ICES WKPELA REPORT 2016

ICES ADVISORY COMMITTEE

ICES CM 2016/ACOM:34

REF. ACOM

Report of the Benchmark Workshop on Pelagic stocks (WKPELA)

29 February–4 March 2016

ICES Headquarters, Copenhagen, Denmark

ICES CIEM
International Council for the Exploration of the Sea
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. 2016. Report of the Benchmark Workshop on Pelagic stocks (WKPELA), 29 February–4 March 2016, ICES Headquarters, Copenhagen, Denmark. ICES CM 2016/ACOM:34. 106 pp. https://doi.org/10.17895/ices.pub.5581

Contents

Executive summary

The Benchmark Workshop on Pelagic stocks (WKPELA) 2016 took place in two parts. A data compilation workshop, 24–26 November 2015, and the main meeting, 29 February–4 March 2016, both at ICES HQ.

There were 23 participants from 7 countries. Out of the 23 participants, 6 were on behalf of stakeholders. Additionally, three external reviewers, two from outside the ICES community, took part in the meeting and reviewed what was presented. The complete participants list can be found in Annex 1.

The purpose of the meeting was to examine the assessment of the Norwegian spring-spawning herring stock (NSSH), both in terms of data and methods. This benchmark meeting was the second, the first benchmark took place in 2008. In 2008 the VPA module in the TASACS toolbox was chosen to be the standard assessment tool in the next years. The assessments showed retrospective problems around 2010–14, but have been quite consistent in the last couple of years. An analysis showed that this problem was mainly caused by ages 13–15 in one survey. The TASACS assessment, however, did not perform well when the plusgroup was lowered to age 12.

WKPELA accepted the findings that there were no indications of increased natural mortality in a period after around 2009. In this respect, biological data from 626 749 individuals of NSSH were analysed. These had been collected internationally in the years 1994–2015 over the full distribution area and over the whole year.

Some of the surveys time-series have now been re-evaluated. The revised data seemed to improve the assessment, and WKPELA accepted these new survey indices. The revision of the survey data leads to higher spawning-stock biomass (SSB) in the years after 2004, compared to the final TASACS assessment in 2015. The upward revision is in the range of 500–800 thousand tonnes.

WKPELA accepted updated values for maturity-at-age in the years 2005–2011. The maturity data were last revised in 2010. WKPELA suggests updating the maturity values each year according to the procedure in WKHERMAT 2010. The revision of the maturity data alters the perception of the spawning-stock biomass in these years, lowering the peak in 2009, but raising the estimate in 2007.

Three assessment models were explored, TASACS, XSAM and one separable model. XSAM is a model template based on a state space model and structural time-series models for fish stock assessments (note that XSAM is not the same as SAM). The main new achievement in this framework is to utilize prior knowledge of sampling errors to a greater extent to improve inference than what has been possible earlier. WKPELA accepted XSAM as the standard assessment tool for the NSSH. However, TASACS, with the same settings as in 2015, shall be used as an alternative assessment, if for some reasons XSAM should fail. XSAM is still under development, so WKPELA suggests allowing for small adjustments if needed. A short-term forecast module is not yet ready, but will be before WGWIDE 2016. The code, a documentation, and an example of how to use the framework will be made available before WGWIDE 2016, at its SharePoint site.

The benchmarked assessment for NSSH uses the same years as the TASACS assessment, but age range is 3–12+ instead of 0–15+. Two surveys on the adult population are used in XSAM, instead of eight surveys on all life stages in TASACS. Broadly speaking,

SSB trends estimated by XSAM and TASACS are similar and the SSB estimated by TASACS lies within the 95% confidence interval from XSAM.

The reference point B_{lim} was re-examined. The evaluation did not change it, B_{lim} = 2.5 million tonnes is still considered appropriate. The F in the management plan was also re-examined. The current value of 0.125 is considered adequate and the management plan can still be considered consistent with the precautionary approach, but it has not been evaluated whether this will lead to maximum long-term yield.

The stock annex was updated according to the decided adjustment to the assessment input data and methodology. A description of the short-term forecast module in XSAM, and a detailed description of how to run XSAM are still outstanding, but will be ready before WGWIDE 2016 when the new assessment model will be used for the first time to provide catch advice.

1 Introduction

ACOM, under the advice of the assessment expert group WGWIDE recommended that the Norwegian spring-spawning herring stock should undergo a benchmark assessment in 2016. This is the second time the stock is in the benchmark process, since the first benchmark in 2008. An issue list (see below) of current assessment/data problems was proposed and formed the basis of the benchmark process.

The benchmark process was divided in two parts, a 3 day data compilation workshop (DCWK) which took place 24–26 November 2015 and a 5 day benchmark meeting 29 February–4 March, both at ICES HQ.

During the data compilation workshop (DCWK) in November a number of issues related to dataseries and assessment were discussed. The main message from that discussion was that the International Ecosystem Survey in May in Norwegian Sea (IESNS) was the backbone of the assessment. A re-calculation of some of the time-series had already begun. This is done with a new software StoX, which provides among other things confidence intervals around the point estimate.

A new model framework (XSAM) was introduced to the group at DCWK. It is a statistical model for estimating fish stock parameters accounting for errors in data. It is not the same as SAM, as the Random walks can be replaced with AR (1) models, the observation model can use estimated standard errors at age and optionally correlation too, and the process error is modelled differently. It was considered feasible to continue developing this framework for the NSSH, as then the observation model would give the datapoints weight. In the previous assessment the datapoints (survey indices) used have been manually chosen by the experts.

Between DCWK and WKPELA scientists continued with their work and for almost all tasks working documents were provided and uploaded to the SharePoint a week before WKPELA. The working documents were presented in the first 2 and a half days of the meeting and the main results are in this report.

At the meeting 3 assessment tools were available, TASACS, XSAM and a statistical catch-at-age model. Not all of them can handle the same input dataseries. The model performances based on different dataseries were explored and compared and after a discussion it was decided to use XSAM as the main assessment tool, however TASACS and the statistical catch-at-age model should be simultaneously run for a comparison. But, on the last day of WKPELA (4 March), the scientists still attending the meeting, realized that some issues remained with the XSAM model framework. Further work and some diagnostic were still needed to decide on which observation model to use, so a task list (see below) was made for the modeller/assessor to work on. 15th of March a WD was uploaded on the SharePoint and the results were discussed at a Skype meeting 17 March. All WKPELA participants were invited, but not all participated in the Skype meeting. The decisions are reported in chapter 3.1.

Issue list:

Issue	Problem/Aim	Work needed / possible direction of solution	Solved
(New) data to be Considered and/or quantified1	Additional M - predator relations	Quantifying the predation on herring 0-group and 1-3 age group by Arctic cod	Not addressed at the meeting.
	Ecosystem driver – inter and intraspecific competition and trophic cascades.	Competition within the pelagic complex – how strongly are the inter- and intra species interactions regulating the herring in population (dynamics and growth)?	This was dealt within WGINOR and Stock-annex updated accordingly, but not addressed at the meeting.
Tuning series	Several tuning series that are no longer updated. Are these still contributing to the assessment? Only the "strong" year classes have been used in the tuning for many years. For how long, that is up to which age, they have stayed in the tuning series has been decided by experts each year. The plus group is now at age 15 (that is 15+). Should it be at a lower age?	Studying the importance of including/excluding the old tuning series from the assessment. Some of the tuning series give misguiding information about year-class strength at age. Which age groups would be used in each series were chosen by "expert" knowledge during the benchmark 2008. During WGWIDE 2014 a presentation was given on how to use statistical methods to take the decision. That work should be done for all tuning series. Work was done regarding the plus group during the benchmark 2008. It should be revised based on the decision taken on how to choose data for the tuning series.	Dealt with at the meeting. 2.2.3
	Can the IESSNS survey be used as a tuning series for NSSH?	Exploring the inclusion of this tuning series.	Explored in DCWK. 2.2.3
	Unexplained variability/changes in the selectivity/catchability of the major fleet used for tuning the assessment. These seem to be causing retrospective patterns in the assessment.	An analysis of variability or changes in the catchability of fleet 5.	Explored in DCWK. Data quality in 2.2.1.

1

Discards	Should slipping be included in the assessment?	Collection of data and estimation of the importance of slippage for the total mortality	Dealt with in the meeting. 2.5
Biological Parameters	Maturity ogives for recent years should be updated following procedures described by WKHERMAT	Maturity ogives were revised in 2010 as a recommendation from WKHERMAT. Values based on year-class strength have been used since then. Those should be updated based on WKHERMAT procedure	2.4
	Hypotheses have been put forward regarding varying condition factor and varying natural mortality of adult herring among years.	Data exists from the feeding areas during early and late summer (survey data). To the extent possible, these hypotheses will be tested based on the available data. In addition, national biological samples of commercial landings are available, which will also be analysed.	2.2.1 and 2.3
	The NSSH assessment has shown a systematic bias, with an overestimation of on average 26% in the period 1997-2011.	Need for understanding what is driving the bias. Is it dependent on assessment model, or the nature of the stock?	Dealt with in the meeting.
	TASACS has been used as the assessment model since the benchmark 2008. ISVPA/TISVPA has been run for comparison.	Could different configurations of TASACS diminish the bias? This should be explored. Alternative assessment models should also bee explored.	Dealt with in the meeting.
	TASAC model currently in use only extends back to year 1988, although data are available for much longer time-series.	The main point is to get a good assessment that advice can be based on. Time-series from 1988 to present should be enough for that. An assessment (VPA) of NSSH exists 1907-1998 (Toresen & Östvedt) so it should be secondary to “get the history” by extending the time-series far back in time.	Dealt with in the meeting.
	NSSH is one of the few stocks in which weighted F's are applied. Can this be discontinued?	Consider the need to continue the use of weighted average F in the assessment and advice.	Considered during DCWK.

Short-term prediction	A determistic short-term program has been used for some years. It is done in excel.	A new short-term prediction model is needed. It will most likely depend on the assessmodel chosen. TASACS has a short-term module that has not been used. Short-term predictions based on SAM exsists too. Other short-term programs should be explored as well. If the assessment model chosen will still show a retrospective pattern, then it should be explored how the short-term model could take account it.	Partly dealt with in the meeting.
Biological Reference Points	These need to be updated if the perception of the stock changes due to changes in the asssessment.		Partly dealt with in the meeting.

Task list for assessment runs and diagnostics to be done before skype-discussion on 17th March 2016:

Need (in addition to diagnostics already discussed):

Add a table that includes, for each model run, the likelihood component values, parameter estimates (including the scaling parameter h), derived quantities, and corresponding standard deviations about the parameter estimates.

- Conduct a set of runs which each leaves out a single data source and include these runs in the table above
- Conduct a set of runs profiling over the scaling parameter h and include in the above table, plot the likelihood profile for the total likelihood, as well as each likelihood component

Add detail and more detailed equations to the description of observation models 0-2 (the correlation and covariance structures) to the working document.

Add equations for each likelihood component to the working document

Trace the causes of any strange or unexpected results (residual patterns, etc.).

2 Data

2.1 Catch data

A general description of the catch data are given in the stock annex. The Norwegian catch-at-age, which represents around 60 % of the total catch, is now estimated with the ECA software (described in Salthaug and Aanes 2014). In addition to point estimates of catch in numbers-at-age, the ECA software provides estimates of precision that also can be implemented in stock assessment models.

2.2 Survey data

Available surveys and their use in the assessment are described in the stock annex.

2.2.1 Evaluation of IESNS

The International Ecosystem Survey in May in Norwegian Sea (IESNS) provides the single most important time-series of fishery-independent data for analytical assessment of Norwegian spring-spawning herring. During this period the herring is registered and measured acoustically from the surface down to 500m depth. Catch samples are taken selectively from the acoustical registration to determine the species composition and length distribution and biology aspects of the target species. The acoustical registrations and biological samples are compiled over all depth ranges thus sampling effort with respect to depth can cause bias in the results if it is not representative of the depth distribution of herring. This potential bias was addressed by using survey data from 2007-2015 to test two hypotheses: (1) Depth distribution of biological samples is not representative of depth distribution of acoustical registrations; (2) Length distribution in uppermost layers differs from deeper layers. The results allowed us to reject both hypotheses for almost all areas, indicating that the biological sampling in the survey is representative for the acoustic registrations with respect to depth. It was only in two cases where such a bias might have appeared, and would have affected length distributions of 22-26% of total acoustic registrations in those two years. The results of this exercise are therefore considered as a quality stamp for the survey in this respect. One of the recommendations arriving from this is that the acoustical data should be delivered to the database at 10m depth channels in the future, which will allow for more thorough analyses and quantification of potential variation in quantity of fish in the acoustic dead zone and fish avoidance from the approaching vessels.

2.2.2 Recalculation of survey data with StoX

Some of the surveys, or parts of the survey time-series have now been re-calculated with the new software StoX, which provide estimates of precision that can be implemented in stock assessment models. A description of these re-calculations are given in ICES (2016) and WD6, WD8 and WD10.

The surveys which have been recalculated are IESNS in the years 2008-2015 (Figure 1.2.1); the IESNS in the Barents Sea, in the years 2009-2015 (Figure 1.2.2); the Norwegian acoustic survey on spawning grounds in February/March (NASF) in the years 1988-1989, 1994-2000, 2005-2008 and 2015-2016 (Figure 1.2.3) and the Norwegian herring larvae survey on the Norwegian shelf (NHLS (Figure 1.2.4).

