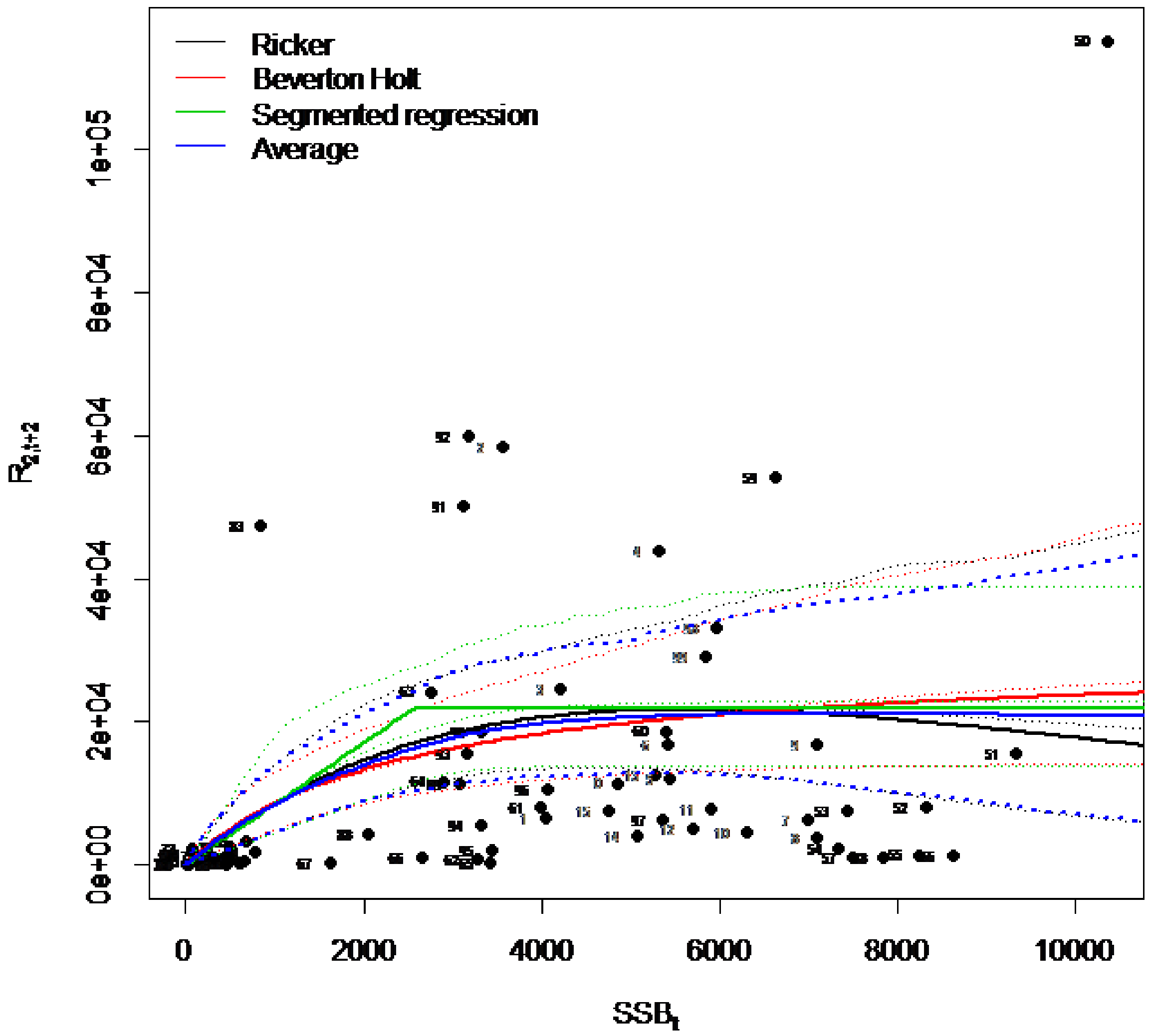


Figure 12. Recruitment (numbers at age 2) versus SSB (two years before) based on data XSAM estimates for 1950-2017. The cohort is indicated alongside the points. The lines are the mean in the fitted recruitment models Ricker (black), Beverton Holt (red), Hockey Stick (green) and a the model average (blue). The model average is based on the AIC-smoothed estimate (see text). The broken lines are 95 confidence intervals of the mean and found by 1000 replicates of pairs of stock recruitment data.

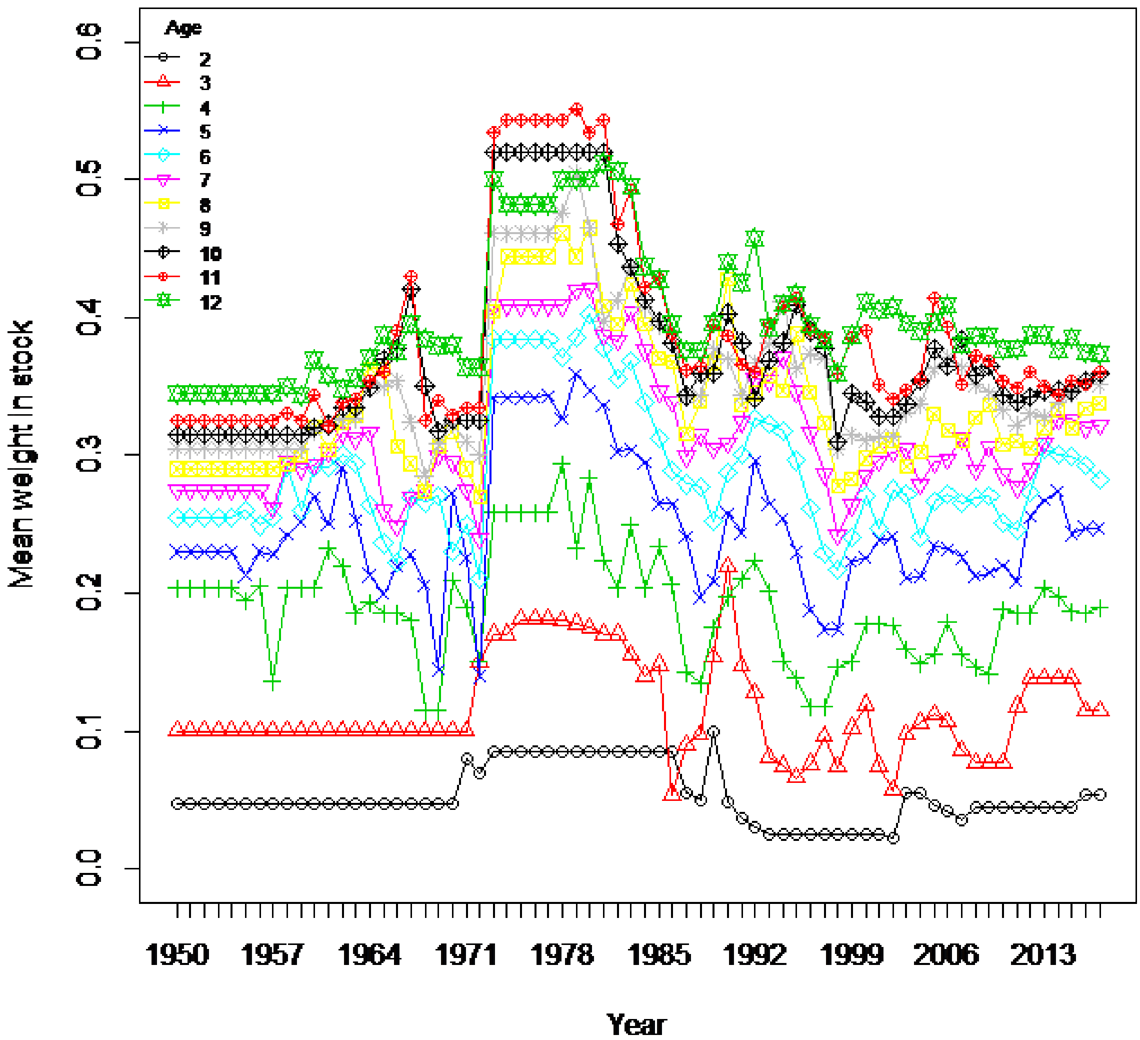


Figure 13. Estimated values of mean weight at age and year in stock for NSS herring as reported by WGWIDE.

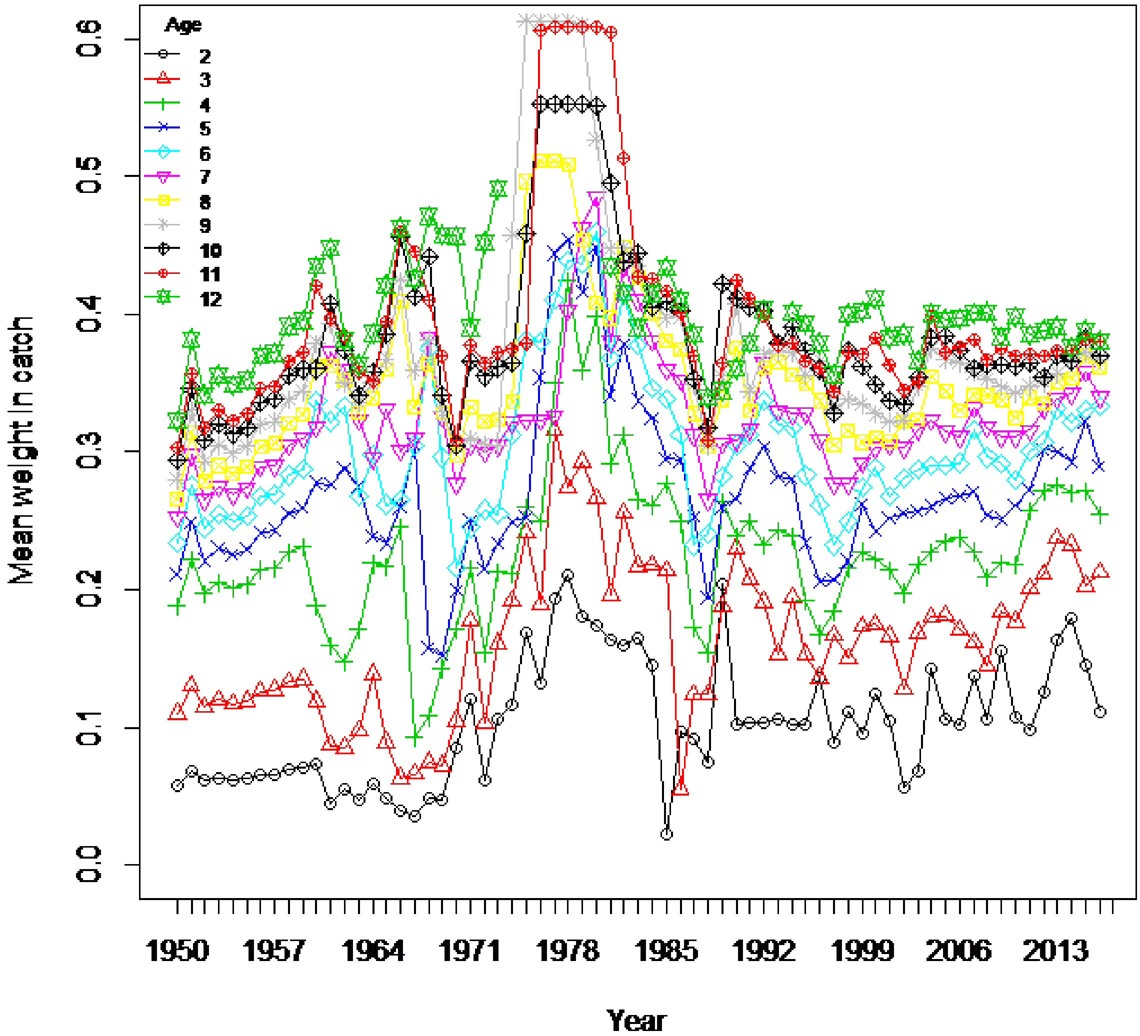


Figure 14. Estimated values of mean weight at age and year in catch for NSS herring as reported by WGWIDE.

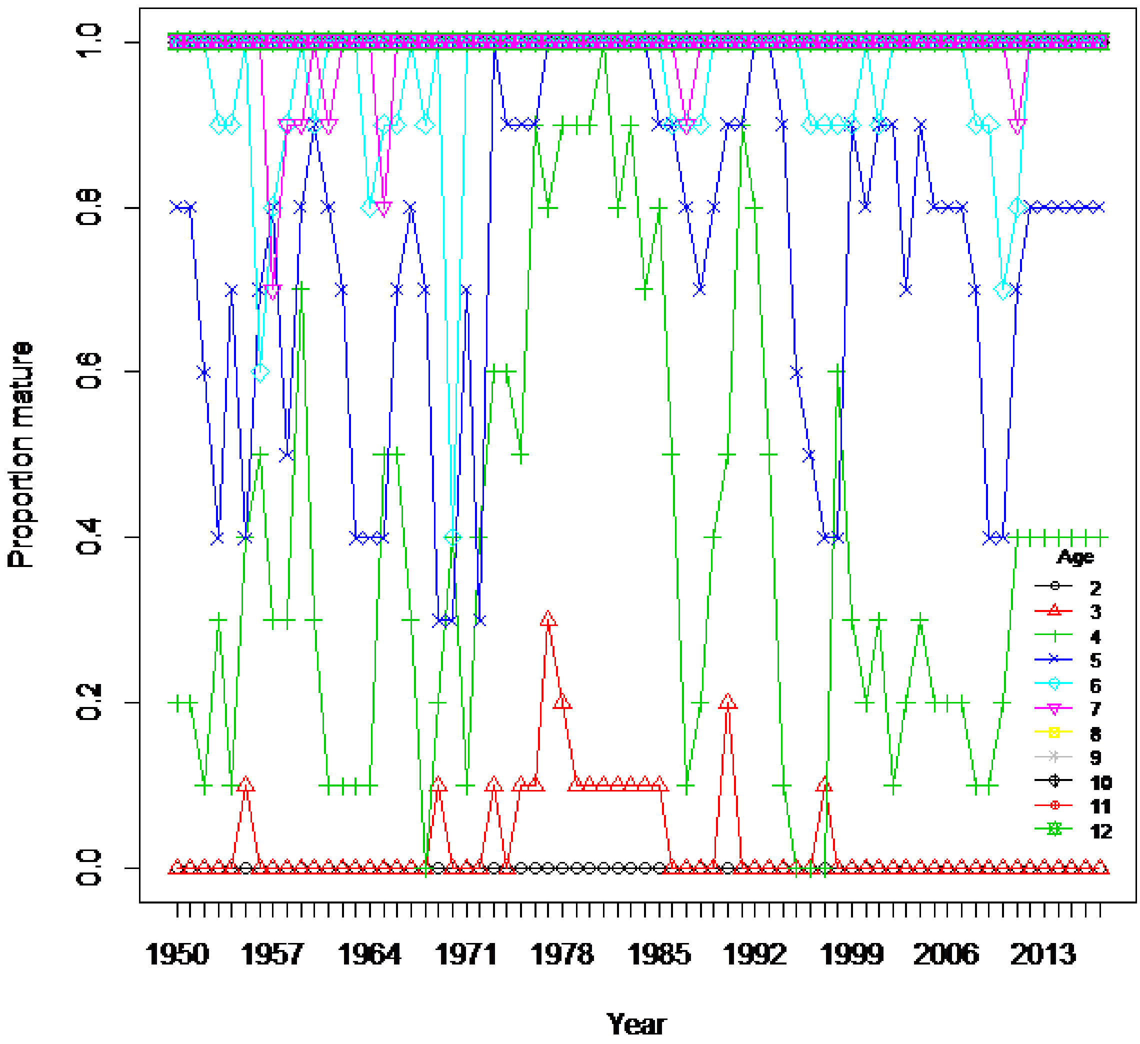


Figure 15. Estimated values of proportion mature at age and year in stock for NSS herring as reported by WGWIDE.