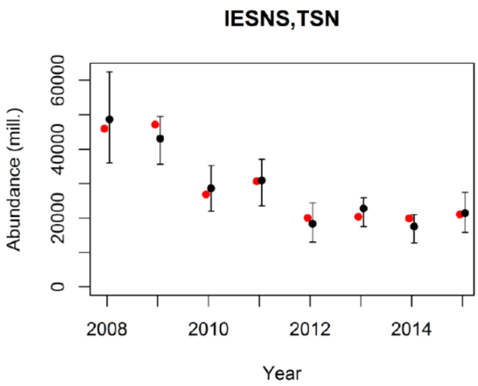

Figure 1.2.1. IESNS in the Norwegian Sea, total stock number. The black dots and error bands are StoX estimates with 95% confidence intervals. Red dots are old estimates (Beam). Figure is borrowed from ICES 2016a.

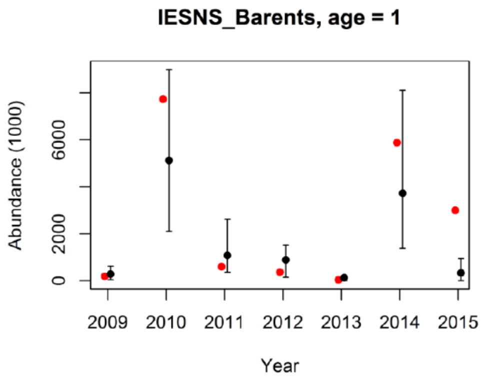

Figure 1.2.2. IESNS in the Barents Sea, age 1. See Figure 1.2.1 for further explanation. Figure from ICES 2016a.

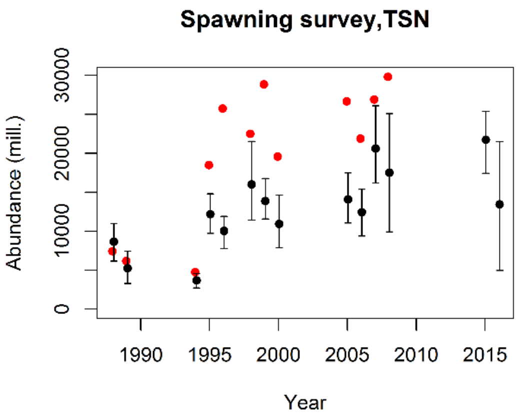

Figure 1.2.3. NASF on the spawning grounds along the Norwegian coast. For explanations see Figure 1.2.1.

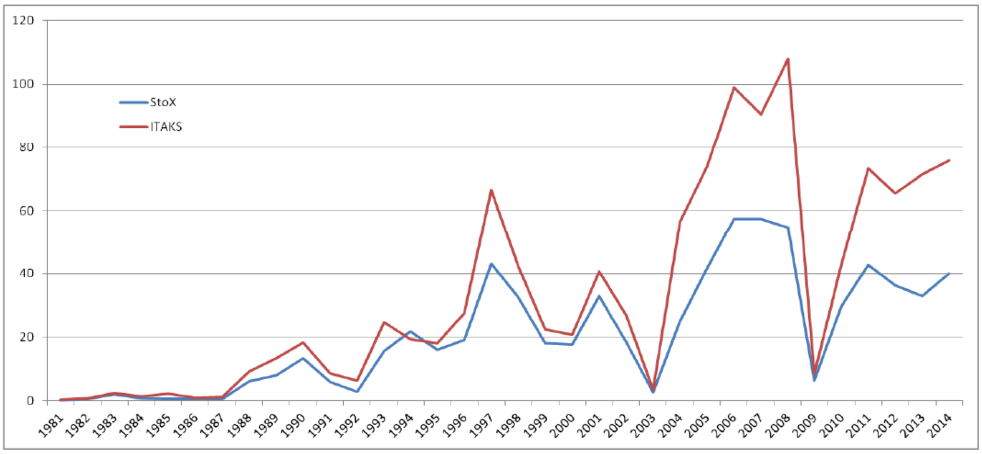

Figure 1.2.4. NSSH indices for herring larvae on the Norwegian shelf (NHLS) 1981–2014 (N * 1012).**

The point values from the re-calculations result in similar values to the old ones (Beam values) for IESNS in the Norwegian Sea. For the IESNS in the Barents Sea the old values lie within the 95% confidence interval in most years. For the NASF the old method seem to have overestimated the stock in the years 1994–2008. For 2015 and 2016 only StoX values are available. The recalculated index from NHLS is considerably lower than the old one, but shows a similar trend (2003 and 2009 are not valid data points).

2.2.3 Validation of time-series

During the benchmark in 2008, the age groups that should be used in the tuning from each survey series were chosen by "expert" knowledge. The key point has been that large year classes provide information while the small ones provide only noise, meaning the big ones were used in the tuning while the small ones not. During WGWIDE 2014 a presentation was given on how to use statistical methods to take the decision. The work has now been done for all tuning series.

Validation of the different survey time-series was done by Salthaug and Johnsen (2014), and the same analyses was carried out on new survey data (addition of years and re-calculated with StoX) give similar conclusions regarding which surveys and ages give valid signals of abundance trends. The main conclusions in Salthaug and Johnsen (2014) were that the surveys designed to measure the adult part of the stock are "approved" for ages (approximately) 3-12, except for IESSNS where no ages were approved. The two young fish surveys in the Barents Sea were approved for ages 1-3.

The working group felt that this was a good guiding tool, but not enough. Data and model are linked together.

2.3 Variations in condition of NSSH

A hypothesis was suggested that natural mortality had increased in a period after 2009. As part of the preparations for the WKPELA 2016 to investigate this hypothesis, a study was initiated on growth and condition of NSSH over the whole distribution area (Homrum *et al.*, 2016). An extensive analysis of biological data from 626 749 individuals of Norwegian spring-spawning herring collected internationally over the full distribution area and throughout the year (from surveys and samples from commercial landings) demonstrated spatial, seasonal and interannual variations in growth and condition during 1994–2015.

The growth, or length-at-age, varied over the period and was negatively related to stock size. The body condition at the end of the feeding season and weight gain over summer increased from record low levels in 1997 towards stable levels above the 20-year mean in the period 2005–2015. This increase in condition happened concurrently with a decrease in somatic growth in terms of length-at-age and an increased investment into gonad weights. These changes in allocation of energy may be linked to both changes in stock size, zooplankton abundance and temperature, but more analyses are needed before one may conclude with regard to main influencing factors.

There were no indications in this extensive data material of a dramatic event with regard to condition of the herring in the period 2009 onwards, whatever month or area analysed, that could have caused increased natural mortality. Around 1997 both length-at-age and condition of herring decreased substantially. This drop in condition was evident both in the east and west in May, and in all autumn months August-November in the east, where the condition in the stock was severely reduced resulting in among other things high levels of atresia in females as previously reported (Óskarsson *et al.*, 2002).

Another interesting aspect of the data are that a fishery on NSS herring occurred west of 2°W as late as September-November since 2004, and has been common ever since. The fish sampled in west in October were in better condition than in east in most years since 2004. This may indicate a prolonging of the feeding season in the west, as the herring normally migrates to wintering areas after feeding has ceased. The potential longer feeding season needs to be investigated with regard to food availability and stomach content analyses in the western areas in autumn.

At the meeting a second presentation was on the same topic. To clarify discrepancies between the two presentations extra work was done during the WKPELA meeting. The main conclusions from this work were included as an appendix to the original working document (Homrum *et al.*, 2016). The extra work done during the WKPELA meeting did not change the conclusion of the original working document (Homrum *et al.*, 2016), that there was no dramatic change in condition that would severely increase natural mortality of NSS herring after 2009.

2.4 Proportion mature

In 2010 the method for estimating maturity-at-age in the stock assessment of NSSH was changed based on work done by the “workshop on estimation of maturity ogive in Norwegian spring-spawning herring” (WKHERMAT; ICES, 2010a). The method which was adopted by WGWIDE in 2010 (ICES, 2010b) is based on work by Engelhard *et al*. (2003) and Engelhard and Heino (2004). They developed a method to back-calculate age at maturity for individual herring based on scale measurements, and used this to construct maturity ogives for the year classes 1930–1992.

The NSSH has irregular recruitment pattern with a few large year classes dominating in the stock when it is on a high level. Most of the year classes are, however, relatively small and referred to as “normal” year classes. The back calculation dataset indicates that maturation of the large year classes is slower than for “normal” year classes.

WKHERMAT and WGWIDE considered the dataset derived by back calculation as a suitable potential candidate for use in the assessment because it is conceived in a consistent way over the whole period and can meet standards required in a quality controlled process. However, the back calculation estimates cannot be used for recent years since all year classes have to be fully matured before included. Therefore assumptions have to be made for recent year classes. For recent year classes, WGWIDE (2010) decided to use average back-calculated maturity for "normal" and big year classes, respectively and thereby reducing maturity-at-age for ages 4, 5 and 6 when strong year classes enters the spawning stock.

WGWIDE also decided that average values should be replaced by back-calculated values when data were available. However, it was not clearly stated if this should be done by including one more year of data in annual assessment or in connection with a benchmark. In the years since 2010, average values have been used.

In WKPELA updated maturity values were available for the years 2006–2011, and Figure 2.4.1. shows assumed values and updated values for these years.

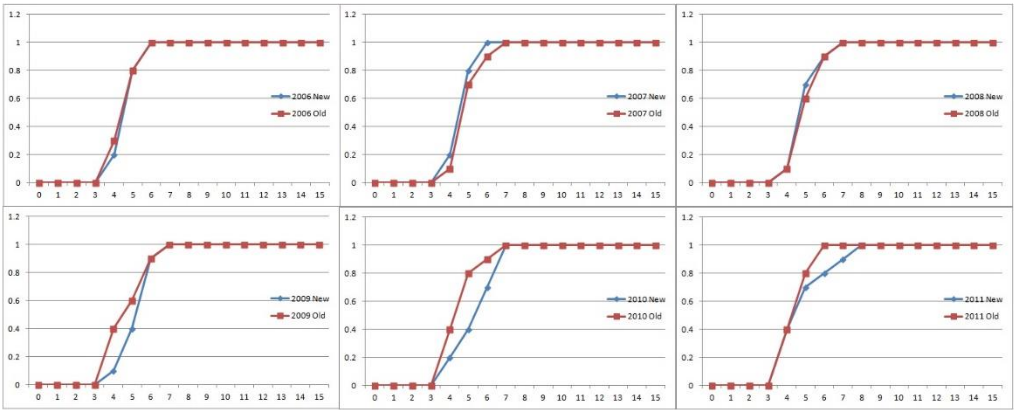

Figure 2.4.1. Old (assumed) maturity-at-age values used in the WGWIDE 2015 assessment (red curves) and updated (back-calculated) values (blue curves) for the years 2006–2011.

This change in maturity values in the years 2006-–2011 alter the size of the SSB in the corresponding years (Figure 2.4.2). Assumed values should be replaced by back-calculated values in the annual assessments for each year where updated values are available.

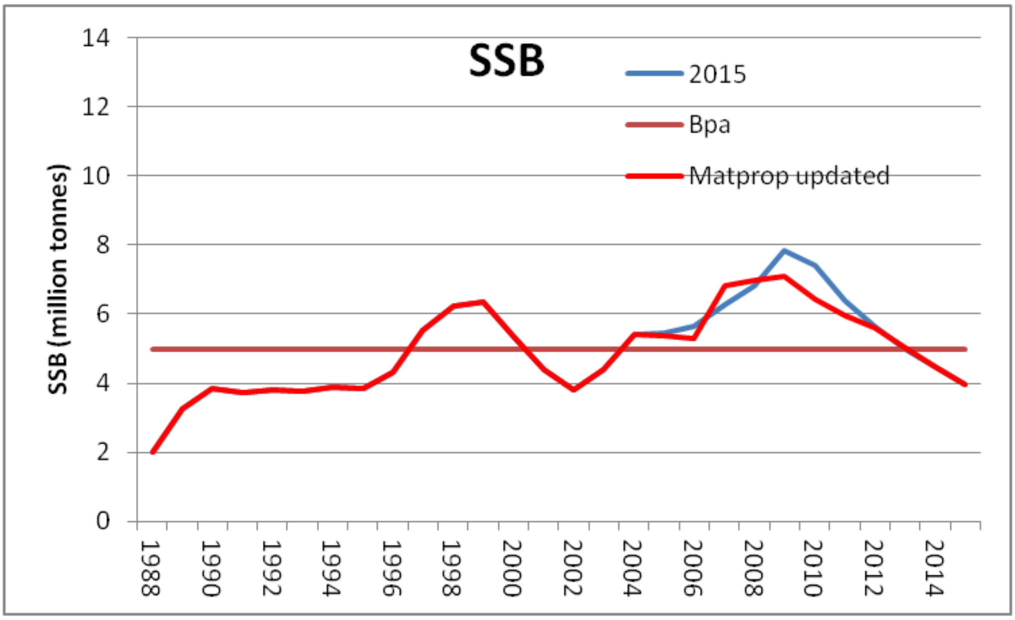

Figure 2.4.2. NSSH. SSB as estimated in WGWIDE 2015 (blue line) and as estimated with the updated maturity values (red line).

2.5 Discards

Ipsos Public Affairs, in cooperation with IMR and the fishing industry, conducted a survey to provide information on fishery induced mortality in the Norwegian fishery for NSSH. The survey was conducted in January/February 2016. The survey was done by phoning skippers and interviewing them. A total of 146 herring skippers participated in the survey, 31 skippers representing the bigger vessel group (ringnot) and 115 skippers representing the smaller vessel group (kystnot).

The data provided an indication that there have been periods of increased occurrence of net bursting. This is seen especially in the period 2007–2010. It does not, however, appear to be a trend in the size of catches where bursting has occurred.

When it comes to slipping, the data show a steady increase in the percentage that has slipped herring from 2004–2012, and then a significant decline in recent years. The variations in the proportion that have dropped herring largely are driven by the skippers on smaller coastal purse-seiners (kystnot). Average size of purse-seine hauls slipped seems to be relatively steady over the period. However, we find that the average size of net hauls slipped is lowest in the recent period.

There was not enough time during the WKPELA meeting to try to estimate the level of slipping/bursting (in tonnes) based on these data, but an attempt will be made before WGWIDE 2016.

3 Assessment

3.1 XSAM

A model template based on a state space model and structural time-series models for fish stock assessment have been developed and are described in Annex 3 and 4 (WD4 and WD12). A structural time-series model is a model which is set up in terms of components which have a direct interpretation. A state space model generally includes a model for the hidden states, a process model, and observation models which link noisy observational data to the process. This framework includes established and documented models. The main new achievement in the framework evaluated for NSS herring compared to other frequently used frameworks is to utilize prior knowledge of sampling errors to a greater extent to improve inference than what has been possible earlier. The model framework has been given the name XSAM to reflect that this is another version (X) of a statistical assessment model and is implemented in R using the R-library Template Model Builder (TMB, Kristensen 2014) and the model is fitted to input data using Maximum Likelihood.

The dynamical (process) model is built on the time-series model described in Gudmundsson (1994) who modelled fishing mortality as a structural time-series model to allow selectivity and effort in a separable model to change over time to allow necessary flexibility in the process. In Gudmundsson's original formulation the time-series models were represented by random walks, whereas in this framework the formulation is extended to allow for replacing random walks with autoregressive models. The framework allows for turning on and off components allowing simplifications such that the model for fishing mortality can be modelled as e.g. a separable model (e.g. Quinn and Deriso 1999), separable model with noise, a separable model with noise where the effort is modelled as a time-series model (e.g. Aanes *et al*., 2007), and may be configured such that the model for fishing mortality is similar to the SAM model as described in Nielsen and Berg (2014). Definitions of variance components in the processes are generic and allow for deviation of the typical independent and identically distributed (iid) assumption such for multivariate processes they are allowed to vary by e.g. age can be correlated, although structures in variances of latent processes may be difficult to grasp and generally requires prior specification.

The observation models are also based on well-established theory and functional relationships linking observations to the processes, but include a flexible formulation of observation error to allow for realistic error structures as well as allowing utilizing prior information about sampling errors. Usually within this class of models, observation models includes simplified parameterizations of the observation errors, and includes parameters that is estimated by fitting the model to point estimates of observational data (e.g. Aanes *et al*., 2007, Nilsen and Berg 2014), usually assuming that errors are independent and identically distributed (iid) for each input dataset. Point estimates of input datasets alone bares little information about the error structures in the estimates, and consequently it should not be expected that information about complex structures of observation errors is inherent in the point estimates and can be estimated by the model. Observation errors generally depend on sampling design, sampling intensity as well as mean values of estimates (Stenevik *et al*., 2015, Aanes and Vølstad 2015) and exhibits considerable variability over time and age which implies that observational data should be weighted accordingly (c.f. Gavaris 1988, Quinn and Deriso 1999). Reliable estimates of sampling errors have generally been lacking due to ad hoc performed monitoring programs and survey design and methods for analysis but have recently gained attraction (c.f. ICES 2011, 2012, 2014, 2015, 2016b, and references therein). Rigorous analysis of survey sample data for both commercial catches and abundance indices is permitd by recent developments in methods and software which accounts for the hierarchical cluster sampling nature of survey sample data as well as potential stratification (see Hirst *et al.*, 2012 for estimation of catch-at-age and Stenevik *et al.*, 2015, ICES 2015 for abundance indices). For the herring data it is found that strong year classes which appear both in abundance indices and catches, are estimated more precisely than weak year classes provided that the sampling intensity is sufficient (Aanes 2016a). Furthermore, it is found that, in combination with low sample sizes, young and small fish that not is fully recruited into the fishery, as well as less abundant old ages are less precisely estimated compared to fish that is fully recruited into the fishery and appear abundant in catch and surveys. Note that utilizing sampling errors implies objective weighting of individual datum in input datasets according to established theory as precise estimates (associated with high abundance, high sampling intensity and well defined sampling design) attain higher weights and increases precision in stock parameters, than variable estimates (associated with low abundance and low sampling effort) which attain lower weights and lowers precision of stock parameters.

The use of a statistical model represents a shift for NSS herring which moves from a VPA-based model to a probabilistic model although statistical models are widely used for other stocks. The main difference between this framework and the SAM framework is that this framework generally not includes the process error term in the cohort equation in addition to potential use of prior information on the observation variances. The configuration of the model and adequacy of model fits is aided by well-known diagnostics such as qq-plots, residual plots, AIC (Annex 3 and 4) and likelihood profiles and is found adequate by the working group.

Utilization of sampling errors represents an aspect which is new for NSS herring and WGWIDE and WKPELA therefore spent much time on reviewing diagnostics and configurations of observation errors from model fits as it effectively opens for various ways of weighting data. This aspect is therefore more closely summarized in the following:

Configuration of observation error and utilization of sampling error is aided by decomposing the observation error into a scaling constant, variances at age and time and correlation at age and time, and certain elements of the structure may be fixed and others estimated when fitting the model to data. Rationale and details for the decomposition of the covariance matrices for observation errors is given in Annex 3 and 4 and summarized as follows. For input dataset O the observation error $\boldsymbol{\Sigma}'^{\boldsymbol{O}}$ is proportional to the sampling variance $\boldsymbol{\Sigma}^{\boldsymbol{O}}$ such that $\boldsymbol{\Sigma}'^{\boldsymbol{O}} = \mathrm{h}_O \boldsymbol{\Sigma}^{\boldsymbol{O}}$ with proportionality constant (scaling factor) h_O. For the herring data, $\boldsymbol{\Sigma}^{\boldsymbol{O}}$ is replaced by an estimate each year. Since the estimates of the correlation structures are variable, we consider two versions of $\boldsymbol{\Sigma}^{\boldsymbol{O}}$, one which includes the covariances, and one which omits the covariances (i.e. assume that the estimates are uncorrelated). This leads to several options for exact configuration which was discussed during WKPELA: At one extreme the errors may be assumed to be iid and unknown, and the parameters (and thus the weights) are estimated by fitting the model to data. This approach increase the risk of introducing bias in both point estimates as well as estimates of variance (see Aanes 2016a) as it does not utilize excessive prior knowledge of the errors in data since it is assumed that data obeys the iid structure. At the other extreme, all parameters in the observation errors can be fixed according to the prior knowledge of sampling variances. This choice faces the risk of underestimating the uncertainty in stock parameters SSB and F since in reality it is reason to believe that observation models may include other sources of variability than estimated from survey sample data. For example, variability of total reported landings or variability of catchability. A compromise between the two is to use an intermediate form between complete ignorance to complete knowledge and estimate certain components of the error structures. The different formulations of the observation error discussed during WKPELA are summarized in Table 3.1.2 and were extensively evaluated for the herring data.

Table 3.1.1 Different configurations of the observation model considered by WKPELA

OBSERVATION MODEL	VARIANCE	CORRELATION	SCALING FACTOR	COMMENT
0	Estimated by fitting the model to data assuming iid error for each input dataset	No	$h = 1$ for all datasets	Common assumption. Implies equal weight on all observations within a dataset across ages and years, but different across datasets
1	Use empirical	No	Estimate h independetly for each dataset	Estimate relative weights between datasets, but keep weights within datasets
2	Use empirical	Yes	Estimate h independently for each dataset	Estimate relative weights between datasets, but keep weights within datasets
3	Use empirical	No	$h = 1$ for all datasets	Error structures and thus weights are completely known
4	Use empirical	Yes	$h = 1$ for all datasets	Error structures and thus weights are completely known
5	Use empirical	No	Estimate the same h across datasets	Use relative weights between and within datasets according to sample data, but allow adjustment of total uncertainty
6	Use empirical	Yes	Estimate the same h across datasets	Use relative weights between and within datasets according to sample data, but allow adjustment of total uncertainty

Based on a careful validation of assumptions and diagnostics of model fits WKPELA agreed that the summarized diagnostics favoured observation model 1 or 5 for the herring data. Comparing qq-plots and residual plots no difference in performance between the two could be detected, but the likelihood profiles provided some support for favouring observation model 5 over 1 based on the complete time-series of data (i.e. 1988-2015) including catch-at-age, Fleet 5 and Fleet 1. More specifically this means that the sampling error component of the observation error is parameterized using empirical estimates of sampling variances for catch-at-age, Fleet 5 and Fleet 1. Comprehensive analysis of the sample data showed that the estimates of the standard errors by age and year were robust and can be determined by the sampling design, sampling effort and mean values of estimates. There is empirical evidence of a correlation structure that exhibits positive correlations among neighbouring ages within a year, but the estimates are imprecise and not stable. This suggests that standard errors could be safely utilized whereas correlation structures should be handled with care. Therefore it was decided that the estimates of the correlations and the effects on the estimates are poorly understood and they are therefore not used at this stage. Furthermore, setting the scaling factor of the sampling variances equal across input datasets implies that each datum is given weight within and across dataset according to the sampling variances, but the scaling factor is estimated to acknowledge that all uncertainty in the observations may not be captured by the sampling variances, and thus reduce the risk of underestimating the uncertainty in SSB and F. In addition to the diagnostics, another reason for not allowing the model to estimate the relative weights across datasets is that there is some evidence that these types of models not are able to appropriately estimate the relative weights, which in this situation results in the model down weighting the surveys too much compared to catch-at-age (see Annex 4 for details).

In parallel with deciding on type of observation model, the specific configuration of the other parts of the model was also decided. It turned out that these choices were not sensitive to the choice of observation model, and were aided by the AIC criterion and examination of qq plots and residuals. From these analyses we found that there is some evidence of a time varying selectivity (significant improvement by AIC and reduction of cohort effects in residuals) although it is necessary to constrain the parameter space for the variance components for selectivity and fishing mortality to reasonable values to obtain convergence. We also find that replacing the random walks in the models for time varying selectivity and effort with AR(1) model significantly improves the fits of the model.

The final configuration of XSAM is

- Time span: 1988–last data year
- Age span: 3-12+
- There is no empirical evidence of $\sigma_3^2 > 0$, therefore σ_3^2 is set to 0 and effectively one level in the hierarchy for the latent state of effort is omitted.
- Effort is modelled as an AR(1) process
- Selectivity is modelled as a multivariate AR(1) process
- $a_m = 11$, i.e. the selectivity in fishing mortality is assumed constant for ages 11 and above
- Observation model 5 which implies estimating a common scaling factor h across datasets

- Point estimates of catch-at-age reported as usual with associated estimates of sampling variances (see Annex 3)
- Fleet 5: StoX estimates, q-plateau at ages above 11
- Fleet 1: StoX estimates, q-plateau at ages above 8

An extract of the diagnostics for the model fit for the final configuration is shown in Figures 3.1.1–3.1.8 and further details are found in Annex 3 and 4. The resulting estimates of spawning-stock biomass and average and weighted average of fishing mortalities (ages 5–10) are shown in Figure 3.1.9 and compared to the estimates of WGWIDE in 2015. Figure 3.1.10 shows the estimated spawning-stock biomass by XSAM and TASACS based on the revised estimates of proportion mature at age (see section 2.4) using the same input data otherwise (i.e. StoX estimates of Fleet 5 and Fleet 1).

The source code for XSAM is currently only available on request (Sondre Aanes) , but the code including documentation and example of use will be made available for WGWIDE at the working group's SharePoint site in due time before the next working group meeting.