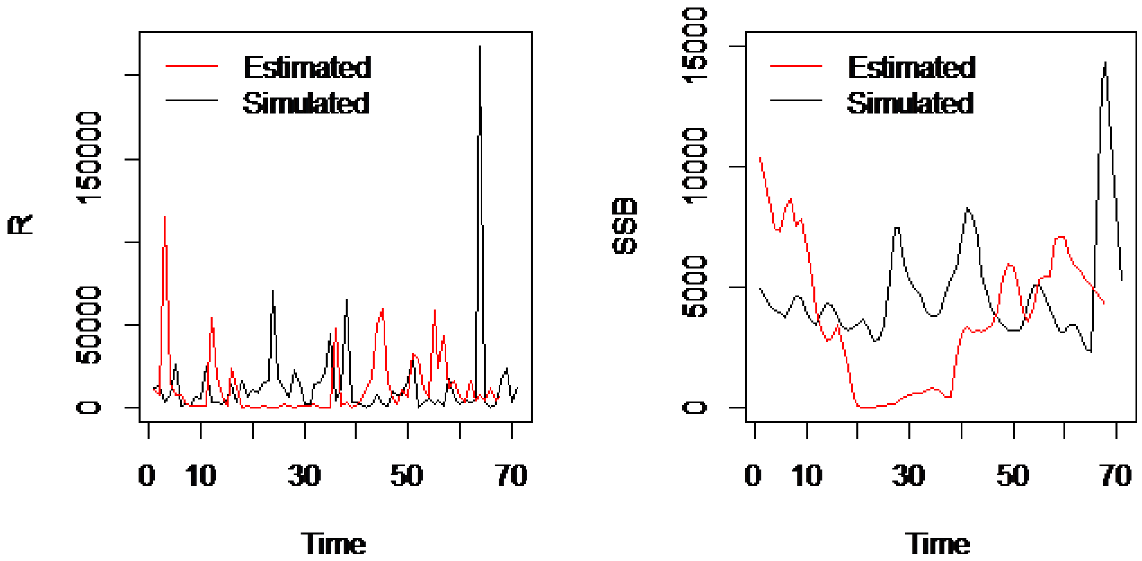


Figure 16. Simulated recruitment and SSB based on the segmented regression model for recruitment. 70 time steps is extracted and plotted with the estimates for comparison.

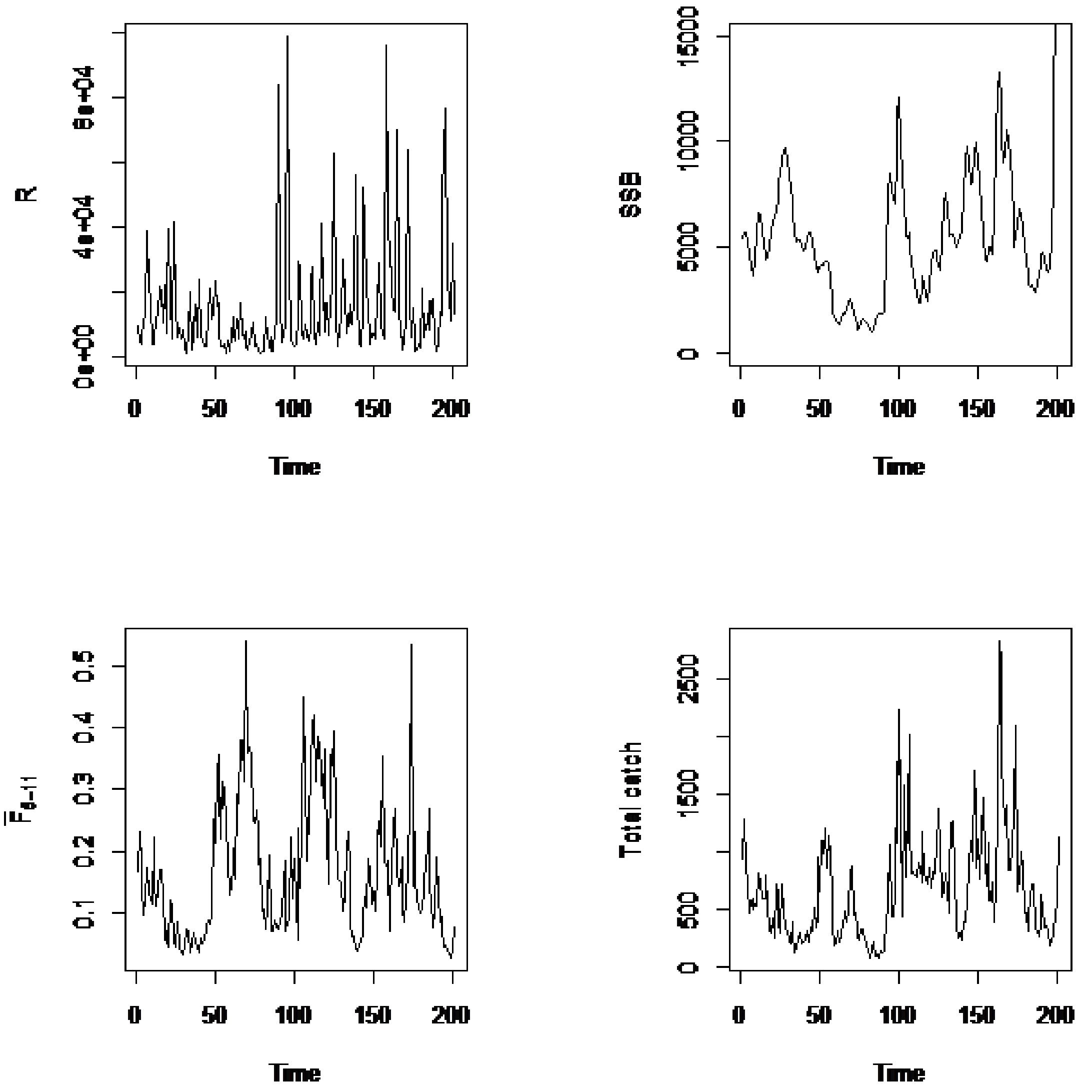


Figure 17. Simulated recruitment, SSB, average fishing mortality and total catch based on the segmented regression model for recruitment where maximum recruitment is restricted to 20% above maximum observed. Data for 200 time steps is extracted for visualization.

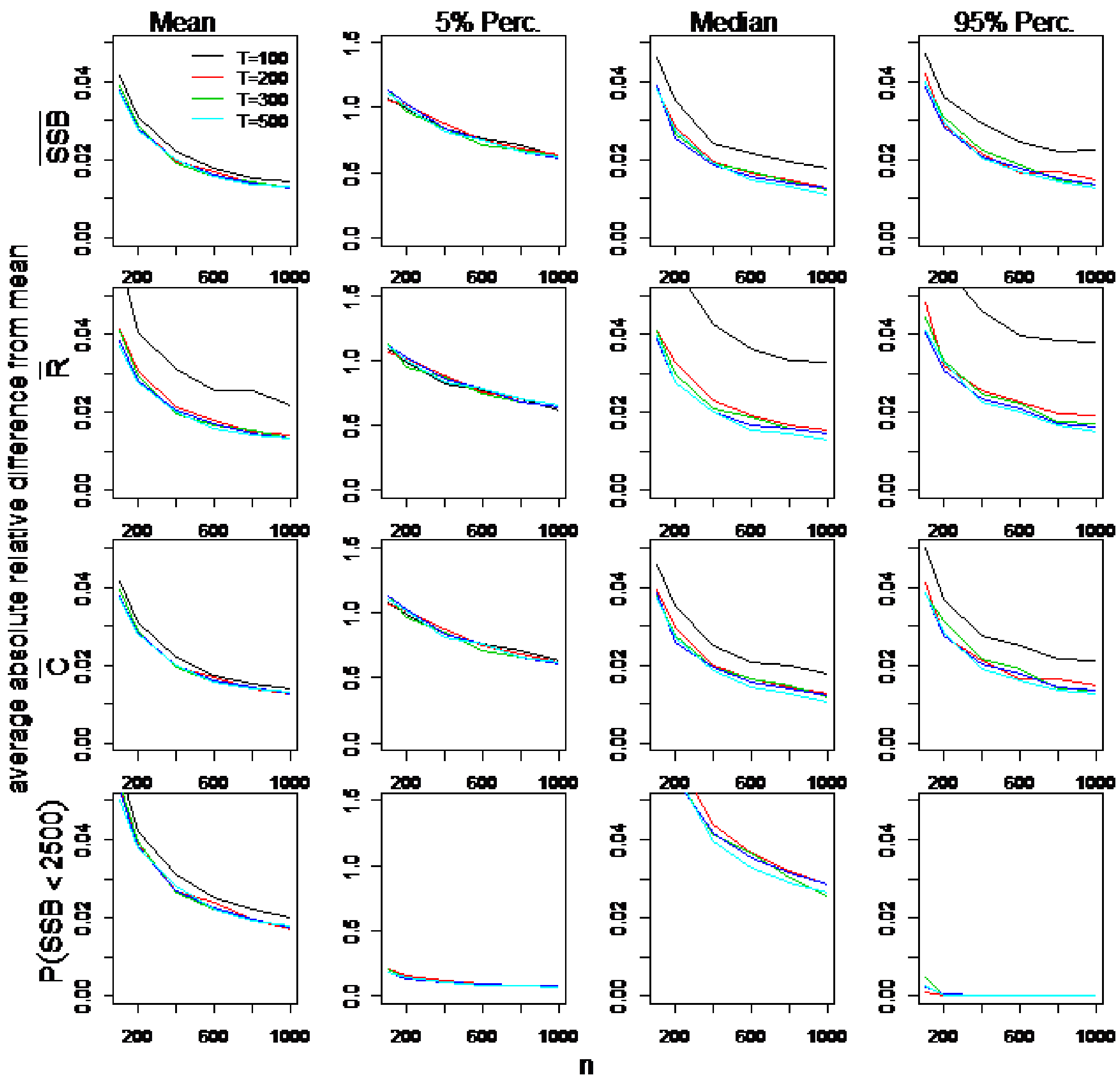


Figure 18. Examining performance statistics as function of number replicates of model parameters $n$ and number of years simulated per replicate $T$ on estimates of mean, 5 % percentile, median and 95% percentile of mean SSB, Recruitment, annual catch and probability of SSB falling below 2500. The performance statistics is average absolute relative difference from the mean. The statistics is derived using variable selection pattern, setting other biological parameters equal to the long term weighted mean and fishing with constant F=0.2.

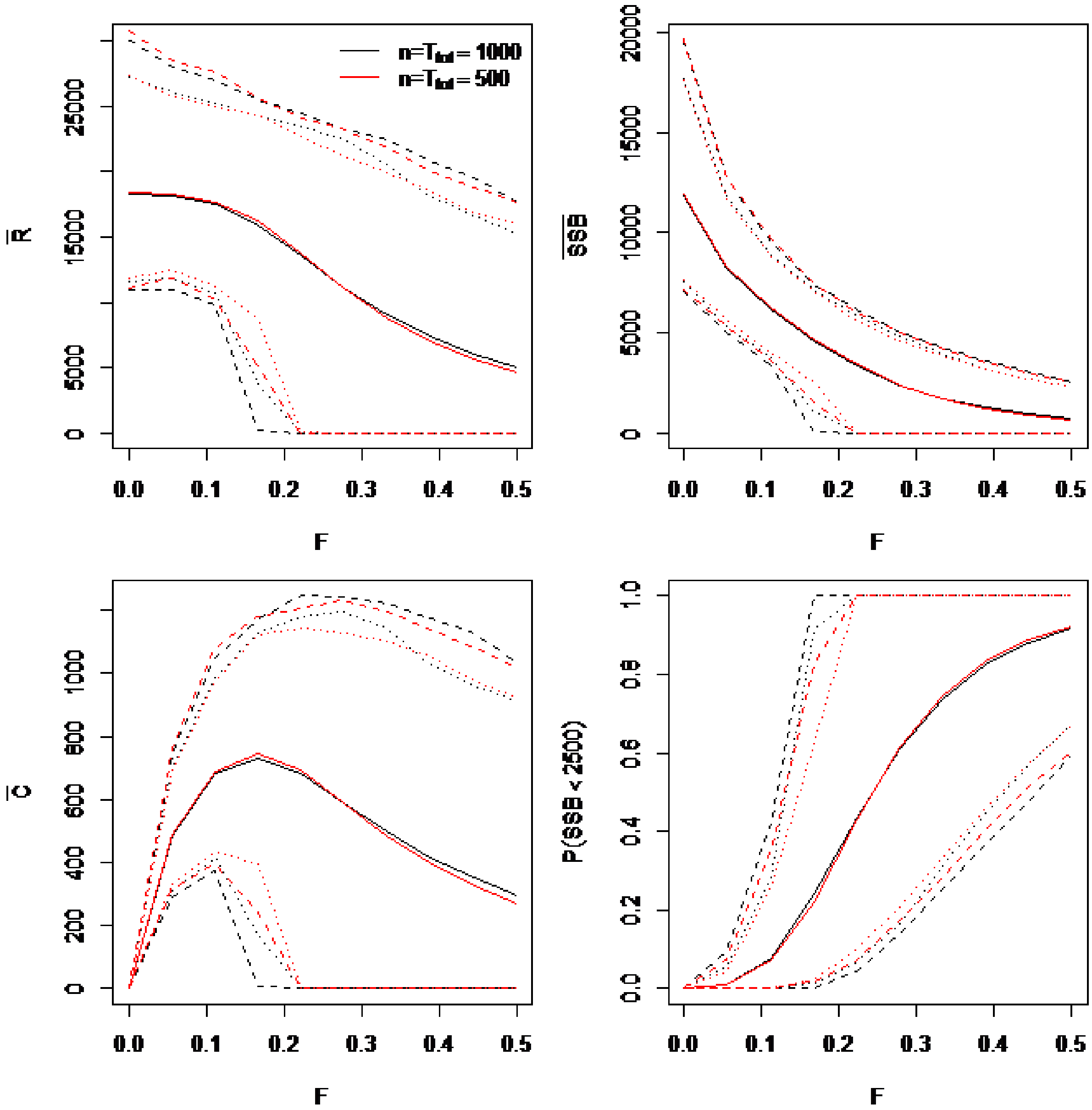


Figure 19. Mean recruitment (upper left), SSB (upper right), annual catch (lower left), and probability of falling below 2500 (lower right) using the AIC smoothed stock recruitment function as a function of average fishing mortality ages 5-11 (x-axis) assuming noise in recruitment follows an AR (1) model. The solid lines are the mean, 90% confidence interval (dotted lines) and 95% confidence intervals (dashed lines) accounting for the variability in stock recruitment parameters (see text). The black lines correspond to the results obtained by 1000 replicates of parameters/models each ran for 1000 time steps, while the results when reducing both to 500 are shown with red lines. In the simulations selectivity in F is according to the XSAM model fit, while weight at age in stock and catch and proportion mature in stock is set constant as the long term weighted mean.