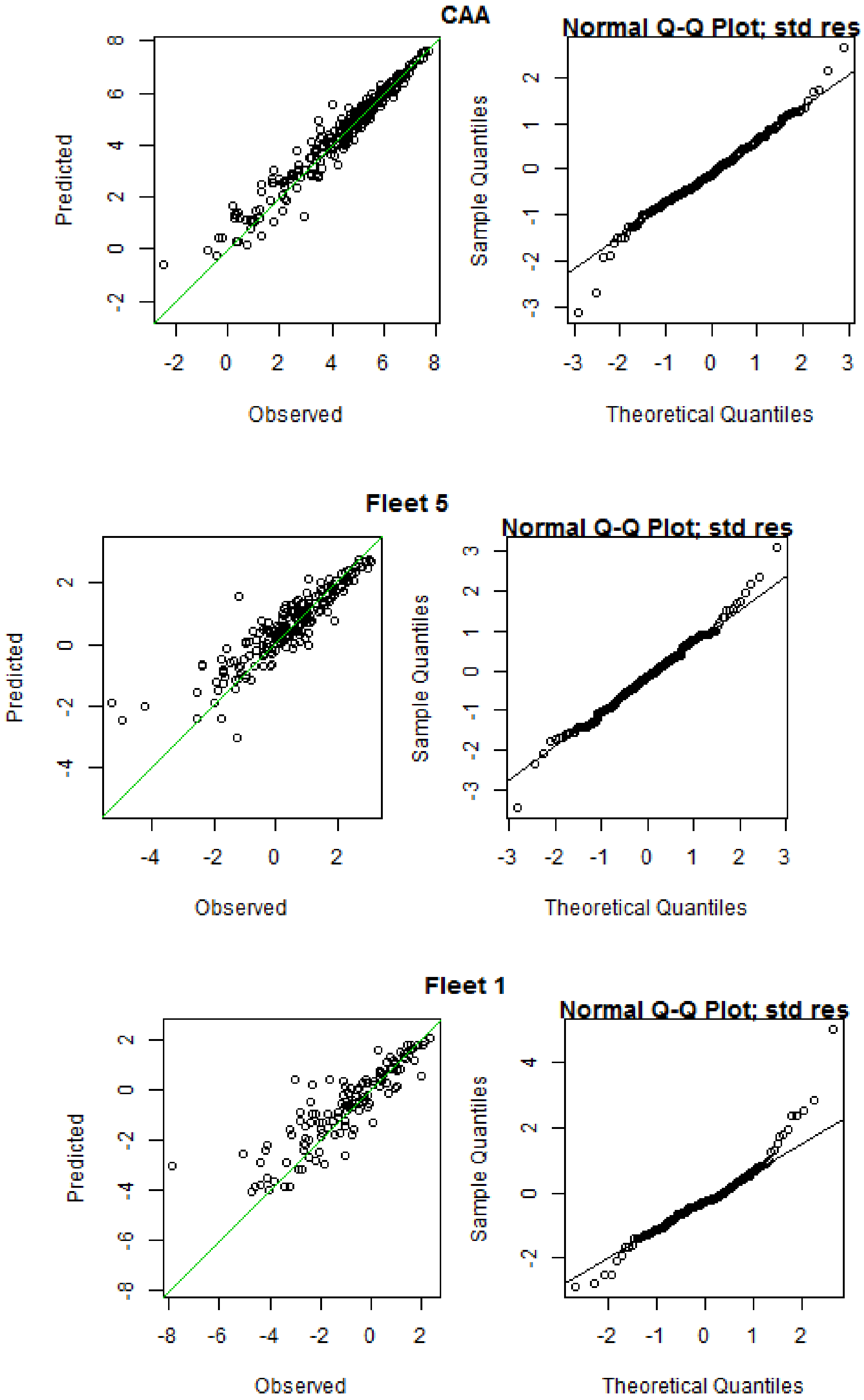

Figure 3.1.1. Observation model 5 including CAA, Fleet 5 and 1. Observed vs. predicted catch-at-age and qq plot of the residuals for catch-at-age.

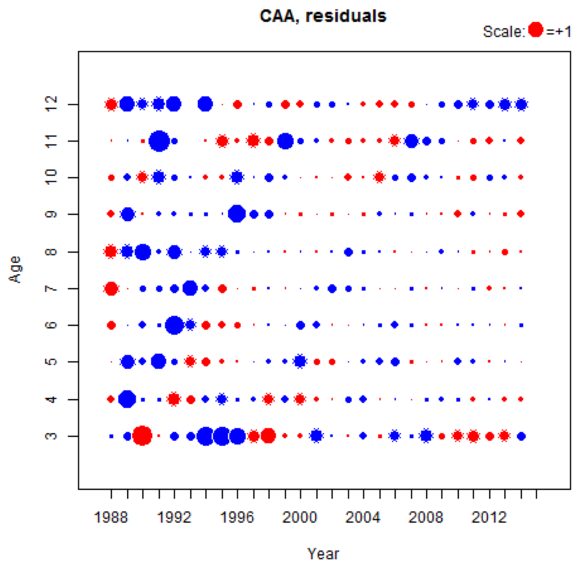

Figure 3.1.2. Observation model 5 including CAA, Fleet 5 and 1. Residuals of catch-at-age.

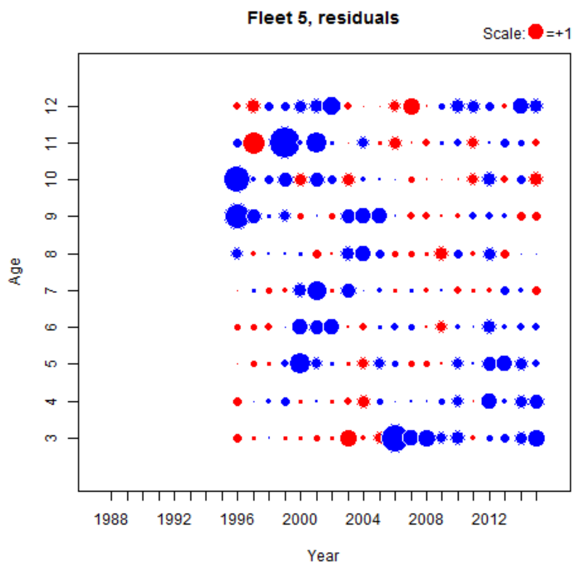

Figure 3.1.3. Observation model 5 including CAA, Fleet 5 and 1. Residuals of Fleet 5.

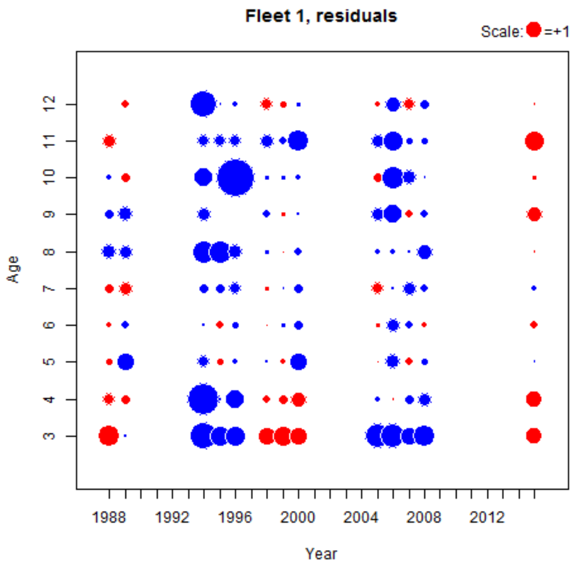

Figure 3.1.4. Observation model 5 including CAA, Fleet 5 and 1. Residuals of Fleet 1.

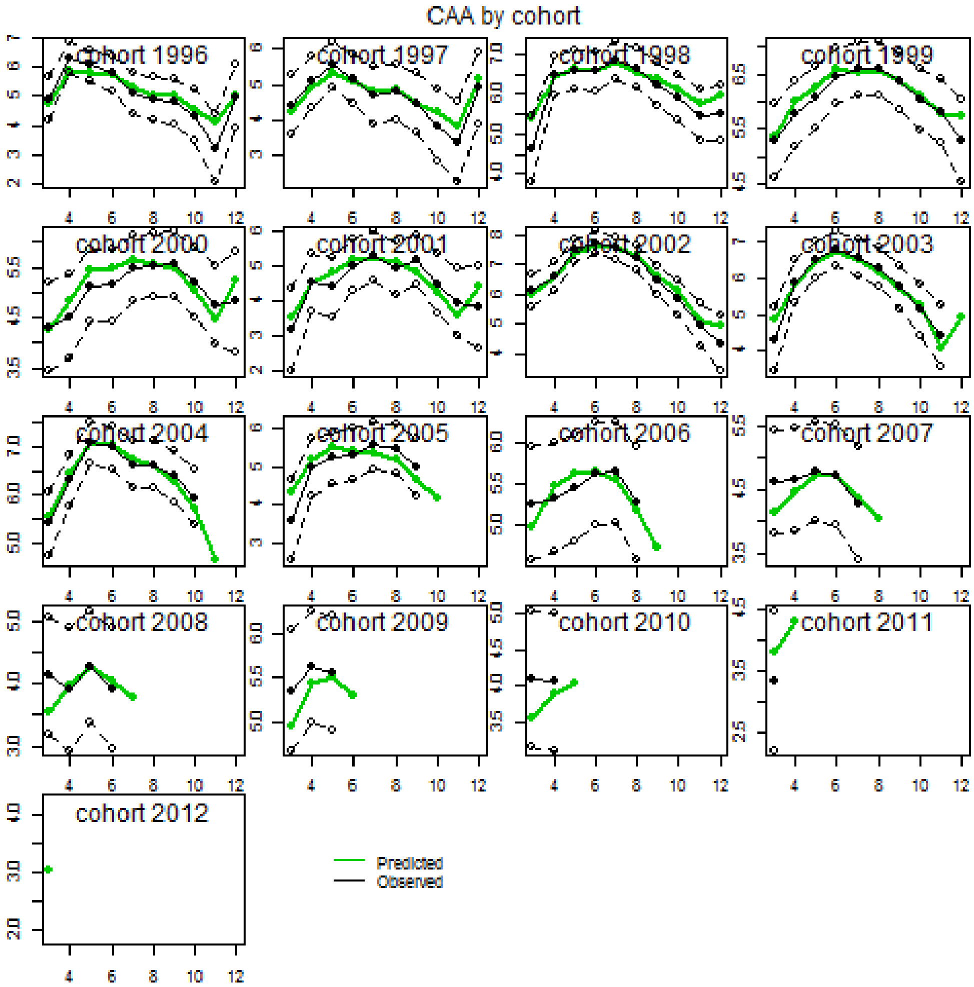

Figure 3.1.5. Observation model 5 including CAA, Fleet 5 and 1. Observed vs. predicted catch-at-age by cohorts 1996–2012. The broken lines with the open circles indicates approximate 95% confidence intervals from empirical analysis of sample data.

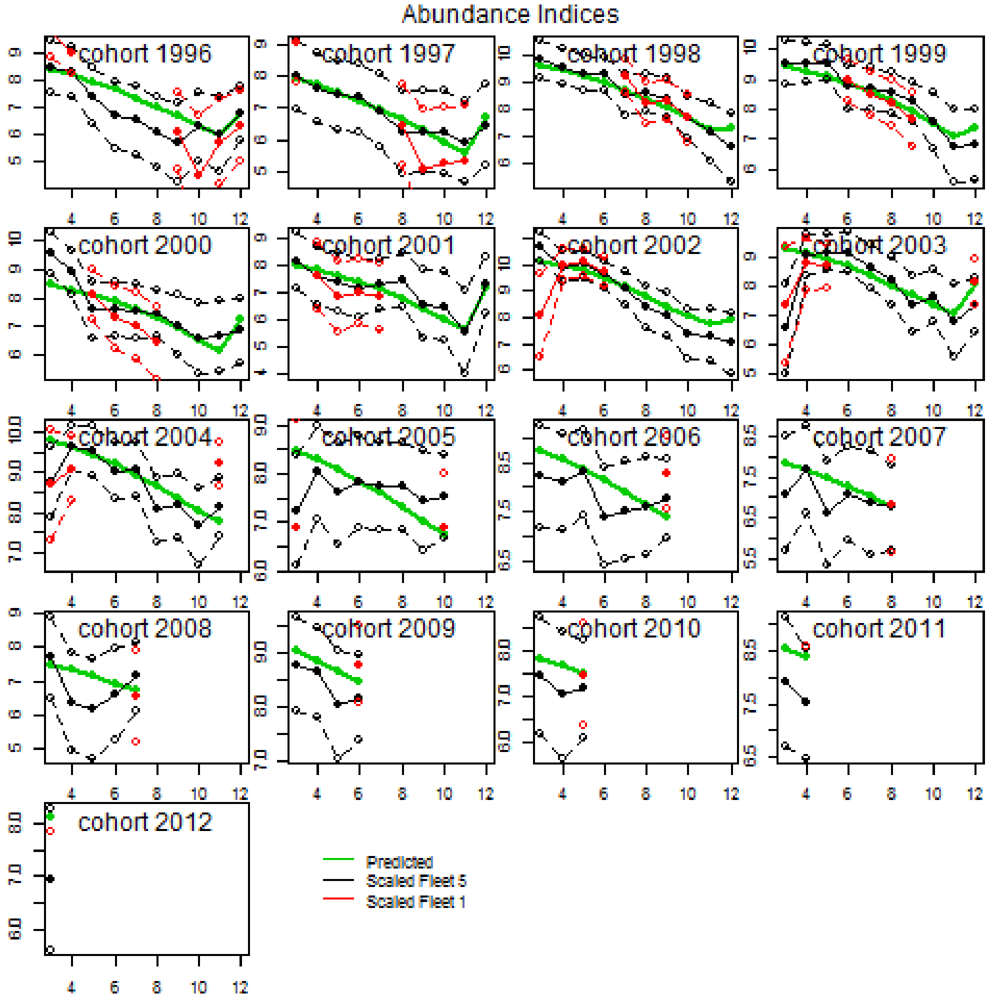

Figure 3.1.6. Observation model 5 including CAA, Fleet 5 and 1. Observed vs. predicted abundance index at age by cohorts 1996–2012 for Fleet 5 (black) and 1 (red). The abundance indices are scaled to the population values using the estimated catchabilities. The broken lines with the open circles indicates approximate 95% confidence intervals from empirical analysis of sample data.

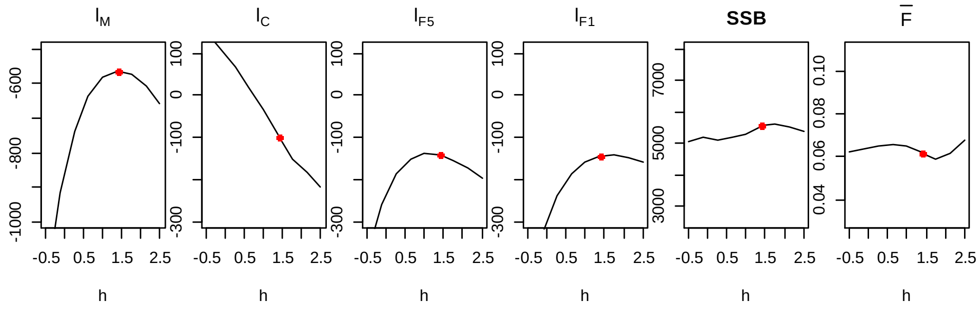

Figure 3.1.7. Observation model 5 including CAA, Fleet 5 and 1. Profiles of marginal log-likelihood l_M, the catch component l_C, Fleet 5 component l_{F5}, Fleet 1 component l_{F1}, point estimate of SSB and average F (ages 5-12+) in 2015 over the common scaling factor for variance in data h. The red dots indicate the value of the respective scaling factors for which the log-likelihood is maximized.

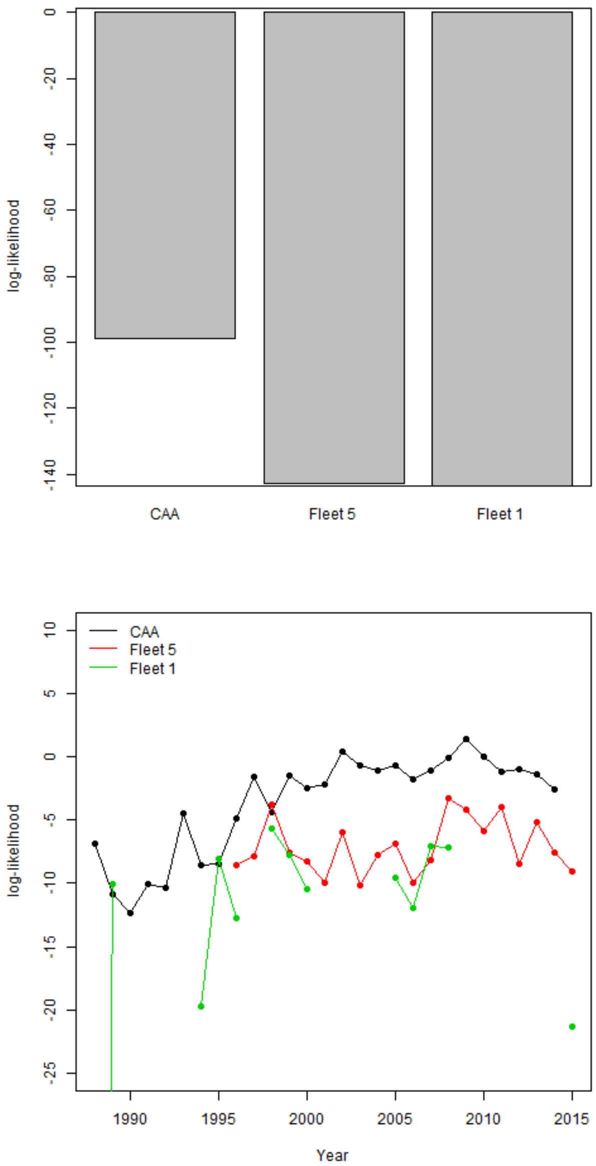

Figure 3.1.8. Log-likelihood values for the data components from the model fit using observation model 1 including Fleet 5 and 1. Total log-likelihood values are in the left panel by year in the right panel.

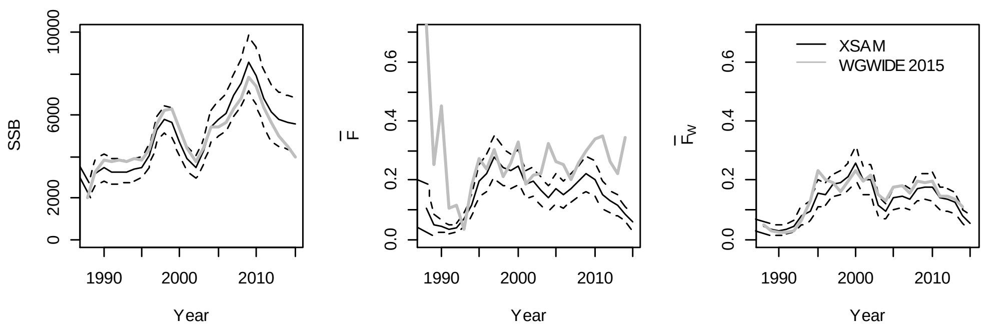

Figure 3.1.9. Estimates of SSB, average F ages 5-10 and weighted average of F ages 5-10 by XSAM using observation model 5 and input data CAA, Fleet 5 and 1. Broken lines show approximate 95% confidence bounds. The corresponding estimates from WGWIDE in 2015 are shown in grey.

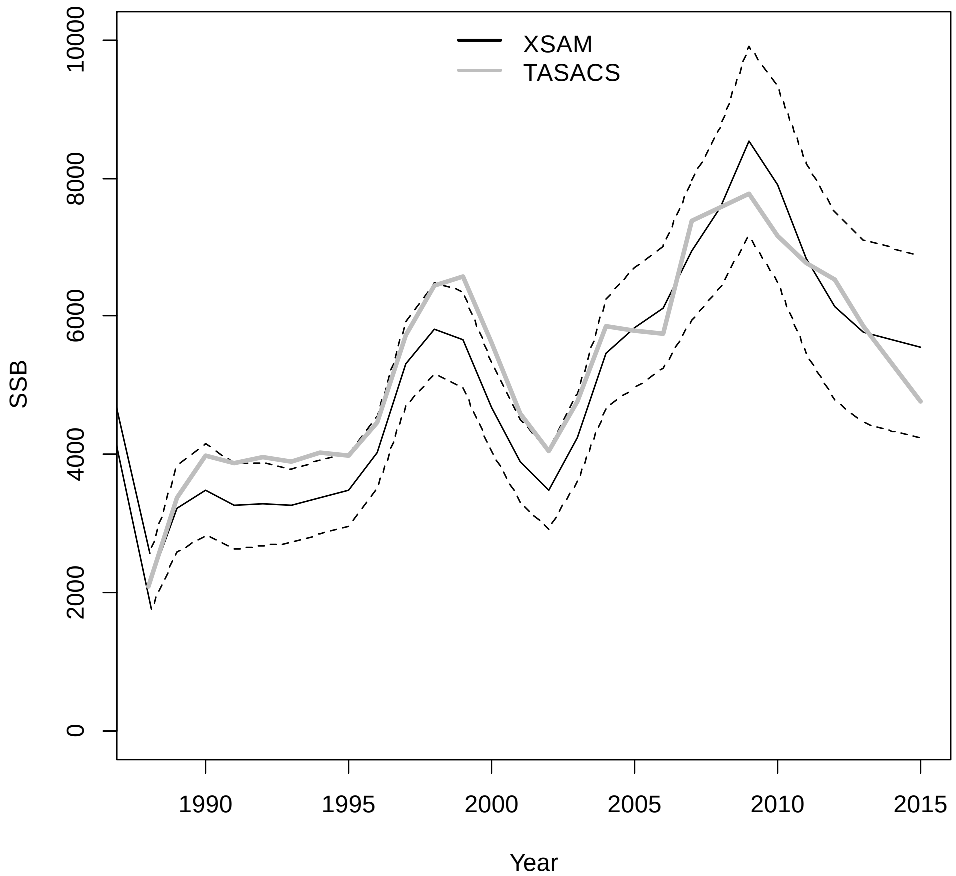

Figure 3.1.10. Estimates of SSB by XSAM using observation model 5 (black) and TASACS (grey) using revised estimates of proportion mature at age and the same input data CAA, Fleet 5 and 1. Broken lines show approximate 95% confidence bounds.

3.2 TASACS

3.2.1 TASACS software and configuration

Since the benchmark in 2008 (ICES 2008), the standard assessment tool has been the VPA module in TASACS. TASACS (Skagen and Skålevik, 2009) is a toolbox with several assessment tools (a VPA, a separable model and ISVPA), with a range of options for objective functions and for handling of the plus group. It also has interface for data handling, archiving and presentation of results. At the 2008 benchmark it was decided to use the VPA option, which performed slightly better with respect to retrospective errors than other options at that time. In line with ICES practice, TASACS has since then been run as decided in 2008, with a few amendments. The mechanizm for handling zero catches was improved in 2013. The new algorithm for deriving the terminal stock numbers, for the year classes where no data were available to estimate the terminal stock numbers, assumes a fixed ratio between F at oldest age and average F in the year, which is equivalent to assuming a fixed selection at oldest age. Similar method is used in the assessment model ICA, and in the separable option in TASACS. The ratio is taken from the selection parameters, as the selection at oldest age relative to the mean over the ages 5–13. There is no standard way to estimate that ratio. However, a sensitivity analysis showed that the exact ratio used has only a minor influence on the estimated numbers in the earlier period and none on the latest part of the time-series. Values between 1.1–1.7 give comparable results. The ratio between the terminal F and the average F over ages 5–13 calculated for all the years where terminal F is estimated is 1.3 (excluding all F = 0), and this was applied in the 2013 assessment.

A bug in the calculation of retrospective error (handling of year classes that are not tuned) was also fixed in 2015. Other additions: a prediction module (improved in 2016) and iterative reweighting of terms in the objective function, have been made but have not been used.

The view of WKPELA is that TASACS should still be kept as a fall-back if XSAM should fail. XSAM uses a subset of the data and another age range. Basically, TASACS would be run as in previous years, but its performance when adopting the data selected for XSAM was briefly examined. Some studies of the performance of TASACS, in particular the retrospective problem were presented to WKPELA and are also briefly discussed here.

The VPA module uses the catch numbers-at-age to reconstruct each year class backwards in time starting with the survivors at the end of the last year with catch data or the oldest true age. The survivor numbers are estimated by fitting the stock numbers-at-age to survey indices at age, calibrated with catchabilities. Below is a more detailed overview of essential features.

TASACS model parameters that are used in the present configuration:

- Stock numbers. N(a,Y) and N(A-1,y): Survivor numbers at the end of each year class cohort.
- Catchabilities at age by fleet: q(a,y,fleet), that model survey indices I as I=q*N, i.e. assuming a linear relation between survey index and stock number.
- Natural mortalities M(a,y)

Options for parameters: Each individual parameter has a 'flag', decided by the user, telling how to handle it:

- Keep it fixed
- Estimate it in the optimization process ('active' parameter)
- Use the same value as previous year
- Use the same value as previous age
- For numbers at oldest age: Use catch numbers and F from younger ages.

The oldest age A is regarded as a plus group and modelled as a dynamic pool with mortality equal to that of the oldest true age: N(A,y+1) = N(A,y)*exp(-M(A,y)-F(A-1,y)) + N(A-1,y)*exp(-M(A-1,y)-F(A-1,y)). It could have been included in the likelihood function, but that is not done here.

TASACS Likelihood components, as decided in 2008.	Observation	Model	Likelihood
Catch numbers-at-age	Cobs(a,y)	Used to back-calculate year classes, assumed error-free.	
Survey indices	Iobs(a,y,fleet)	Imod(a,y,fleet) = N(a,y)*q(a,fleet)	◎a,y,fleet{logIobs(a,y,fleet)/Imod(a,y,fleet)}2
Biomass (SSB) indices	Iobs(y,fleet)	Imod(y,fleet) = SSBmod(y)*q(fleet)	◎a,y,fleet{logIobs(a,y,fleet)/Imod(a,y,fleet)}2
Total likelihood			Total likelihood is the sum of all components.

The 'likelihood function' is just a sum of squared log residuals. In the present configuration, they are not weighted according to their variance, and thus are not likelihoods in the true sense. Terms can be weighted manually, both individual terms and whole surveys.

The optimization, finding the minimum of the negative likelihood as a function of the 'active' parameters is done by a searching routine.

The standard procedure, agreed in 2008 has;

- Fixed natural mortality-at-age equal for all years (M=0.9 for ages 0–2, 0.15 for older ages).
- Equal weight to all individual terms in the likelihood function, except data that are excluded by giving them zero weight.
- Survey catchabilities constant over time but dependent on age except for the oldest ages (see below).

- For some small year classes, no attempt is made to estimate their magnitude. Their survivor numbers are just fixed at a small value and the year class is back-calculated with the catches.
- Some survey data are excluded by giving them zero weight. That is the case for some small year classes and for some years and ages where the survey index in question is regarded as not representative for the stock abundance. Criteria for excluding survey data were established in the 2008 benchmark and have been followed since.

In the 2008 benchmark, it was decided to start the year range in 1988. Data do exist further back in time. Previous Working Groups used 1950 as starting year, and stock estimates going back to 1907 were published by Toresen and Østvedt (2000). The motives for using 1988 were problems with the VPA for the period in the 1960ies and 1970ies when the fishery was closed and most catches-at-age were zero except for the very youngest ages. Moreover, it was considered that the period form 1988 onward would represent the present situation of prime interest to management, and that including earlier years should have little affect on the present estimates, and that if a longer time-series is needed, previous estimates of the earlier period should still be valid.

3.2.2 Effect of data revisions on TASACS performance

As part of the benchmark, two revisions were made to the data. The major surveys were revised using the StoX software, which provides variance estimates but also revised point estimates (Section 2.2.2). In addition, the maturity ogive was revised for the years after 2007–2011 (Section 2.4), in line with previous revisions. TASACS was run with the revised data, using the same conditioning as the standard assessment by WGWIDE, to compare with the results using the previous data.