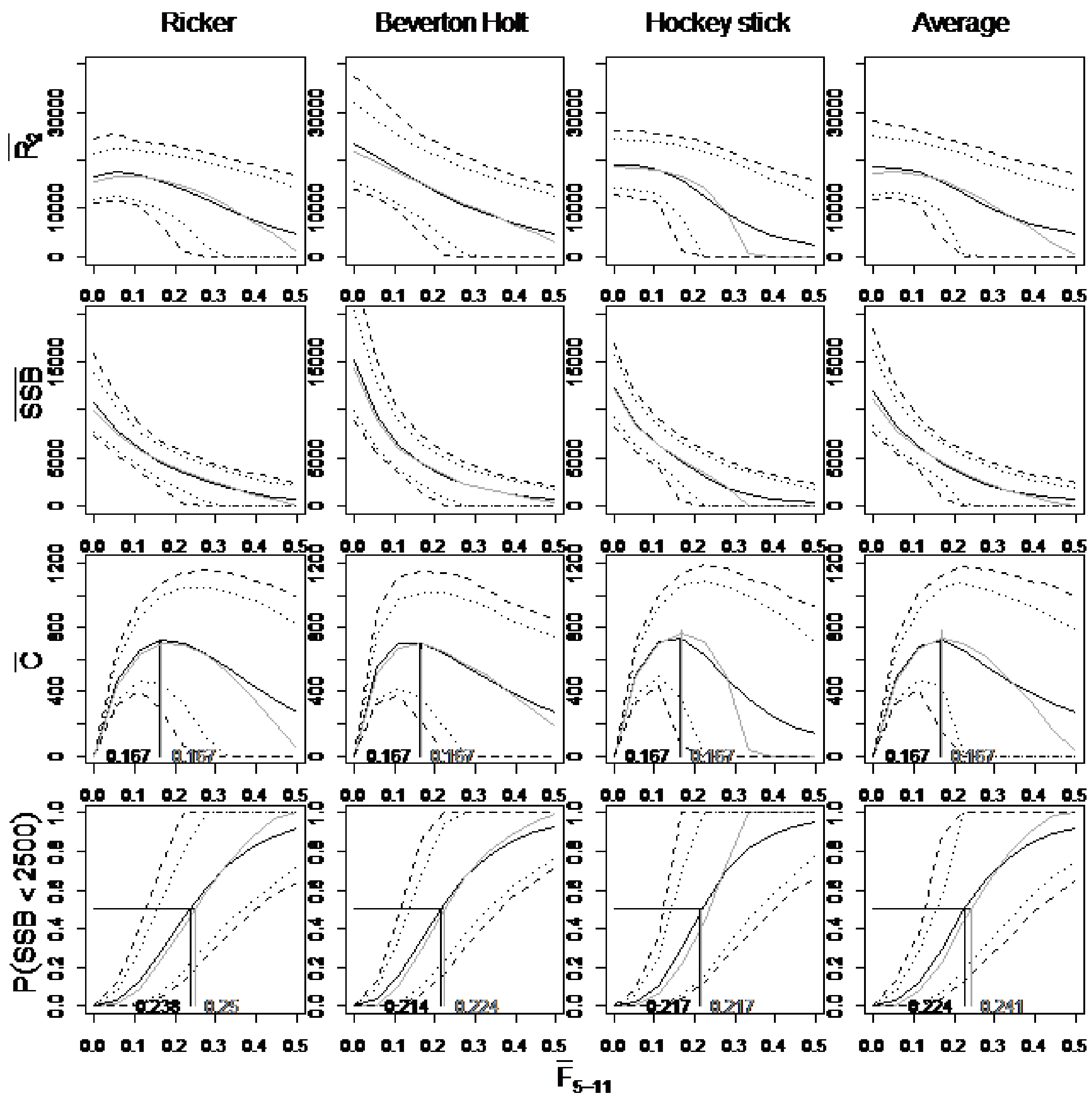


Figure 20. Mean recruitment (top row), SSB (second row), annual catch (third row), and probability of falling below Blim (=2500) by recruitment model Ricker (left), Beverton Holt (second column), Hockey stick (third column) and model average (last column), as a function of average fishing mortality ages 5-11 (x-axis) assuming noise in recruitment follows an AR(1) process. The gray lines correspond to the median values while black lines are the mean (solid line), 80% confidence interval (dotted lines) and 90% confidence intervals (dashed lines) accounting for the variability in stock recruitment parameters (see text). The vertical lines for mean annual catch indicates maximum catch and for probability of falling below 2500 where it intersects 0.5 on the y-axis. In the simulations selectivity in F is according to the XSAM model fit, while weight at age in stock and catch and proportion mature in stock is set constant as the long term weighted mean.

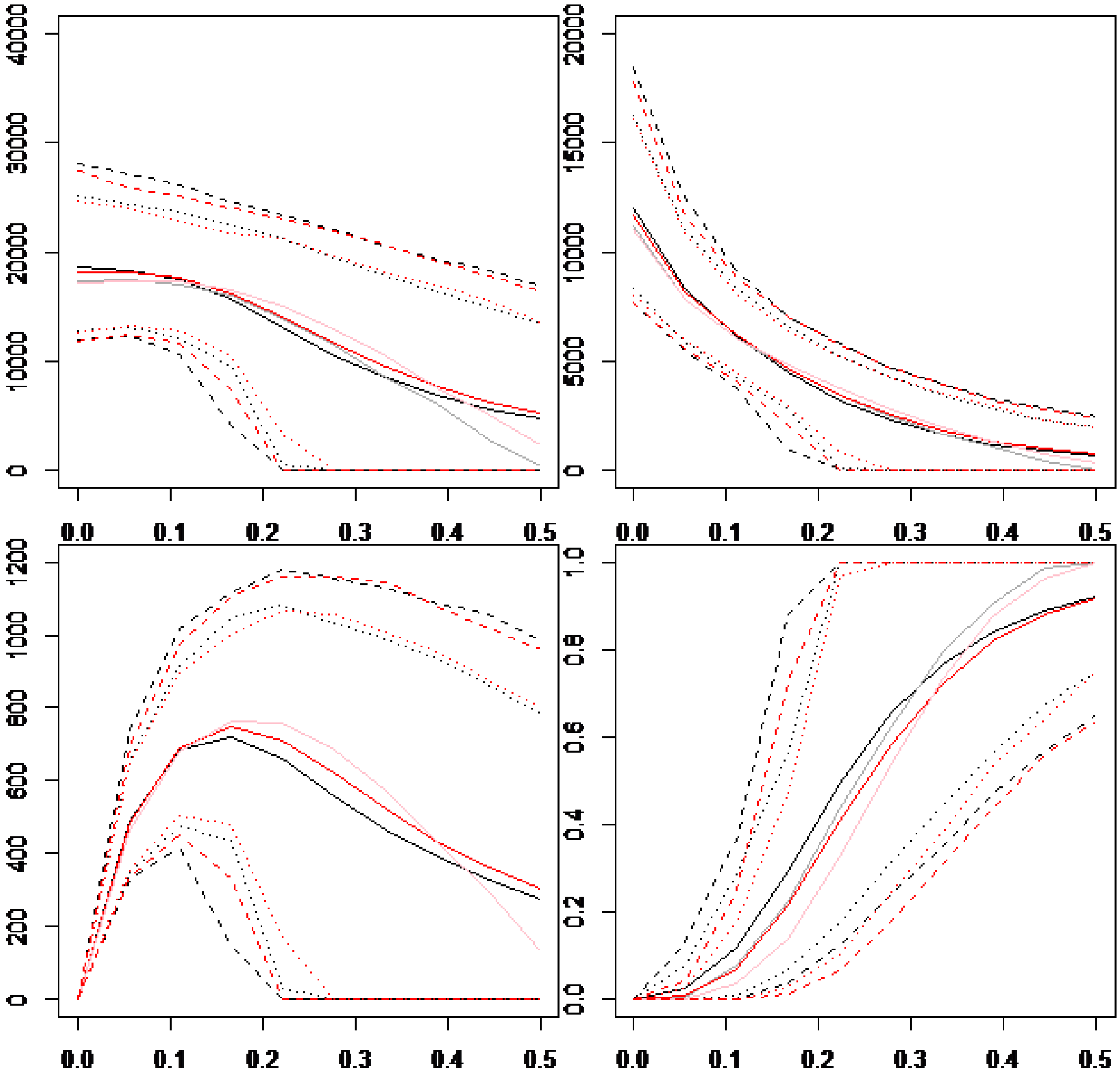


Figure 21. Mean recruitment (upper left), SSB (upper right), annual catch (lower left), and probability of falling below 2500 (lower right) using the AIC smoothed stock recruitment function as a function of average fishing mortality ages 5-11 (x-axis) assuming noise in recruitment follows an AR (1) model. The solid lines are the mean, 90% confidence interval (dotted lines) and 95% confidence intervals (dashed lines) accounting for the variability in stock recruitment parameters (see text). The black lines correspond to the results obtained by using the AR(1) model for the noise in the recruitment process while the red lines corresponds to assuming no serial dependence in recruitment parameters.

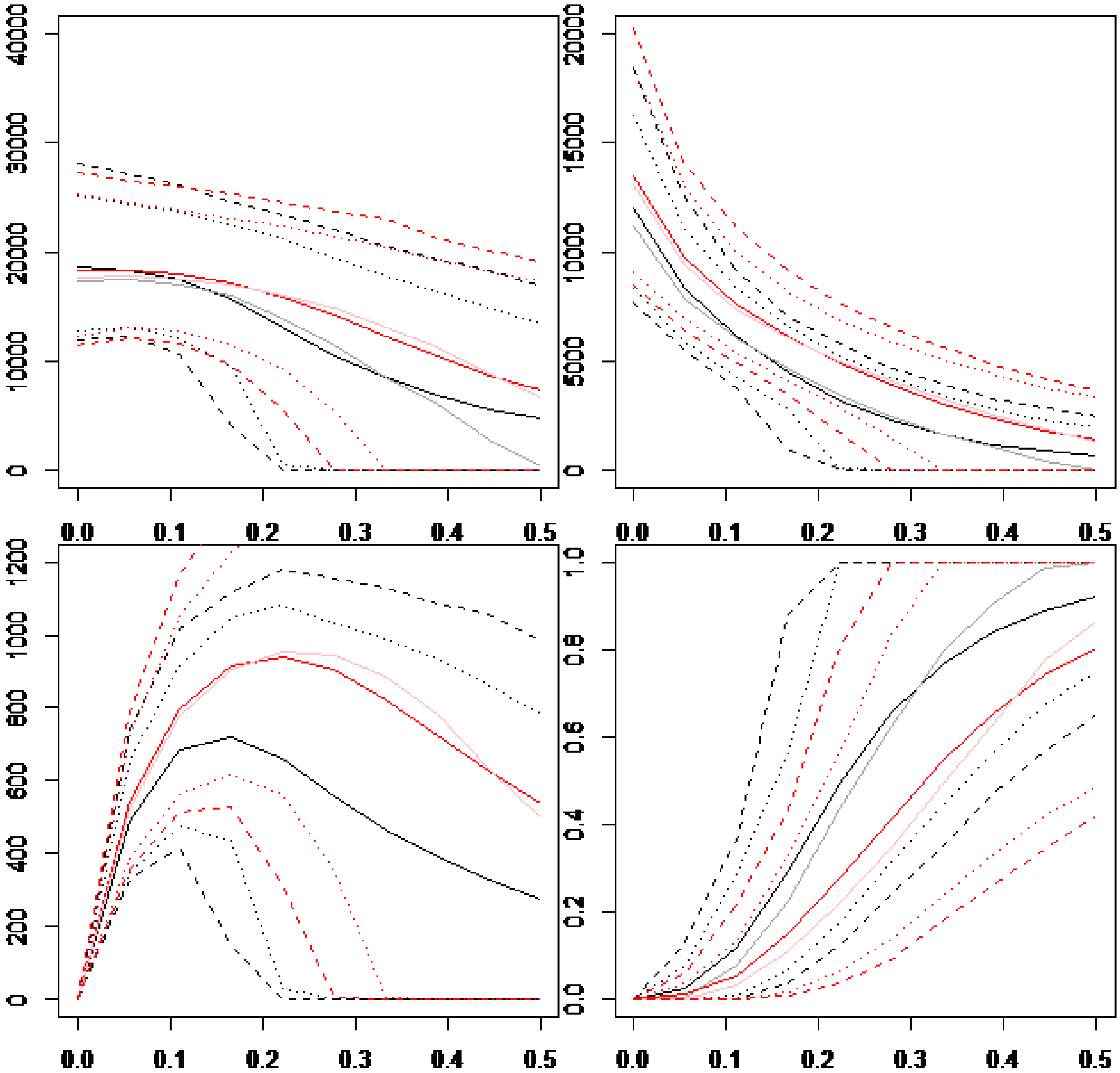


Figure 22. Mean recruitment (upper left), SSB (upper right), annual catch (lower left), and probability of falling below 2500 (lower right) using the AIC smoothed stock recruitment function as a function of average fishing mortality ages 5-11 (x-axis) assuming noise in recruitment follows an AR (1) model. The solid lines are the mean, 90% confidence interval (dotted lines) and 95% confidence intervals (dashed lines) accounting for the variability in stock recruitment parameters (see text). The black lines correspond to the results obtained by using the weighted long term mean of weight at age and proportion mature at age while the red lines corresponds to replacing the weighted long term mean with the average of the last three years.