It was realized that two observations in Survey 4 (Age 1 in 2013 and age 2 in 2012) created very large residuals. In both cases, the indices were far smaller than one would expect. It turned out that both these indices were based on very few fish caught (3 and 1 respectively). Therefore, it was decided to ignore them.

Figure 3.2.1 shows the SSB estimates using the revised input data with the old and new maturity ogives, and the SSB estimate by WGWIDE in 2015 for comparison. In addition, the figure shows a run where the Larval survey (NHLS), which is used as an SSB indicator, is ignored. The signal in this survey is in conflict with the rest of the assessment, as indicated by a strong trend in the residuals and an almost flat q-q plot. The revision of survey data leads to somewhat higher SSB estimates in the years after 2004. The revised maturity data have some effect for 2006–12, altering the shape of the peak in 2009. Most of this effect is due directly to altered maturity, the difference caused by the fit of revised SSBs to the larval survey SSB index is <0.5% in all years (not shown).

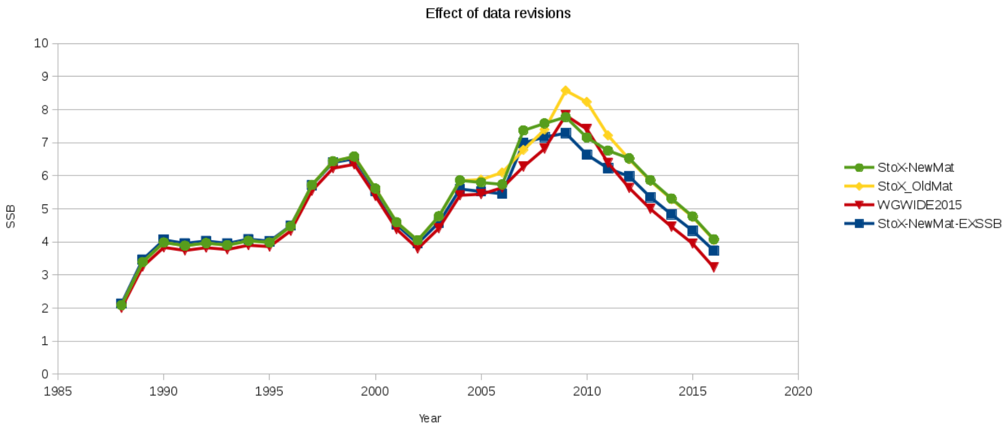

Figure 3.2.1. Time course of SSB with various revisions of data. See text for explanation.

3.2.3 Retrospective errors with TASACS

Retrospective errors were a problem with the TASACS assessment, in particular in the period around 2010–14. In the last few years, the assessments have been quite consistent. Some of this problem could be amended by a revision of the handling of small year classes, made in 2013 A working paper was presented which attempted to trace the cause of the retrospective problem (Skagen, WD02 It was found that the retrospective problem could be removed by excluding survey data for survey 5 for the ages 13–15, in particular for the year classes 1997 and 1999 (1998 might also come in the same category, but the data for old age were already excluded for that year class). Apparently, these year classes were reduced very rapidly in the survey data, leading to a downward revision of the estimate of the history of the year class each year. The problem was also visible as a cluster of positive residuals for Survey 5 in 2006–2009, and a correspondingly high ratio between survey index and estimated stock numbers. Some of these deviations, in particular the larger ones, were caused by data for ages and years that are not used in the assessment. The cause of the steep reduction in the survey indices, and whether the recent low indices or the previous high indices are the most correct ones, could not be decided, but it was noted that catchability estimates were far from stable, even for fleet 5.

3.2.4 Adapting to the conditioning of XSAM

The standard TASACS assessment used ages 0–15+ and years 1988–2015. Altogether 7 acoustic surveys are used, although some years and/or year classes or individual data are excluded from some surveys. In addition, a larval survey is included as an indicator of SSB. The survey 5 is regarded as the backbone of the present assessment. The effect of excluding surveys was examined in the November 2015 data preparation workshop, and it was demonstrated the stock estimates for the years after 2000 were de-stabilized without that survey, leaving the juvenile surveys as the only calibration of the recent year classes.

XSAM is currently set up for ages 3–12 (which may be changed to 0–12) and uses only the surveys 1 and 5. It was briefly examined how TASACS could work with only those data. The investigation concentrates on SSB, which can be regarded as a key measure of abundance. Fishing mortality was related to SSB as expected throughout and is not shown.

All runs were made with the revised data as recommended by WKPELA (see section 2.2.2 and 2.4) as discussed in Section 3.2.2.

Runs were made with:

- 0-12+ with all surveys
- 0-12+ with surveys 1 and 5

These were compared with

- 0-15 with all surveys
- 0-15 with surveys 1 and 5

In the 0–12 cases, two alternatives were examined for the ages above which the catchability was assumed to be flat:

- The age used as a standard in the current assessments,
- The ages suggested for XSAM.

The 0–15 runs were made with catchabilities conditioned as in the standard assessment.

The ages are shown in the text table below:

	Fleet 1	Fleet 2	Fleet 3	Fleet 5
TASACS standard	10 (max age)	10	10	11
Suggested for XSAM	8	8	7	11

All these examples were without the larval survey. If the larval survey is included, the SSB estimates in the distant past is slightly lower and those in the most recent years slightly higher (see Figure 3.2.1). This is in line with the signal in that dataseries and in the residuals, which is an increasing trend that is not seen in other data.

Lifting the youngest age from 0–3 years makes surveys 4, 6 and 7 redundant. Excluding these surveys leaves no information about ages 0–3, so the stock number-at-age 0 becomes just an expansion of the estimate at age 3. For year classes that have reached old age, the data at ages 0–2 have limited influence, although excluding them makes some difference to some year classes. The exception is year classes that still are less than 4 years old, for which there is no other information than the juvenile surveys. Without the juvenile surveys, the recruitment of these year classes is undetermined.

Excluding surveys 2 and 3, which were terminated long ago, was not expected to have much effect on the estimate of the present state of the stock, but should have some affect on the early period. This may become more important as some old year classes were covered by remaining surveys only at the oldest ages. In particular, the 1983 year class is first seen by survey 5 at age 13. It is still seen at ages 5 and 6 in survey 1, though.

The results show that the SSB estimate for the years 1988–1995 indeed was substantially lower when the plus age was reduced and this was the case whether the surveys 2 and 3 were included or not. The dominating year class in that period is the 1983 year class, and its estimate is far lower. A detailed examination of the N-values of the 1983 year class corresponding to the various sources of information revealed that this is both because some data (survey 5 in particular) are lacking, and because of different catchability estimates. In particular, the catchabilites for surveys 2 and 3 look unstable at old age (Figure 3.2.2).

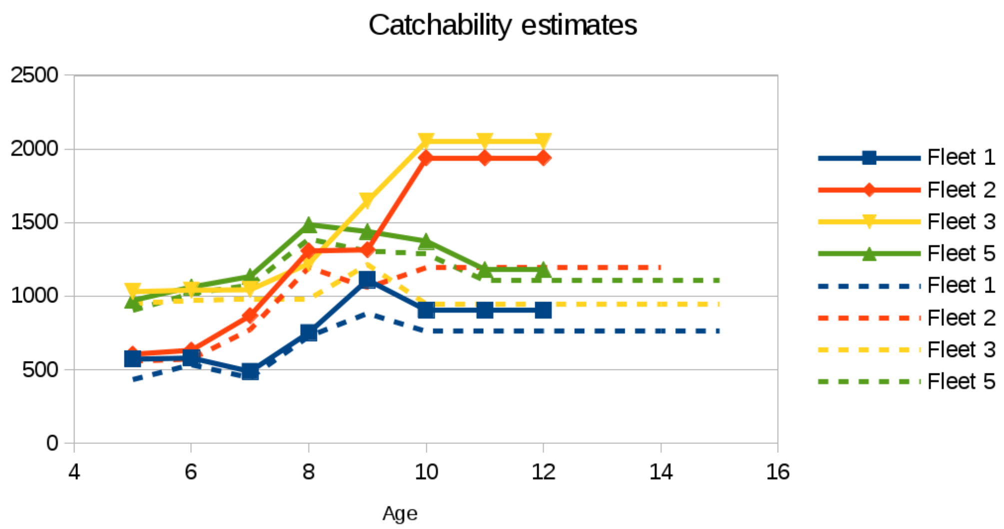

Figure 3.2.2. Estimates of catchabilities at age for the survey fleets covering adult herring. Whole lines is when using 12 as plus age, hatched lines are using 15 as plus age.

For the SSB estimates in recent years, it turned out that reducing the plus age from 12–15 years had some effect. However, the effect was quite sensitive to which data were included - using all survey data or only fleets 1 and 5 made some difference. Also, the age above which the catchability was assumed flat mattered, in particular when all surveys were included (Figure 3.2.3). Comparable effects were seen on recent Fs and recruitments, except for average Fs dominated by non-estimated year classes and recruitment of very recent year classes when survey information is skipped.

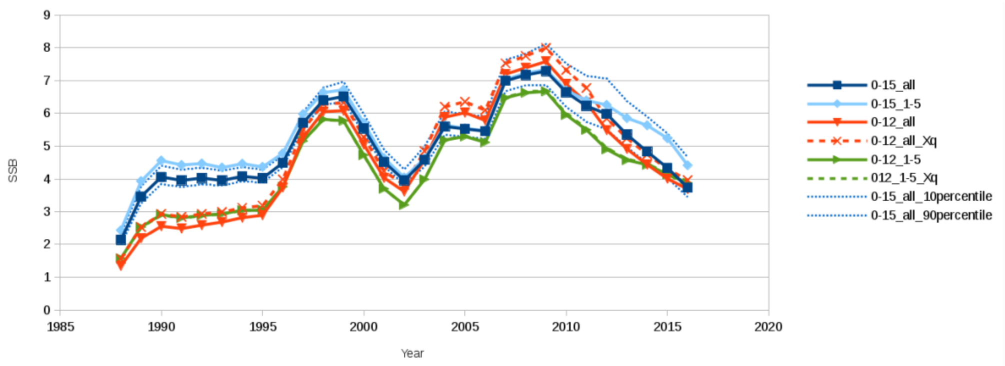

Figure 3.2.3. TASACS estimates of SSB for various uses of survey data. 0-12 and 0-15 are the age ranges used. All indicates using all surveys except the larval survey, 1-5 is using only surveys 1 and 5. Xq indicates assuming flat survey catchability above the same ages as used in XSAM. The 10 and 90 percentiles are from a bootstrap run of the 0-15_all option.

Accordingly, the results in the recent period are sensitive to both which surveys are included, which ages as used and the ages above which catchability is assumed to be flat. The effect of these options does not seem to be systematic. For example, using only surveys 1 and 5 rather than all surveys reduces the SSB estimate when taking 12 years as plus age, but increases it when the plus age is 15 years. Hence, the key issue may seem to be that the assessment is quite sensitive to noise in the data, rather than systematic effects of specific surveys or ages. The sensitivity of the results to noise in the surveys, was discussed further in Section 3.2.3 as a possible cause of retrospective problems.

In summary, using TASACS as before, but with data and assumptions in line with what is practised in XSAM, makes some difference, with results close to the 10 percentile of a bootstrap run with the previous use of the data. Reducing the plus age to 12 destabilizes the catchability estimates of fleets 2 and 3, leading to somewhat variable results depending on the assumed lowest age for a flat catchability. These surveys have some un-predictable affect on the estimates of the abundance even in the recent period. Likewise, the SSB index from the larval survey tended to lift the SSB estimate for the recent period slightly, linked to a marked trend in the residuals.

3.3 Statistical catch-at-age model

The model which is described in appendix 5 is all written in AD-model builder but divided in different submodels. First the Historical assessment model is run, estimating biological parameters and selection pattern of the fisheries with confidence intervals on parameters, stock size and fishing mortality calculated from the inverse Hessian matrix. The inverse Hessian matrix is then used as proposal distribution in MCMC simulations where the number of simulations are 2 million and the parameters from every 1000th run are saved to a file (done with the command **ashcatage -nox -mcmc 2000000 -mcscale -mcsave 1000**). The saved sets of parameters are then used in 2000 stochastic runs, in each run the assessment model is run, feeding directly into the prognosis, observation model and Harvest Control rule that in the program are just simple functions in the prognosis function. The stochastic simulations are done with the command ashcatage -mceval which reads the file ashcatage.psv storing the 2000 sets of parameter values stored in the mcmc run. The model is written in such a way that it must do prediction for at least 4 years, even in the estimation mode. In the stochastic simulation mode the number of years simulated is usually increased from around 5 to 50-100 but running the estimation with 50 years will increase the computation time as each mcmc evaluation involves 2 million function evaluations. In the estimation phase nondifferential functions are not allowed, and stochasticity in biological parameters is not allowed, at least not in values that affect the "likelihood function". In AD model builder code the stochastic simulation phase is identified as **mc_eval_phase** and some functions are only active in this phase (checked in code with **if(mc_eval_phase())**

The historical assessment part is either a statistical Catch-at-age model or a VPA model. In the SCAA model selection can be allowed to change at specified periods. The VPA model operates based on Popes equation. Various methods for treatment of the oldest age group have been tested but what was used here is to use the fishing mortality from the catch-at-age model of the 2 oldest age groups.