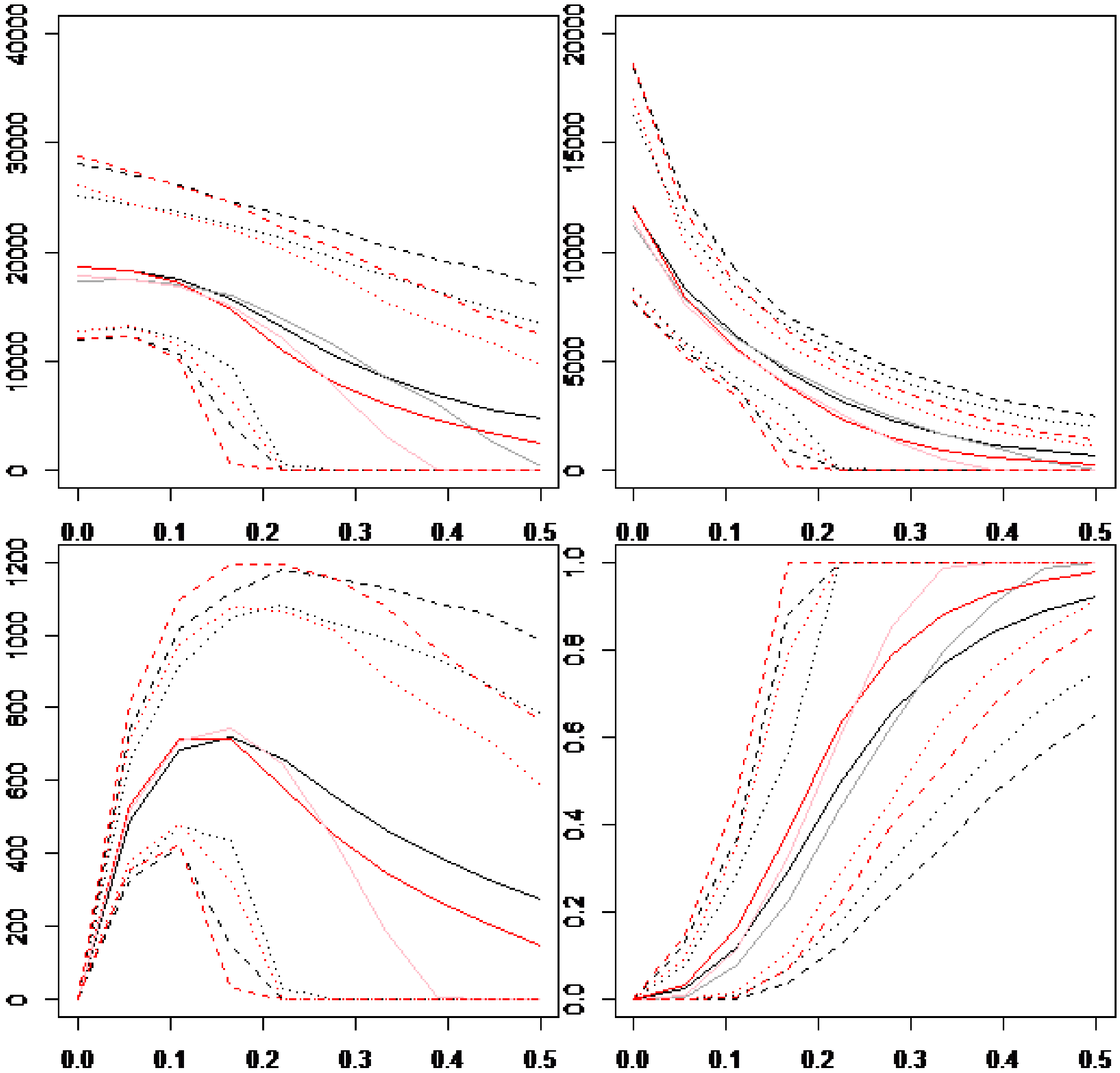


Figure 23. Mean recruitment (upper left), SSB (upper right), annual catch (lower left), and probability of falling below 2500 (lower right) using the AIC smoothed stock recruitment function as a function of average fishing mortality ages 5-11 (x-axis) assuming noise in recruitment follows an AR (1) model. The solid lines are the mean, 90% confidence interval (dotted lines) and 95% confidence intervals (dashed lines) accounting for the variability in stock recruitment parameters (see text). The black lines correspond to the results obtained by using the unweighted F over ages 5-11 while the red lines corresponds to using the weighted F over the same age range.

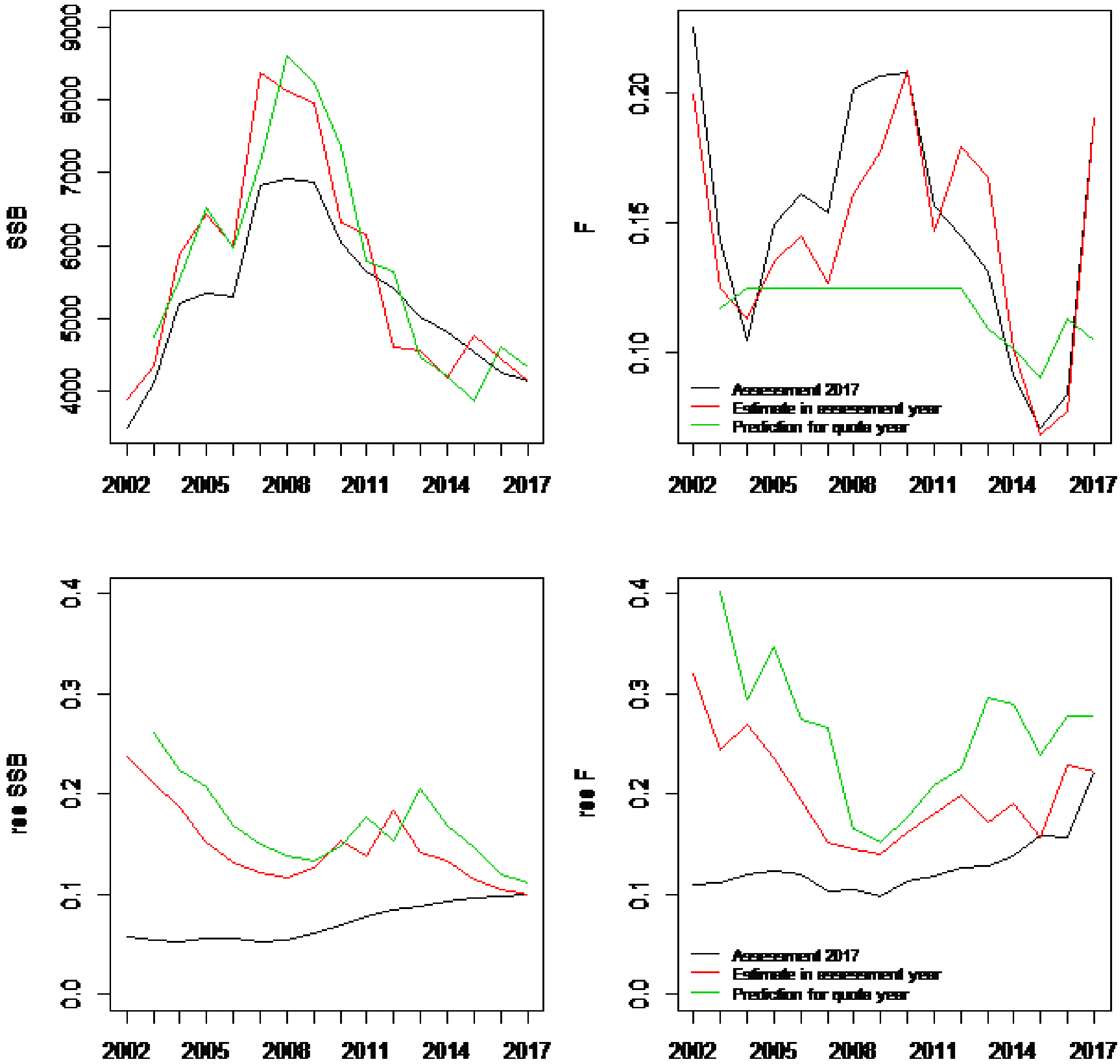


Figure 24. Assessment and prediction error by the XSAM model for SSB (left column) and average fishing mortality (right column). Black lines are from the XSAM assessment in 2017. The red lines are the estimates in the respective assessment year obtained by a retrospective run of XSAM. The green lines shows the resulting predictions (estimates in the quota year) made one year before, obtained by a retrospective run by XSAM making predictions as in last assessment.

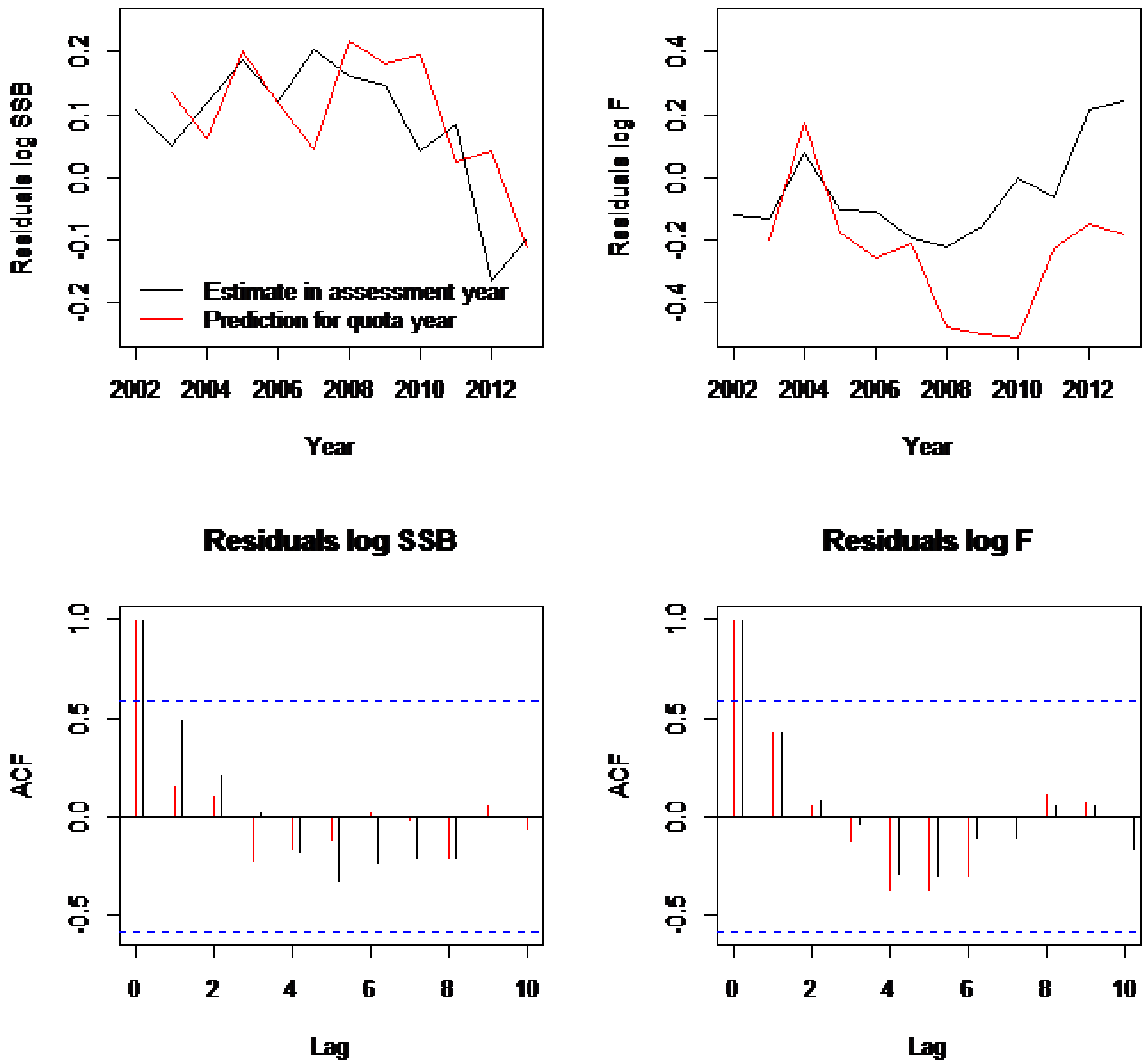


Figure 25. Residuals for log estimates in assessment year (black) and log prediction for quota year (red) for log SSB (top left) and log average fishing mortality (top right). Values for SSB and F are taken from the current assessment of historical values and residuals calculated from retrospective run of the assessment model including predictions. The two lower panels show the corresponding estimated auto correlation function. The correlation between residuals for log SSB and log F in the assessment year is -0.86 ($p<0.001$), while for the prediction is -0.52 ($p=0.100$). Mean residual for log SSB in assessment year is 0.080 with standard deviation 0.111. Mean residual for log SSB for prediction is 0.101 with standard deviation 0.101. Mean residual for log F in assessment year is -0.047 with standard deviation 0.152. Mean residual for log F for prediction is -0.247 with standard deviation 0.197.

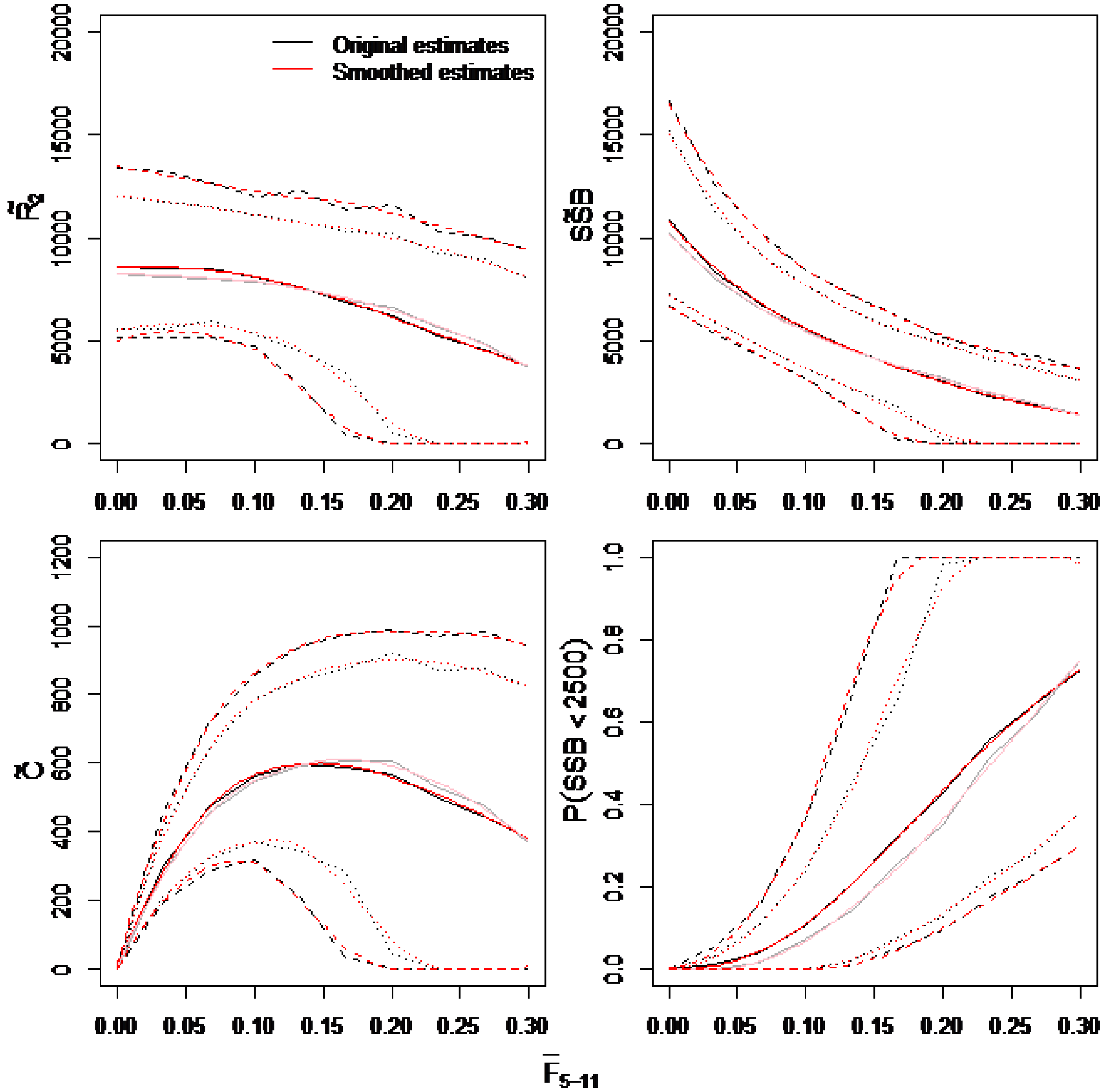


Figure 26. Comparing the original mean and percentiles in the estimated distribution of recruitment, SSB, annual yield and probability of SSB falling below 2500 in a specific year by specific values of F (equidistant sequence from 0-0.3 of length 10) (black) with smoothed estimates of the same for ~continuous F (equidistant sequence from 0-0.3 of length 100). Solid black and red lines are means, gray and pink corresponding median values, dotted lines correspond to 80% confidence and dashed lines correspond to 90% confidence intervals, respectively.

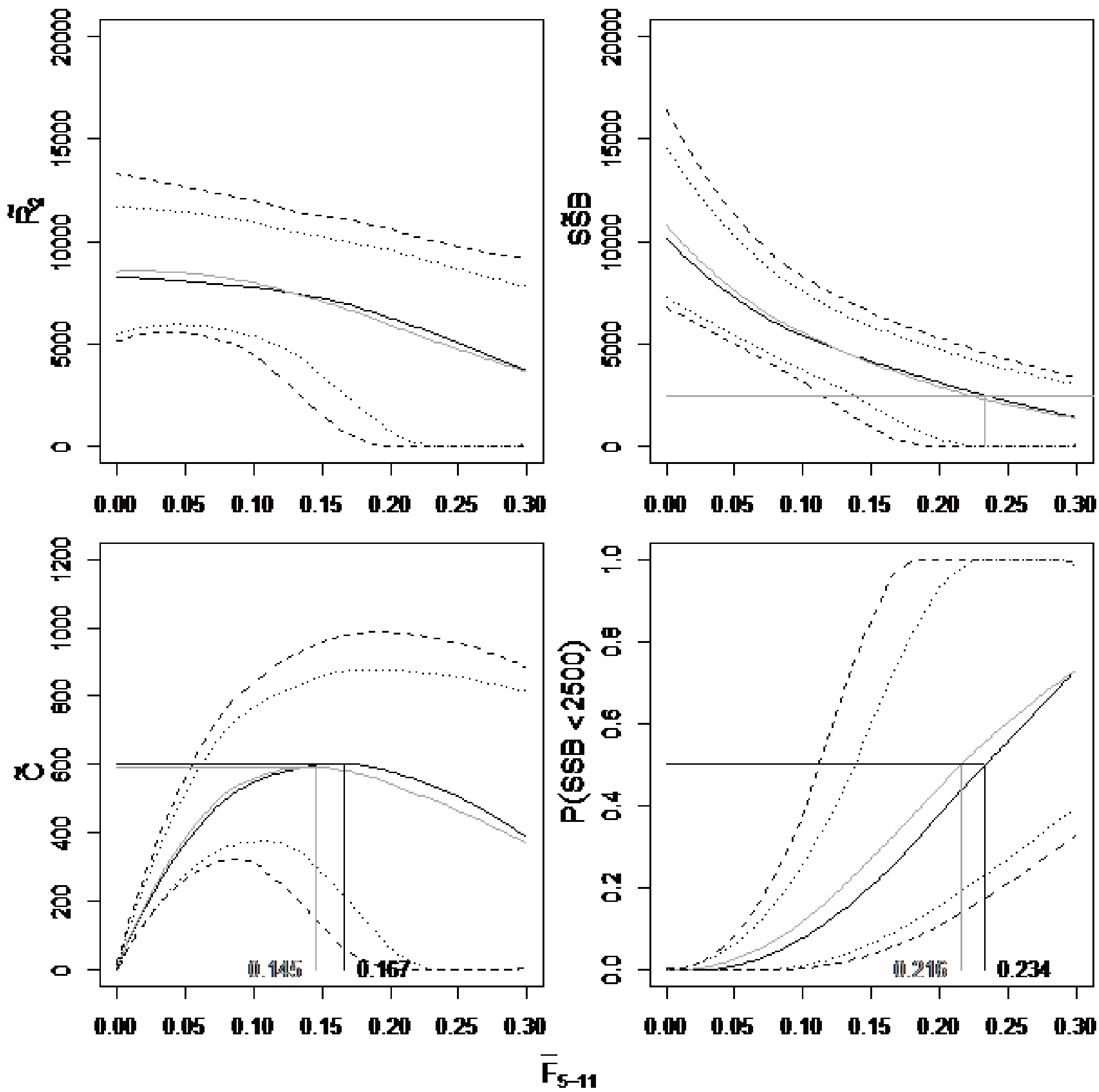


Figure 27. Median recruitment, SSB, annual catch, and long term probability of SSB being below 2500 a specific year as a function of a constant F (average over ages 5-11). Solid black line is the median, while the gray line is the mean. 90% and 80% confidence intervals are shown as dashed and broken lines, respectively. The simulations are run without assessment/prediction error and no $B_{trigger}$ point.

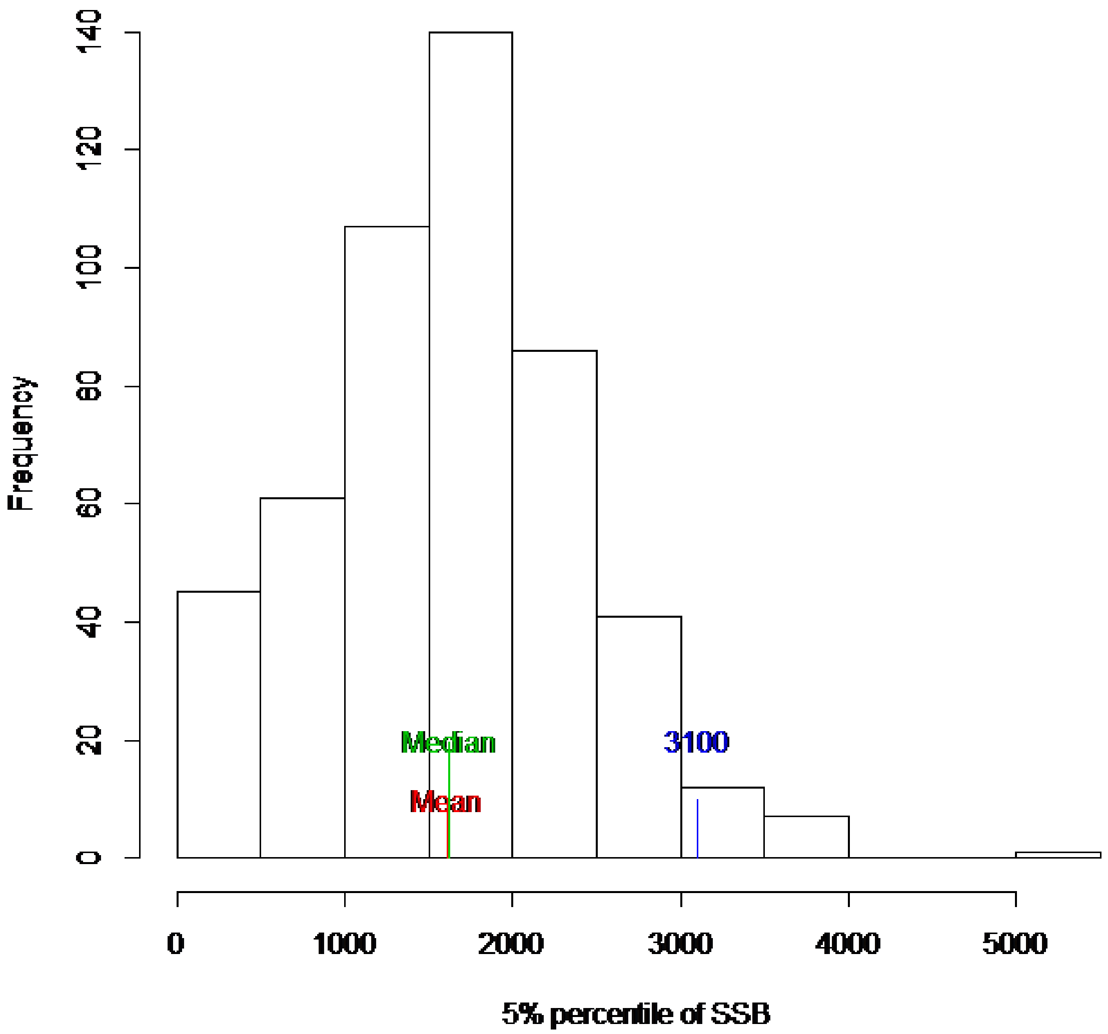


Figure 28. Histogram of 5% percentiles of SSB obtained by applying a constant F=0.15 (corresponding to the average over ages 5-11). The simulations are performed without assessment/prediction error and and no $B_{trigger}$ point. The mean and median of the percentiles is indicated as well as the proposed value of Blim=3100.

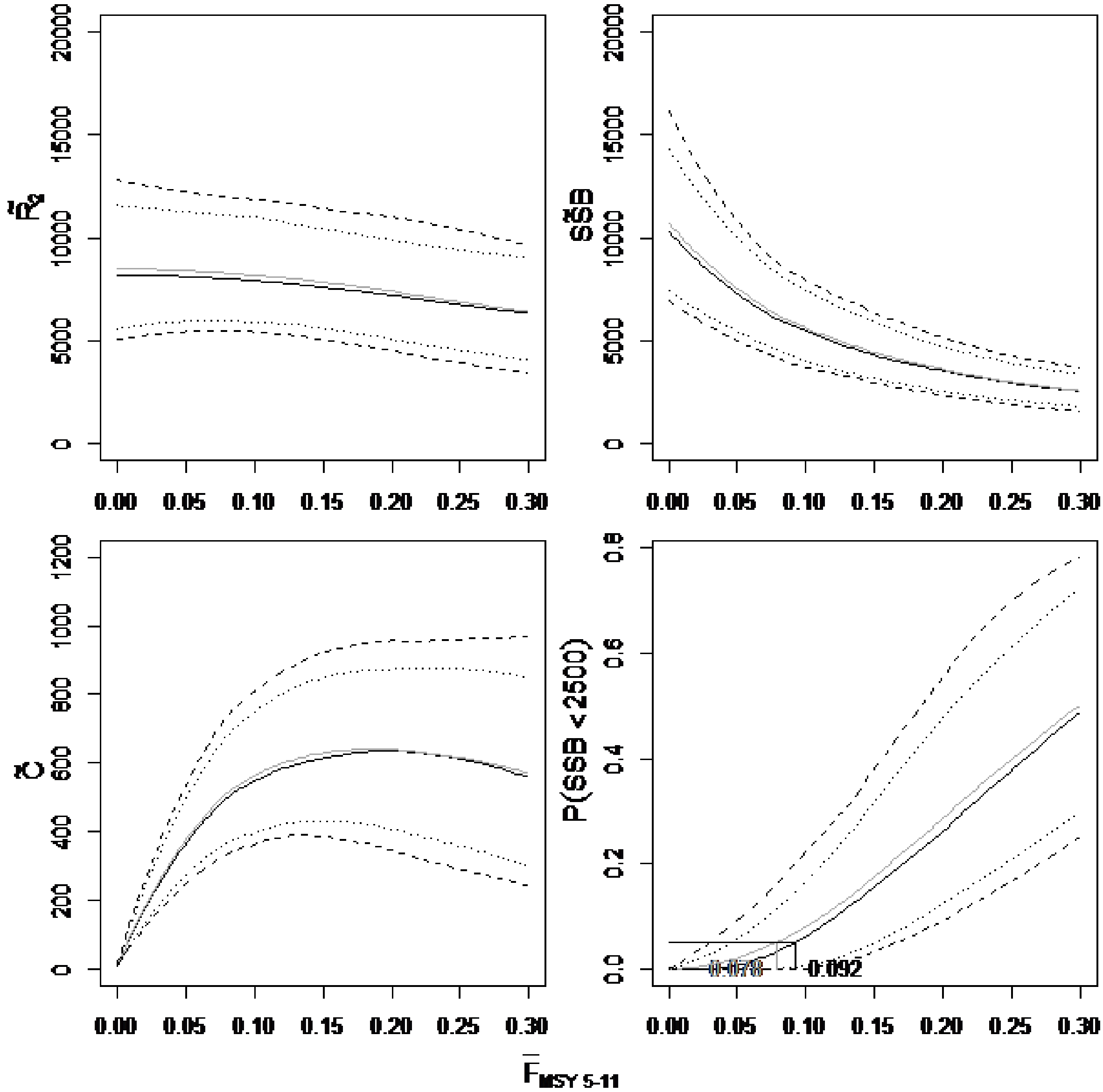


Figure 29. Median recruitment, SSB, annual catch, and long term probability of SSB being below 2500 a specific year as a function of a constant F (average over ages 5-11). Solid black line is the median, while the gray line is the mean. 90% and 80% confidence intervals are shown as dashed and broken lines, respectively. The simulations are run with assessment/prediction error corresponding to CV F=0.26 and CV SSB=0.16 and $B_{trigger}$ =3100.

# $B_{lim}$ for Norwegian spring spawning herring

Gjert E. Dingsør

The stock perception of Norwegian spring spawning herring (NSSH) has changed with the introduction of XSAM, both before and after the stock collapse (figure 1). Different models give different perceptions. Thus, there is a need to revisit the $B_{lim}$ value. The present $B_{lim}$ value of 2.5 million tonnes was set equal to the old minimum biological acceptable level (MBAL) in 1998. The MBAL value was based on the first VPA performed on this stock (Dragesund and Ulltang 1978). This study indicated that the decline in spawning stock biomass caused the poor recruitment from 1967 and onwards. Dragesund and Ulltang (1978) argued that a ”*critical level for the Norwegian spring spawning herring may be of the order of 1-2 million tonnes*.”

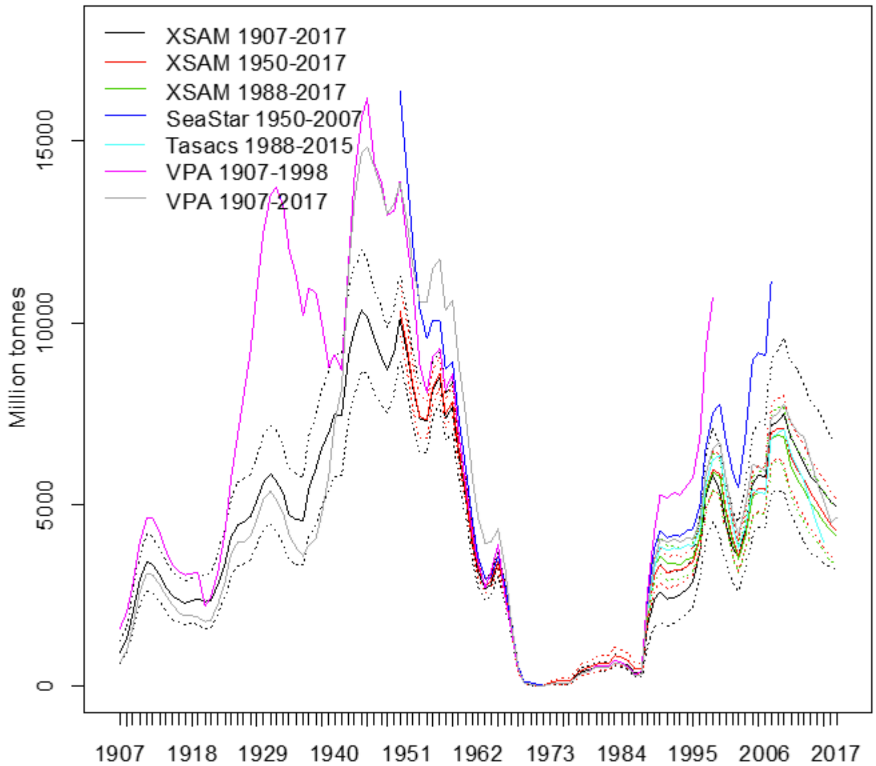


Figure 1. Estimated spawning stock biomass by different methods and time series of data indicated by the legend. The XSAM model fits includes approximate 95%confidence intervals shown by the broken lines. Aanes pers. com. (2018).