The log likelihood of the survey is based on "modified log residuals" i.e $(\frac{log(I+\delta)}{log(\hat{I}+\delta})$ where δ is a small number selected to reduce the effects of sampling errors. Typical value of δ would correspond to index or CNO based on 3-5 otoliths. Including this factor makes the model more robust to small values or zeroes and relatively robust to inclusion of older age groups.

The pattern of CV with age is given in the input files but a multiplier estimated for each survey. In the VPA mode CV is estimated independently for each survey and age group.

The model estimates correlation of residuals in the survey. The correlation is by a first order AR model based on absolute value of difference in age. High value of the correlation parameter approaches a yearfactor downweighting the survey compared to other data.

Standard settings of the model were to use ages 1–13 from survey 4 (Barents sea)+survey 5 (Norwegian sea) . This means in practice that ages 1 and 2 are from survey 4, age 3 from both and ages 4–13 from survey 5. Results from survey 1 were also included in some runs.

Results from the model and plots referred to can be found in the WD9. A summary plot showing some results is shown below (figure 3.3.1).

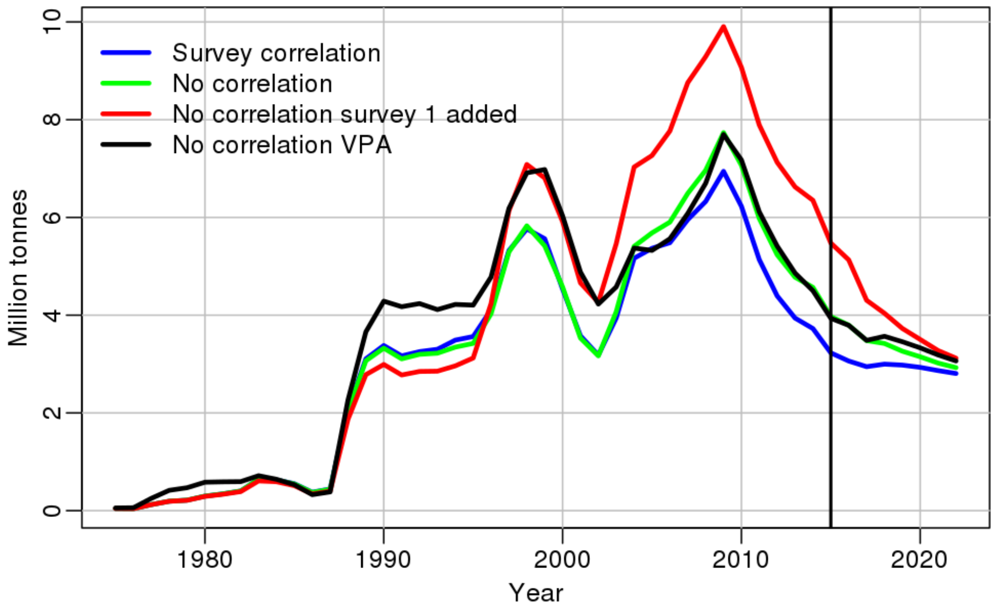

Figure 3.3.1. Results from model with 4 different settings.

Different formulations of the model give spawning stock 2015 between 3.2–5.3 million tonnes (figure 3.3.1). The lower estimates are from runs where correlation of residuals by age is estimated in survey 5, the higher values where it is set to 0. Allowing the selectivity to change occasionally does not affect the model results.

One of the most important results from the model is that the catch-at-age data indicate greater mortality than the survey data. This can also be noted by plotting catch in numbers by age and survey indices by age.

Including survey 1 is a challenge for this model, as the survey was conducted in many years from 1988 to 2008, but not conducted from 2009–2014, to be started again in 2015. The old series will give information about q and CV but there is only one value that affects the most recent estimate.

The model has problems with survey 5 results from last 4 years where survey indices of some cohorts indicate "negative Z".

4 Short-term forecast

4.1 XSAM

A short-term forecast module is not yet ready in the XSAM framework. It will however be ready before WGWIDE in late summer 2016.

The XSAM framework includes a forward dynamical model which allows for utilizing prediction of total catch with specified variance in the assessment year to improve the mortality estimates in the assessment year while accounting for the uncertainty in the prediction of total catch in weight. This approach is justified by empirical evidence that the working groups prediction of the total catch in the assessment year agrees very well with the actual catches with low error (Error coefficient of variation of ~5% for the years 1996–2014, Annex 3). This also means that the forecast of the abundance (and SSB provided mean stock weight at age) at the beginning of the quota year automatically is provided. Due to the time-series model for fishing mortality which is decomposed into a component for selectivity and a component for effort, the prediction of both selectivity and effort in the quota year is directly available. This means that a prediction of the fishing pattern for the quota year is available. By adjusting the predicted level of F for the quota year to match the predetermined F, the corresponding total catch in the quota year is determined, and the forecast for abundance as basis for advice is also available. TMB provides the estimated hessian for parameter estimates and states which is used to approximate their distributional properties. The forecast can therefore be made by simulating from the simultaneous distributions of the parameters and states to maintain the uncertainty in the forecast, and hence include the uncertainty originating from variability of starting conditions as well as uncertainty in the fishing pattern. The approach was discussed and approved during WKPELA but has not yet been coded into the framework for practical use. This requires some minor coding into the framework and will be made available until the next WGWIDE. In general, forecasting of the population also implies forecasting the recruits. For a short-term forecast, the level and variability of recruitment will be of minor importance since they will not affect estimates of SSB or average values of F used for management advice. Hence, the forecast of recruitment can be based on the model estimates provided by XSAM (see Annex 3 for details).

4.2 TASACS

TASACS has a module for short-term predictions that has not been used because of some bugs that reduced the functionality. These have now been fixed. So far, short-term predictions have been made on a spreadsheet.

In connection with the revision of the prediction module, the options for selection at age were revisited.

The following options are now available. Options 1–4 all use fishing mortalities as finally estimated in the assessment and reported in the summary.txt file.

1) From assessment: The estimated fishing mortalities at age in the last assessment year, which is the year prior to the first (intermediate) year in the prediction.

2) Weighted average fishing mortalities for each age over a year range: Sel(a) $= \Sigma_y F(a,y)*N(a,y)/\Sigma_y(N(a,y)$

3) Geometric mean fishing mortality at each age taken over a year range as specified

4) Un-weighted average fishing mortality at each age taken over a year range as specified.

5) Specify a number for each age.

Option 2 is new. The background is that some fishing mortalities at age behave strangely, partly because the algorithm is a VPA with noisy catch data, partly because some year classes start with assumed survivor numbers (Figure 3.2.4).

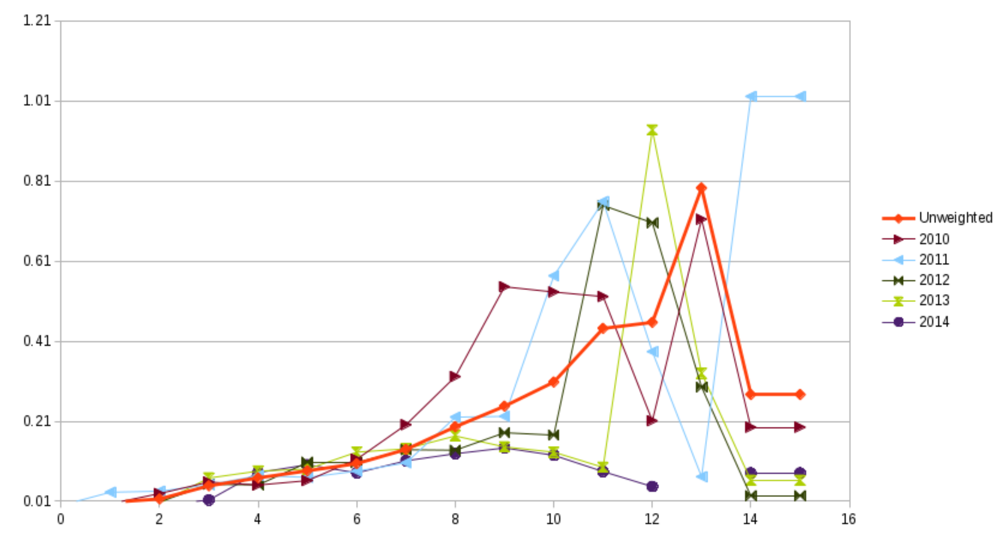

Figure 3.2.4. F at age by year in the standard assessment by WGWIDE. Age 13 in 2014 is not shown, but included in the unweighted average, it was very large.

Some suggestions were discussed by WKPELA, as outlined in Figure 3.2.5. The option 2 emerged from that discussion, and could be further modified by assuming a flat selection above some age (age 8 in the figure). WKPELA accepted the selection weighted flat from age 8.

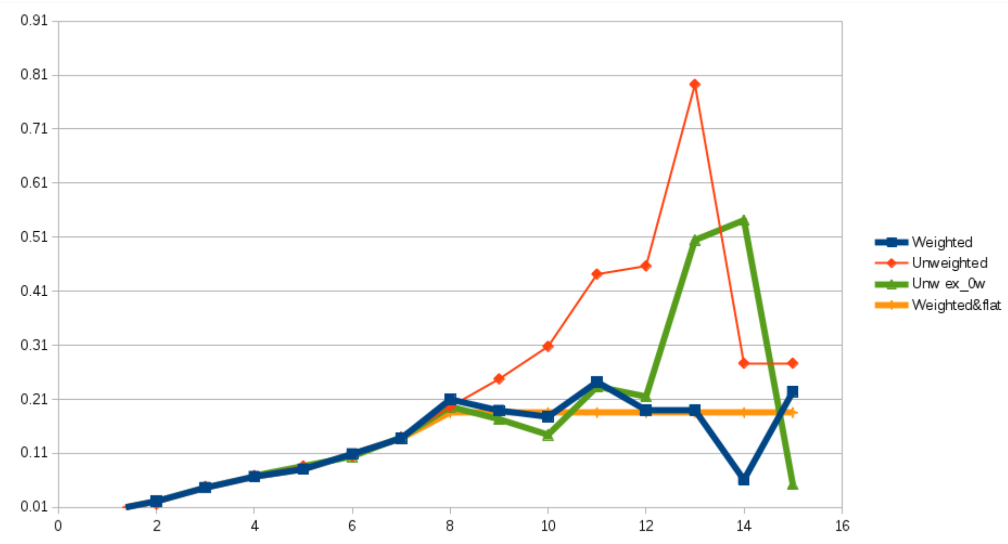

Figure 3.2.5. Suggestions for calculating selection at age in the prediction module in TASACS. The thick blue line is the weighted average (Option 2). The thick green line is unweighted average, but excluding the ages and years where Survey 5 data are ignored. The thin red line is the unweighted average. The orange line is the weighted average, but smoothed by taking a flat line from age 8 onward.

5 Reference points and stock–recruitment functions

The model based on data from 1907 was used for evaluations of reference points. The version used was the VPA model, due to variability of selection, especially before and after the collapse. The survey tuning was with surveys 4+5 without modelling correlation.

Two biomass reference points have been defined for this stock; B_{lim} = 2.5 million tonnes and $B_{trigger}$ =5.0 million tonnes. Here estimate of those reference points is revisited based on estimation of a Hockey stick stock - recruitment function.

The first step was to run the model based on data from 1975. Plotting stock vs. recruitment on log scale demonstrates the stock - recruitment relationship clearly, the small year classes become even smaller when the spawning stock is small (figure 5.1)

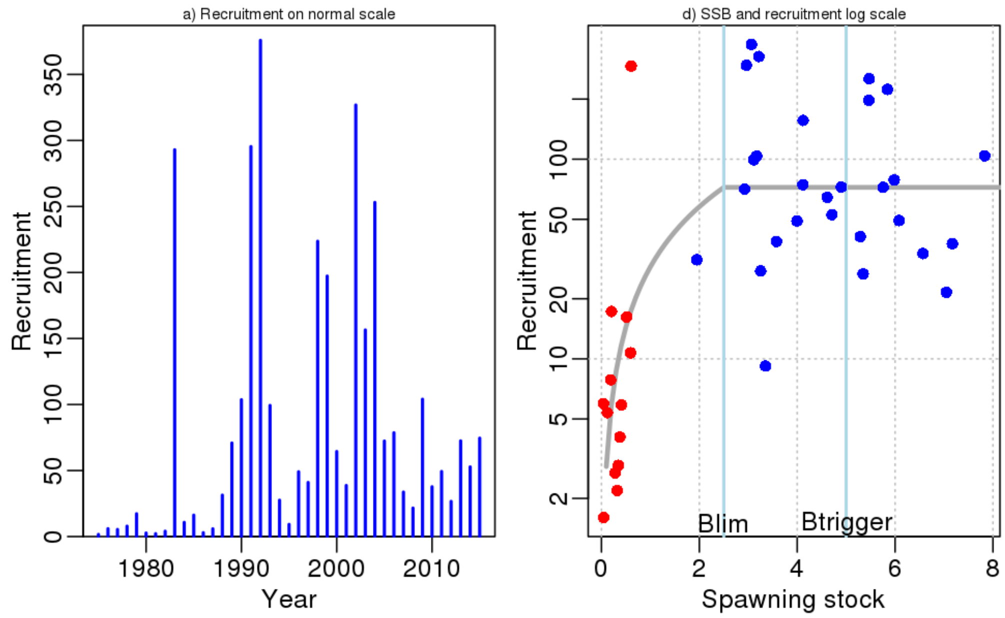

Figure 5.1. Estimated recruitmentvs.time on normal scale (left). Recruitment on log scalevs.SSB (right). The red points are data before 1988, blue 1988–2014

What is apparent from the figures (5.1 and 5.2) is that SSB_{break} is somewhere between 600 thous. and 3000 thous. tonnes, but there is only one data point in between as the spawning stock increased from 600–3000 million tonnes in 2 years (1987–1989) when the large 1983 year class became mature.