When determining MBAL values it was common to include a buffer and this makes MBAL more comparable to the definition of $B_{pa}$. The definition of $B_{lim}$ is more precise and $B_{lim}$ is not supposed to include a buffer; ”*A deterministic biomass limit below which a stock is considered to have reduced reproductive capacity*.” Thus, the biological basis for the current value of 2.5 million tonnes is poor, outdated and questionable.

The NSSH stock dynamic at low fishing pressure is determined by large fluctuations in recruitment (figure 2). NSSH has highly variable recruitment with occasional very strong year classes. The ICES guideline (ICES 2017) classifies stocks according to their historic recruitment and uses this classification to determine the reference point $B_{lim}$. Stocks with occasional strong year classes (spasmodic stocks) are classified as Type 1 and according to the guideline, $B_{lim}$ should be "*based on the lowest SSB where large recruitment is observed*". This advocates that $B_{lim}$ for NSSH should be based on the 1983 year class (figure 3).

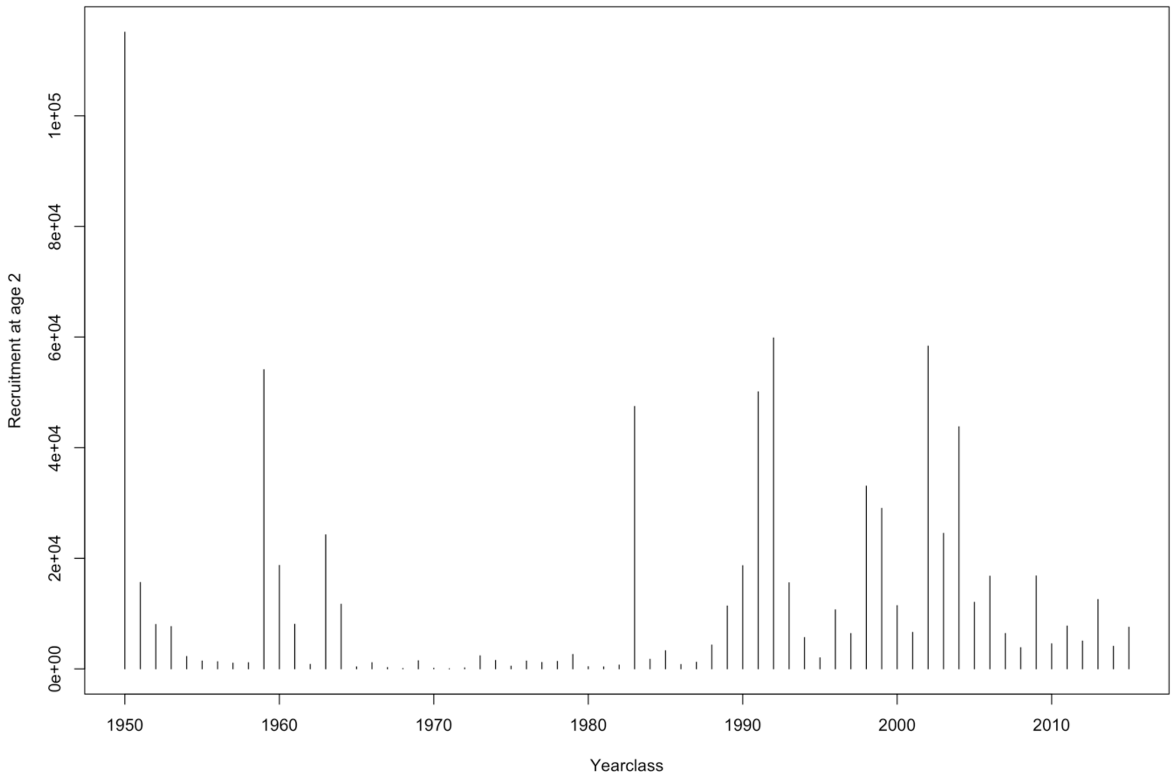


Figure 2. Norwegian spring spawning herring recruitment at age 2 (XSAM 1950-2017).

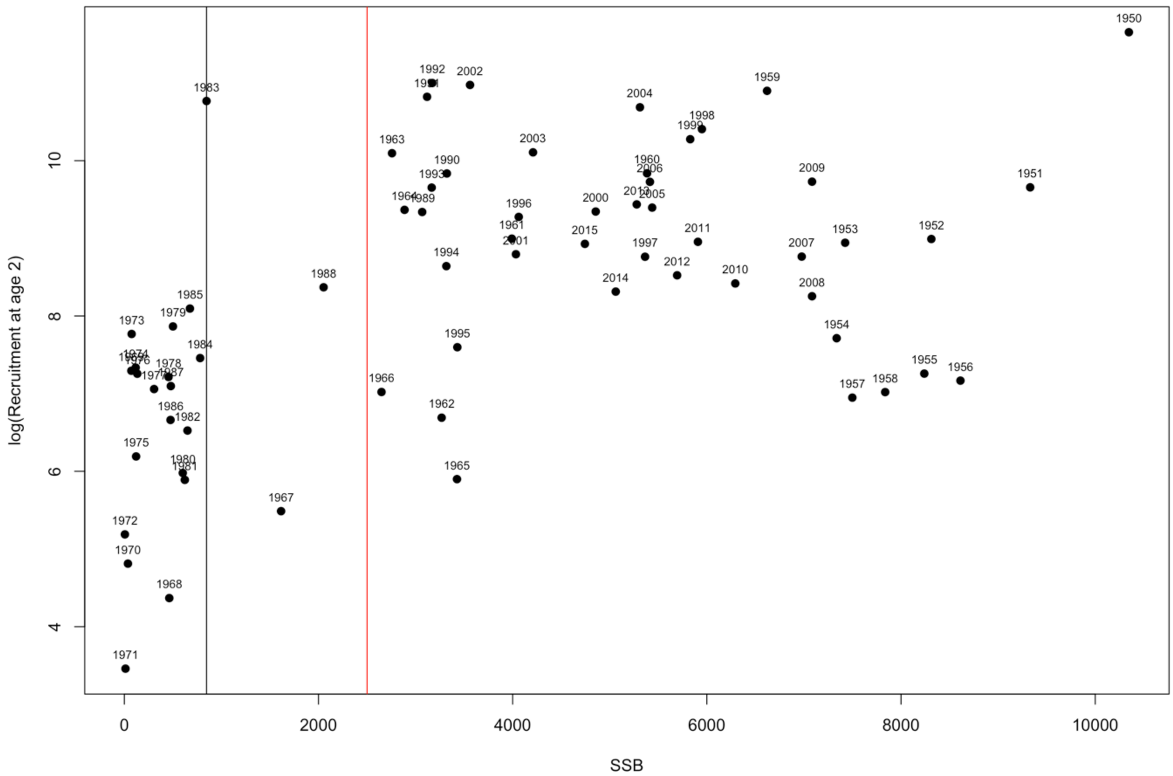


Figure 3. Norwegian spring spawning herring log recruitment – spawning stock relationship, SSB in 1983 = 847 thousand tonnes (black vertical line) and current $B_{lim}$ = 2500 thousand tonnes (red vertical line), XSAM 1950-2017.

Studies have shown that 1983 was a special year with rare environmental conditions (e.g Skagseth et al 2015) and this is used as an argument against using the 1983 SSB as the level for Blim. However, the 1984 and 1985 year classes that originate from about the same SSB levels, do also fall within the expected range of recruitment for a stock that is not impaired (figure 3 and 4) and support a $B_{lim}$ of 850 thousand tonnes.

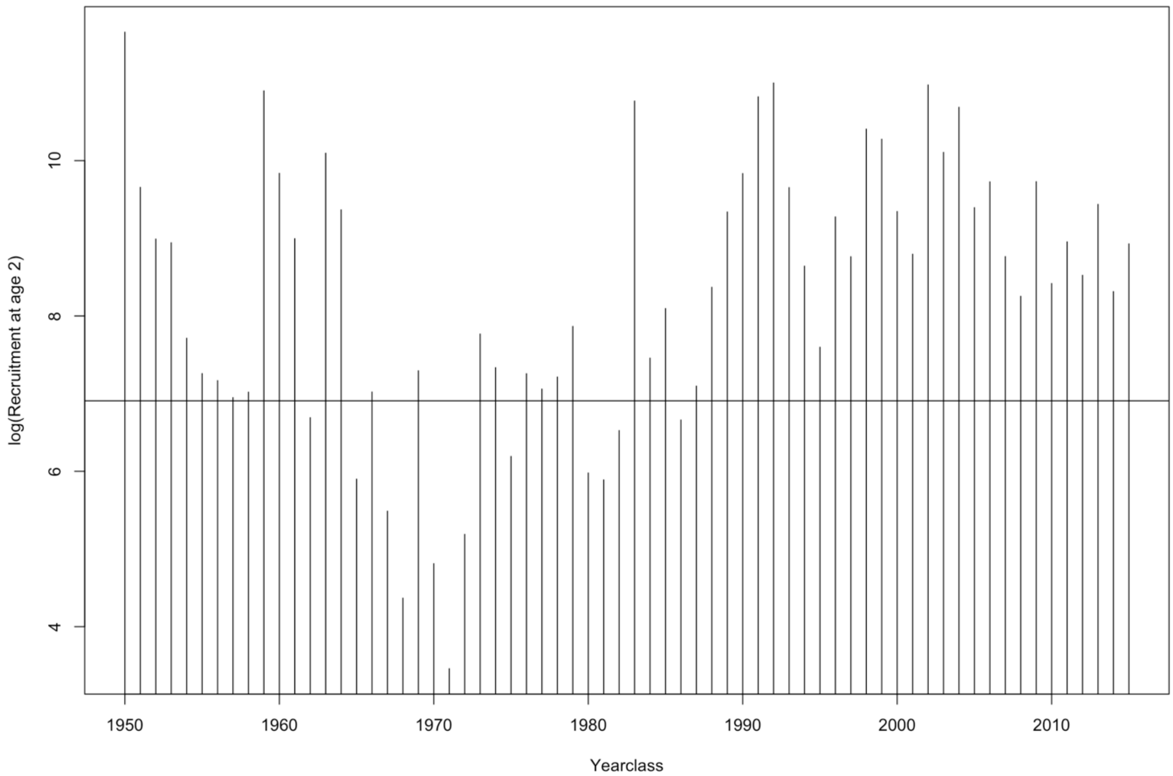


Figure 4. Norwegian spring spawning herring log recruitment at age 2, horizontal line show recruitment at $10^9$ individuals (XSAM 1950-2017).

It has been argued that because NSSH has gone through a stock collapse that resulted in impaired recruitment, the stock should be classified as Type 2. This argument is not supported by the guideline, on the contrary, it states that if the full range of SSB has not been explored, it is a Type 6. A Type 1 stock may have shown impaired recruitment, but the S-R is difficult to describe because of the occasional strong year classes that drives the stock dynamics.

The recommended method for determining $B_{lim}$ for a Type 2 stock is segmented regression. There have been several attempts to apply segmented regression on NSSH (e.g. ICES 2007; 2013; 2016) and all show that the estimation of the break point is unstable and varies from 1-3 or 2-4 million tonnes. The segmented regression technique requires a well-defined break point and WGWIDE states in the stock annex that the use of segmented regression is not appropriate for this stock, with reference to WKREF (ICES 2007). Still, at every review of $B_{lim}$, segmented regression has been used. It needs to be acknowledged that segmented regression is not an appropriate method for this stock and an alternative method should be used. This is also stated in the guideline "*If the performance of the segmented regression analysis is found to be unsatisfactory, …, alternative approaches for estimating $B_{lim}$ should be investigated.*" The definition of Type 1 has the best fit to the NSSH stock – recruitment relationship.

A less justified alternative is to use the data from after the rebuilding period, disregarding all data prior to 1988. The stock does not show any S-R relationship for this period and $B_{loss}$ (SSB in 1988) can be used to define $B_{lim}$, a Type 5 approach (figure 5).

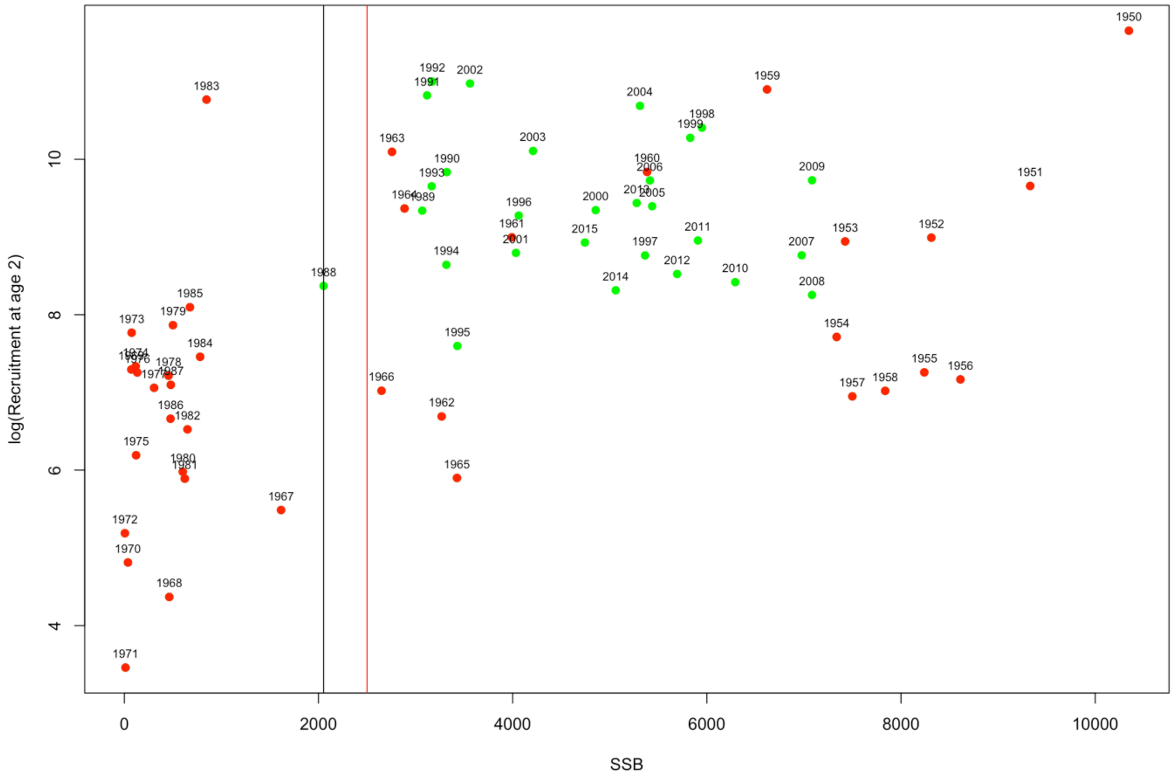


Figure 5. Norwegian spring spawning herring log recruitment – spawning stock relationship, 1950-1987 red points, 1988-2015 green points, SSB in 1988 = 2052 thousand tonnes (black vertical line) and current Blim = 2500 thousand tonnes (red vertical line), XSAM 1950-2017.

My conclusion is that NSSH should be classified as a Type 1 stock and SSB in 1983 should be used as $B_{lim}$.

## References

Dragesund, O., and Ulltang, Ø. 1978. Stock size fluctuations and rate of exploitation of the Norwegian spring spawning herring, 1950-1974. Fiskeridirektoratets Skrifter Serie Havundersøkelser, 16, 315–337.

ICES. 2007. Report of the Workshop on Limit and Target Reference Points (WKREF), 29 January – 2 February 2007, Gdynia, Poland. ICES CM 2007/ACFM:05. 89 pp.

ICES. 2013. Report of the Blue Whiting/Norwegian Spring-Spawning (Atlanto-Scandian) Herring Workshop (WKBWNSSH), 11–13 March 2013, Bergen, Norway. ICES CM 2013/ACOM:69. 88 pp.

ICES. 2016. Report of the Benchmark Workshop on Pelagic stocks (WKPELA), 29 February – 4 March 2016, ICES Headquarters, Copenhagen, Denmark. ICES CM 2016/ACOM:34. 106pp.

ICES. 2017. ICES Advice Technical Guidelines. 12.4.3.1 ICES fisheries management reference points for category 1 and 2 stocks. DOI: 10.17895/ices.pub.3036

Skagseth, Ø., Slotte, A., Stenevik, EK., and Nash, RDM. 2015. Characteristics of the Norwegian Coastal Current during Years with High Recruitment of Norwegian Spring Spawning Herring (Clupea harengus L.). PLoS ONE 10(12): e0144117. doi:10.1371/journal.pone.0144117

# Norwegian spring spawning herring
Estimation of reference points.

Working document 03 for WKNSSHREF 2018

Höskuldur Björnsson

April 6th 2018

The work shown here is just an update of the work done 3 years ago and described in working documents 13, 9 and 1 in WKPELA 2016 (WD 9 and 13 were also put on the sharepoint for WKNSSH-2018). The model that is described in WD-13 has been used for HCR evaluation for many other stocks both, last time NEA mackerel. The model has not changed sinces 2 years ago but the data have changed as 2 more years of data were added and survey 1 was not included in the work 2 years ago. The prediction part for a F rule has changed from 2 years ago when the F was implemented as a F-multiplier in the advisory year (the year following the assessment year). Now the stock in the assessment year is multiplied by an assessment error and the "perturbed stock" simulated one year using the TAC from last year. The predicted "perturbed stock" is then used to calculate the TAC for the advisory year. In the end the "real stock" is projected one year using the TAC generated last year. The assessment error used here is therefore the uncertainty in the stock biomass in the beginning of the assesment year.

Most of the runs done here were just updated since 2016 but few more options added. As an example the simulation periods in earlier work where either 1975-2014 or 1907-2014 but here the periods 1975-2016, 1950-2016 and 1907-2016 were investigated . The period 1935 - 2016 could also be investigated but a problem with the data before 1935 is that mean weights at age are constant. Also some of the runs were based on age 12 as plus group to be in line with Sondres work and the use of 1950-2016 is also to be in line with Sondres work. Runs with variable number of selection patterns (in time) are shown and also VPA runs that use $F_{15}$ from forward running model as F for the oldest group. The VPA is always run with 15 as plus group. VPA models do not handle plus groups well and when they are used the plus group should be relatively small. For this stock assumptions (or estimate) of the abundance in the plus group have often large effect on stock size (even when the plus group is 15) and relatively large difference in historical stock size and recruitment between different runs can be observed.

WD-9 from 2016 shows more details about the runs, stock - recruitment functions etc, what is shown here are mainly summaries.

## 1 Reference points

For this stock $B_{lim}$ was set to 2.5 million tonnes in 199?. After the collapse the first large yearclass (1983) increased the spawning stock from 600 thous. to 3 million tonnes in 2 years so relatively little information is available from recent data on exactly where the break point in a Hockey stick function is.

Therefore, older data are used with the known limitation that selection pattern in earlier period is very different from what is has been last 3 decades, with substantial targeting of ages 0-2 that are not at all caught recently. Catches of age 0 are were not included in the runs from 2016 but they were tested to have relatively small effect on estimated reference points while including age 1 changed more. Including age 0 could though have more effect when running from 1950 (not done in 2016) as the catch of age 0 was relatively high in the period 1950-1965. The value of assumed M for ages 0-2 (0.9) might have an effect here, high M makes the effect of fisheries on ages 0-2 less.

| | FirstY | nsel | age | rmax | ssbbr | cvbr | CV | acf | cvacf |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1975 | 1 | 1-15 | 69.4 | 2402 | 0.11 | 0.99 | 0.00 | 0.00 |
| 2 | 1975 | 1 | 1-15 | 69.4 | 2402 | 0.12 | 1.00 | 0.20 | 0.78 |
| 3 | 1975 | 4 | 1-15 | 64.6 | 2242 | 0.10 | 1.01 | 0.00 | 0.00 |
| 4 | 1975 | 4 | 1-15 | 64.6 | 2233 | 0.11 | 1.01 | 0.21 | 0.73 |
| 5 | 1975 | 1 | 1-12 | 71.8 | 2494 | 0.35 | 1.00 | 0.00 | 0.00 |
| 6 | 1975 | 1 | 1-12 | 72.1 | 2535 | 0.41 | 1.01 | 0.21 | 0.73 |
| 7 | 1975 | 4 | 1-12 | 66.5 | 2382 | 0.11 | 1.02 | 0.00 | 0.00 |
| 8 | 1975 | VPA | 1-15 | 75.2 | 3153 | 0.37 | 1.08 | 0.00 | 0.00 |
| 9 | 1975 | 1 | 1-15 | 68.8 | 2306 | 0.34 | 0.97 | 0.00 | 0.00 |
| 10 | 1950 | 5 | 1-15 | 56.5 | 2238 | 0.36 | 1.33 | 0.00 | 0.00 |
| 11 | 1950 | 5 | 1-15 | 58.1 | 2324 | 0.13 | 1.35 | 0.30 | 0.40 |
| 12 | 1950 | 5 | 1-12 | 59.9 | 2443 | 0.49 | 1.36 | 0.32 | 0.38 |
| 13 | 1950 | VPA | 1-15 | 65.8 | 2485 | 0.34 | 1.26 | 0.00 | 0.00 |
| 14 | 1907 | 6 | 1-15 | 69.0 | 2327 | 0.11 | 1.14 | 0.00 | 0.00 |
| 15 | 1907 | 6 | 1-15 | 68.7 | 2294 | 0.31 | 1.16 | 0.30 | 0.32 |
| 16 | 1907 | 6 | 1-12 | 70.9 | 2380 | 0.20 | 1.17 | 0.31 | 0.30 |
| 17 | 1907 | VPA | 1-15 | 77.4 | 2887 | 0.11 | 1.09 | 0.00 | 0.00 |

Table 1: Parameters of a hockeystick stock-recruitment function for various model settings and data

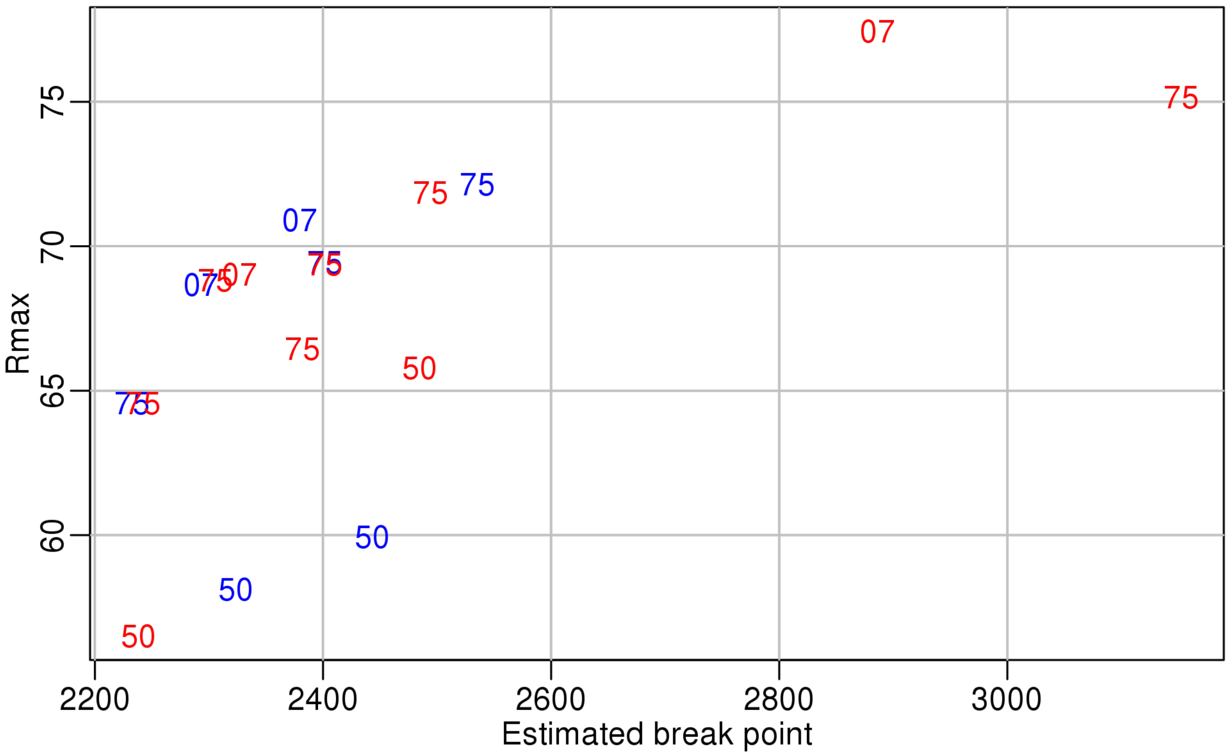


Figure 1: $R_{max}$ as function of $SSB_{break}$. Text shows first data year. Blue values indicate runs with first order autocorrelation estimated

Looking at the relationship between $SSB_{break}$ and $R_{max}$ the usual positive relationship appears (figure 1). The runs starting in 1950 show lower estimated $R_{max}$ indicating relatively low productivity in the period 1950-1975, something that is probably expected (exclusion of age 0 from the catches might explain part of the difference).

The VPA runs give the highest values of $SSB_{break}$ and $R_{max}$. Estimation of $\rho$ turned out to be relatively unstable in connection with VPA so no VPE runs with estimated $\rho$ are presented.

Standard error in $SSB_{break}$ is somtime relatively low ($\approx 0.1$). This is the standard error obtained from the Hessian matrix, standard error from mcmc simulations is somewhere around ($\approx 0.3$). The reason for this problem is not clear.

The main conclusion is that $B_{lim}$ is close to the current value of 2500 thous. tonnes. It could be argued that taking into account positive correlation between $SSB_{break}$ and $R_{max}$ higher $B_{lim}$ should be used in high

$R_{max}$ runs, something that does not fit well into current framework for advice.

## 2 Assessment results

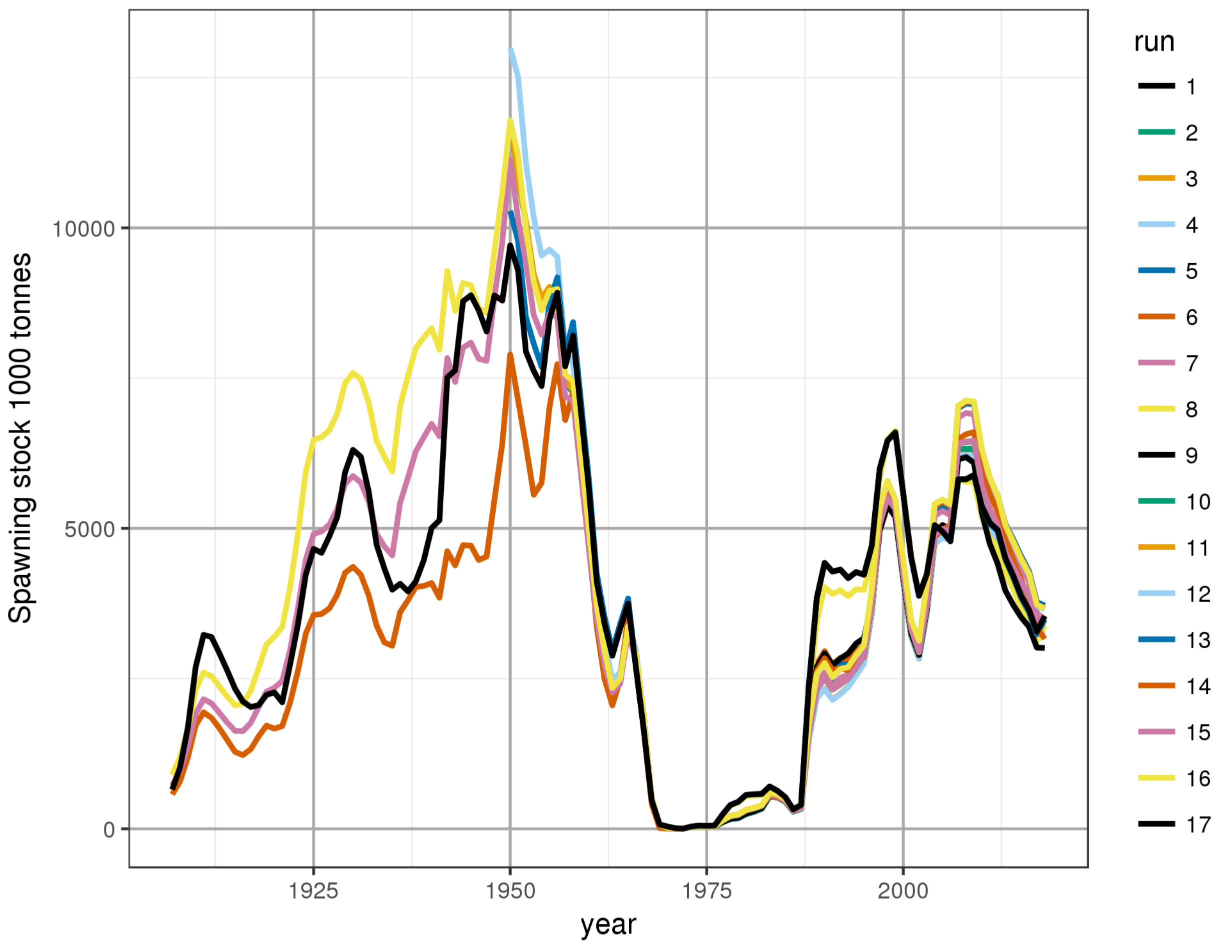


Figure 2: Spawning stock from different runs. Numbers refer to table 1

Spawning stock from differrent runs is shown in figure 2. Many of the runs lead to exactly the same historical results (those with and without estimated $\rho_{rec}$). The runs with the highest histororical biomass are the VPA runs.

## 3 Estimation of $B_{pa}$ and MSY $B_{trigger}$

The formula for $B_{pa}$ is $B_{pa} = B_{lim} \times e^{1.645*\sigma}$ where $\sigma$ is the standard error of estimated $SSB$ in the assessment year. $\sigma$ is approximately 0.12 based on a model tuned with surveys 1 and 5 but 0.14 when only tuned with survey 5. The increased precision obtained by survey 1 is questionable, are the surveys independent in the context that part of residuals have nothing to do with the surveys but rather the fact that the model is wrong (structural error). Also survey 1 has well settled estimate of $q$ from early years but some very large gap in the timeseries, the largest from 2008-2014. Therefore 0.15 or higher would be the correct value to use. That will lead to $B_{pa} = 3200kt$. The highest value used for $B_{trigger}$ in the simulations presented was 3000 kt, lower than candidate $B_{pa}$.

## 4 Estimating $F_{msy}$

Simulations were conducted based on the model configurations shown in figure 2 and table 1. CV of assessment error was set to 0.2 based on estimated model uncertainty and analytical retros (work done in 2015 excluding survey 1). This assessment error applies to biomass in the assessment year but the model takes care of the "amplification of uncertainty" through the assessment year. Autocorrelation of assessment error was set to 0.7

based on analysis of retrospective pattern. Autocorrelation of recruitment was set to 0.35 (estimate in R) or as estimated when estimation of first order AR model was included. Mean weight at age was stochastic around the average of last 20 years.

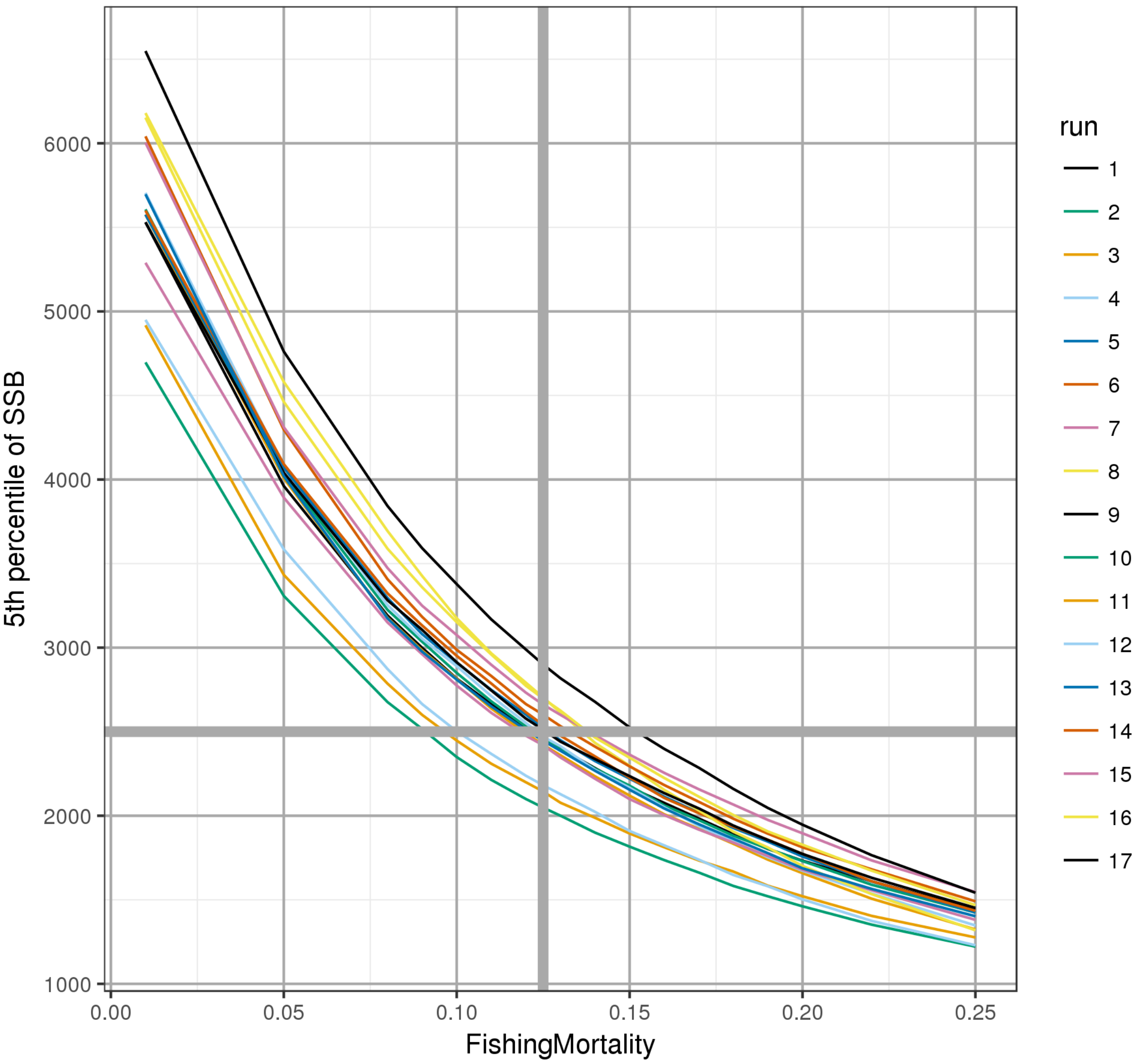


Figure 3: Fifth percentile of SSB as function of target fishing mortality, using $B_{trigger}$=3 million tonnes Numbers refer to table 1

$P(SSB < B_{lim}) < 0.05)$ is the limiting criterion in determinition of $F_{msy}$ for this stock. Based on $B_{trigger} = 3000$ (closest to $B_{pa}$) the range of estimated $F_{05}$ is between 0.1 and 0.15 but most of the values are close to 0.125.

This is further shown in table 2 where the results of the simulations are summarized. In this table $Fmsy1$ is F leading to maximum median yield, $Fmsy2$ F leading to maximum average yield, $F05a$ F leading to fifth percentile of the spawning stock $= B_{lim}$ when $B_{trigger} = 0$, *catchmed* maximum median catch and *catchmean* maximum average catch. Those values are all based on no $B_{trigger}$ while $F05$ is fishing mortality leading to fifth percentile of spawning stock $= B_{lim}$ when $B_{trigger}$=3 million tonnes. $F05$ would in all cases be what would be defined as $F_{msy}$ as it is lower than the values maximising median catch.

The VPA runs have the tendency to give highest yield and highest $F_{msy}$. The difference is most notable using the period since 1975. The difference between VPA and forward running models is largest when fishing mortality is low as it was when the 1983 yearclass was going through. Using the period since 1950 leads to lowest $F_{msy}$ but the catch is not nessecarily less. What makes the period from 1950 onwards special is extremely large contribution of one cohort (1950) and including that cohort leads more variablity in predicted recruitment. Unusually large catches of age 0 (not included) in that period might also have an effect.

$F = 0.125$ has been the target for over 20 years and is near the middle of candidates for $F_{msy}$. Continuing using $F = 0.125$ and reducing $B_{trigger}$ to 3 million tonnes does therefore seem like plausible option according to

| | FirstY | age | nsel | acf | Fmsy1 | Fmsy2 | F05 | F05a | catchmed | catchmean |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1975 | 1-15 | 1 | 0.35 | 0.204 | 0.204 | 0.121 | 0.113 | 852 | 1059 |
| 2 | 1975 | 1-15 | 1 | Est | 0.202 | 0.199 | 0.123 | 0.115 | 896 | 1174 |
| 3 | 1975 | 1-15 | 4 | 0.35 | 0.180 | 0.183 | 0.120 | 0.113 | 826 | 1109 |
| 4 | 1975 | 1-15 | 4 | Est | 0.184 | 0.183 | 0.123 | 0.111 | 893 | 1237 |
| 5 | 1975 | 1-12 | 1 | 0.35 | 0.199 | 0.198 | 0.127 | 0.119 | 869 | 1093 |
| 6 | 1975 | 1-12 | 1 | Est | 0.206 | 0.190 | 0.128 | 0.119 | 937 | 1206 |
| 7 | 1975 | 1-12 | 4 | 0.35 | 0.187 | 0.188 | 0.119 | 0.109 | 845 | 1040 |
| 8 | 1975 | 1-15 | VPA | 0.35 | 0.145 | 0.152 | 0.136 | 0.124 | 1002 | 1397 |
| 9 | 1975 | 1-15 | 1 | 0.35 | 0.194 | 0.192 | 0.125 | 0.119 | 818 | 1025 |
| 10 | 1950 | 1-15 | 5 | 0.35 | 0.202 | 0.227 | 0.090 | 0.085 | 818 | 1094 |
| 11 | 1950 | 1-15 | 5 | Est | 0.196 | 0.204 | 0.096 | 0.088 | 905 | 1292 |
| 12 | 1950 | 1-12 | 5 | Est | 0.194 | 0.212 | 0.100 | 0.090 | 891 | 1242 |
| 13 | 1950 | 1-15 | VPA | 0.35 | 0.188 | 0.206 | 0.121 | 0.112 | 892 | 1156 |
| 14 | 1907 | 1-15 | 6 | 0.35 | 0.206 | 0.225 | 0.132 | 0.125 | 921 | 1124 |
| 15 | 1907 | 1-15 | 6 | Est | 0.204 | 0.218 | 0.139 | 0.128 | 966 | 1184 |
| 16 | 1907 | 1-12 | 6 | Est | 0.202 | 0.212 | 0.137 | 0.129 | 973 | 1196 |
| 17 | 1907 | 1-15 | VPA | 0.35 | 0.191 | 0.203 | 0.152 | 0.141 | 939 | 1132 |

Table 2: Summary HCR/Fmsy evaluations

those analysis except bias in assessment in last decades is taken into account. Short term considerations might also lead to some lowering of F if type III risk is considered.

Looking at the results in table 2 maximum median catch is between 800 and 900 thous tonnes in the alternatives where data from 1950-2016 are used (same as in XSAM) but the mediancatch in XSAM is $\approx 15\%$ lower. Part of this difference seems to be caused by inclusion of age 1 that was heavily caught before the collapse. Adding age 0 has less effects.

Agverage catch from those runs is quite high (table 2) and might

# 5 Measures of fishing effort

Currently advice for this stock is based on weighted average fishing mortality of ages 5-11 where the fishing mortality is weighted by stock numbers. At the meeting other measures were discussed like unweighted fishing mortality or harvest rates. 3 different measures are shown in figure 4 all showing similar main trends. Deviations are related to large cohorts recruiting to the stock.

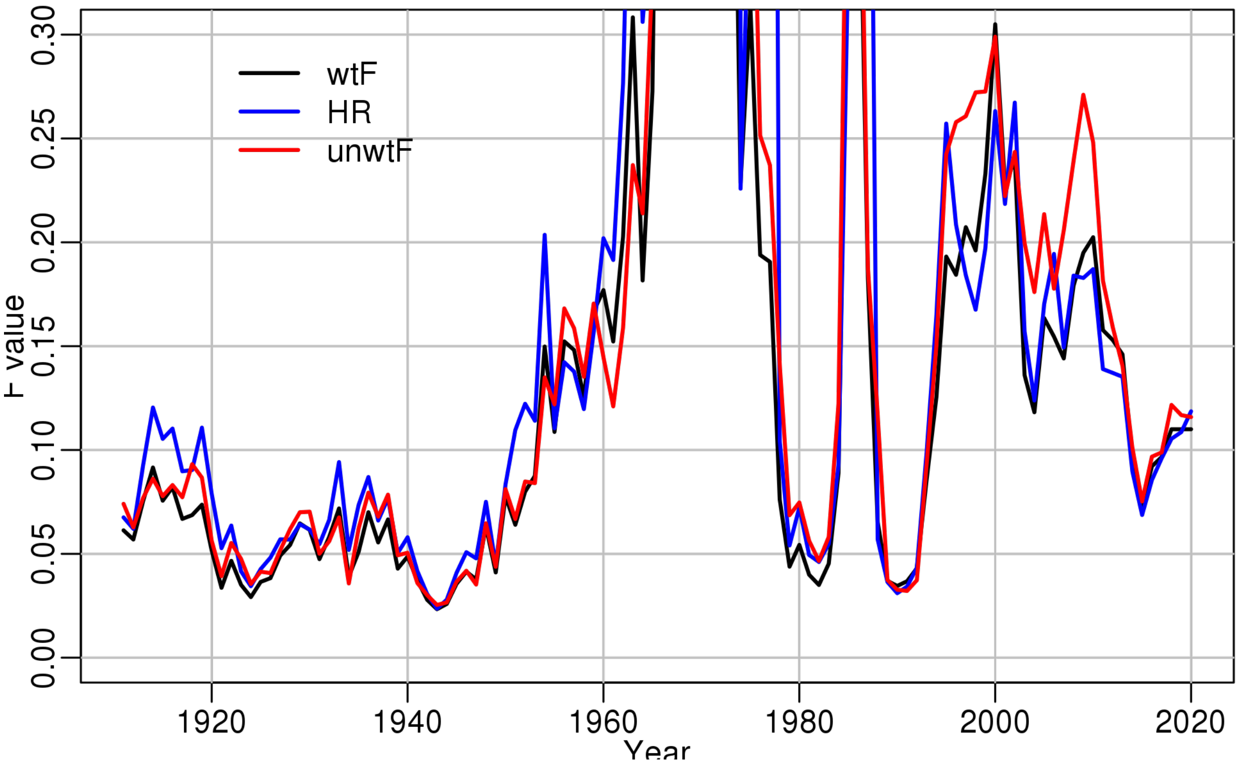


Figure 4: Development of different measures of fishing effort since 1907. High values outside any plausible management plan fall outside the plot. The measures shown are $F_{5-12}$ weighted by stock numbers, $F_{5-12}$ unweighted and harvest rate based on $B_{5+}$

The harvest ratio in figure 4 is shown as proportion of $B_{5+}$ but $B_{5+}$ is a reasonable proxy for the fishable stock and SSB. If the advice was based on biomass one year earlier (the assessment year) $B_{4+}$ might be a better candidate and some version of the HCR for Icelandic cod could be used.

Delay of maturity data would make $B_{5+}$ a good candidate for $B_{trigger}$, it is not the correct SSB but relatively close and it is available at the time of assessment. Still criteria in HCR simulations would be based on "real SSB".

# 6 Conclusions

Using estimated breakpoint from a Hockey stick fit as candidate for $B_{lim}$ is not a perfect solution but does a better method exist. For this stock the value of the break point turns out to be relatively robust to model settings but standard error of the estimate is close to 0.3. Compared to most other stocks the breakpoint is relatively well defined.

Basing runs on the timeperiod 1950-2016 makes the results sensitive to inclusion of catches of ages 0 and 1. Having to include those agegroups is in itself a problem as the effect of the fisheries on ages 0 and 1 depend much on the assued $M$ for those ages. In the runs shown here (table 2) maximum median yield is reasonably constant for different estimation periods.