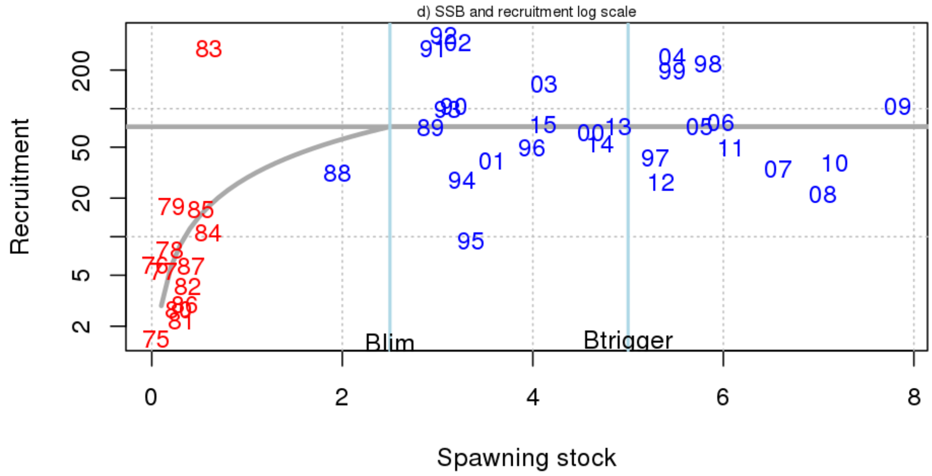

Figure 5.2 Spawning stockvs.recruitment. The text indicates year classes. Hockey stick function with breakpoint at B$_{lim}$ is shown.

The uncertainty in the reference point can be seen in the scatter of estimates of R_{max} and SSB_{break}. (figure 5.3). The scatter of points is with SSB_{break} between 2 –3.2 million tonnes and positive correlation between SSB_{break} and R_{max}, as usual.

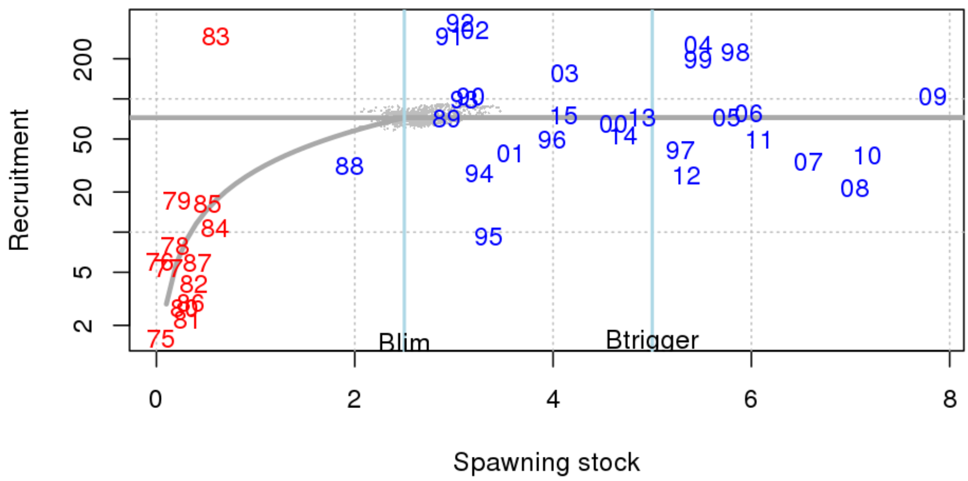

Figure 5.3 SSB-recruitment function with scatter, of R$_{max}$, SSB$_{break}$ pairs shown

The average value of the breakpoint is 2.67/2.68 million tonnes. The former value is geometric mean and latter arithmetic mean. The deterministic values is 2.56 million tonnes.

Adding the data since 1907 fills some of the gaps but there are still few data points with SSB between 1–2 million tonnes (Figures 5.5 and 5.7). Those data must be taken with caution as in many of those years substantial amount of young herring (ages 0–2) is caught and value of M used for those age groups becomes decisive for how large the assumed recruitment really was in those days compared to recently. The lower the value of M the more influential were the fisheries of small herring in depletion of the stock. The results shown here are from a VPA model with F of the oldest age groups obtained from a separable model (with 5 selection periods). Assumptions about F of the oldest age group does in many cases have relatively large effect on the results. Estimated breakpoint in the hockey-stick relationship is 2.89 million tonnes for the VPA model but 2.38 million tonnes for the separable model, but these estimates are not significantly different.

Average recruitment is 129 milliards in the period 1922–1966, but 106 milliards after 1988, not a large difference taking into account the change in selection pattern. Relatively small change in M of age 0–2 would make the recruitment in the latter period higher.

According to the model results spawning stock was relatively small at the beginning but maturity-at-age is also very low in the early part of the series. The exact size of the fishable stock at the beginning of the series is mostly dependent of the assumption about F on the plus group, and the SSB obtained here was lower than the values by Toresen and Østvedt 1997.

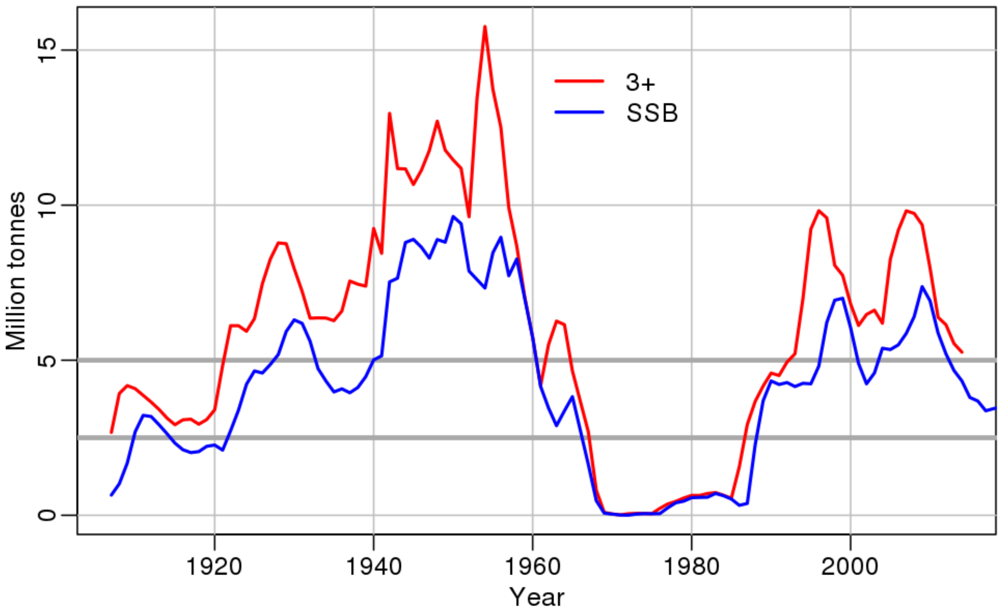

Figure 5.4. Development of spawning stock and biomass 3+ according to the model

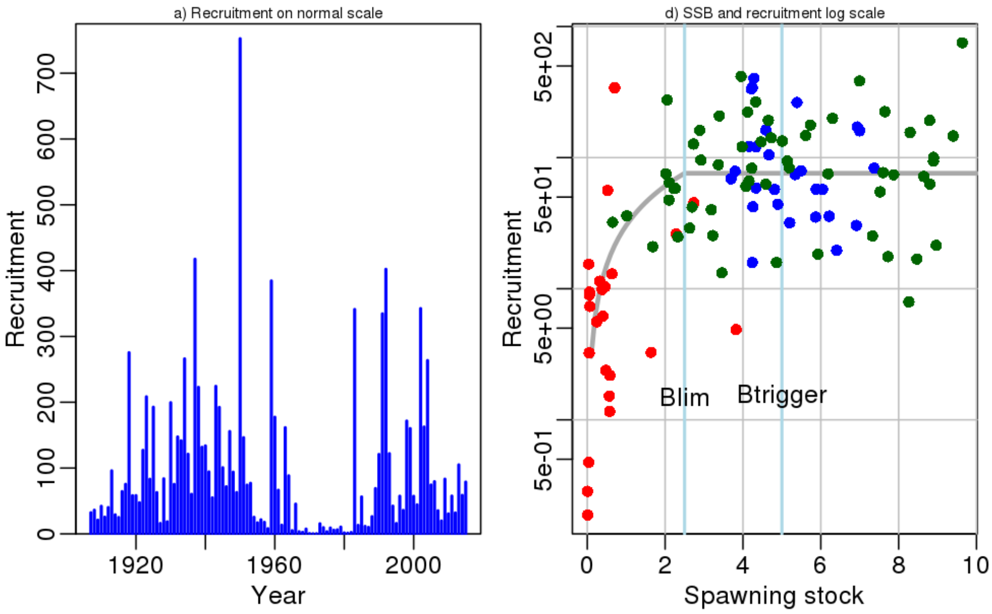

Figure 5.5. Estimated recruitmentvs.time on normal scale (left) and recruitment on log scale vs. SSB (right) . The dark green points are data before 1965, red points 1966-1987 and blue points 1988–2014

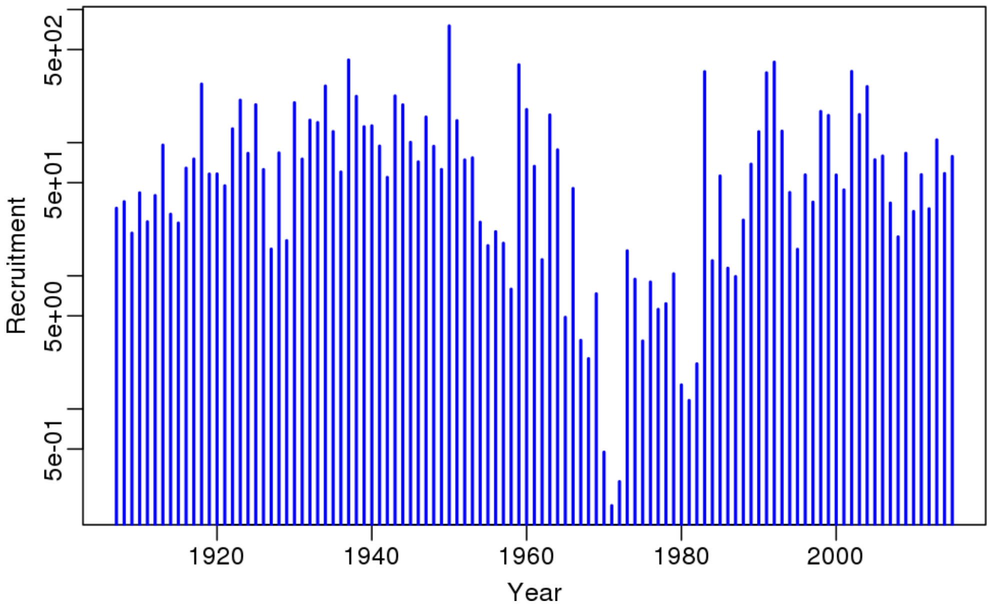

Figure 5.6. Estimated recruitment 1907–2014 on log scale

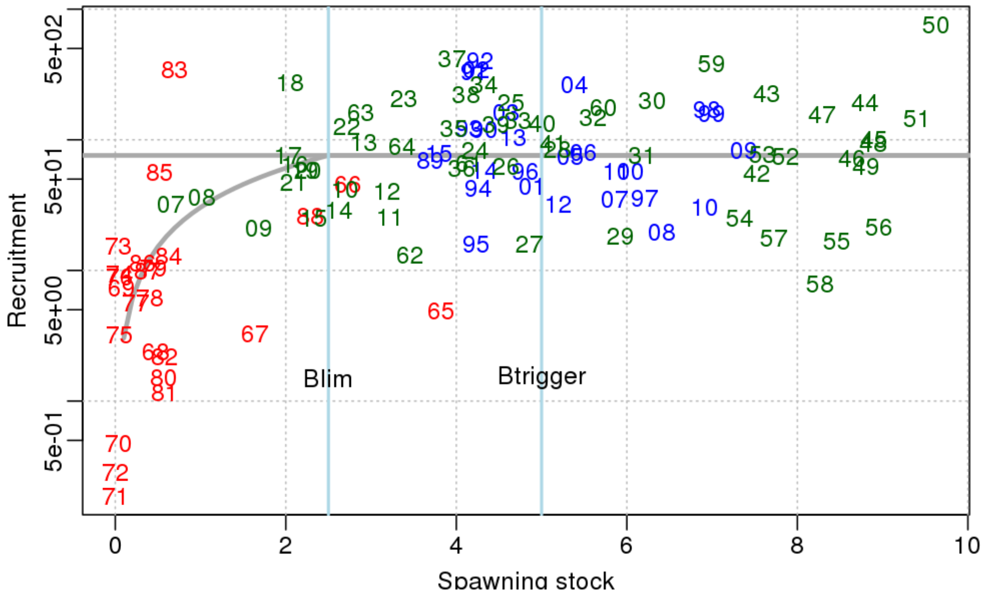

Figure 5.7. Spawning stockvs.recruitment. The text indicates yearclasses. Hockey stick function with breakpoint at Blim is shown for reference. Values before 1965 are dark green, 1966–1988 red and 1989–2014 blue.

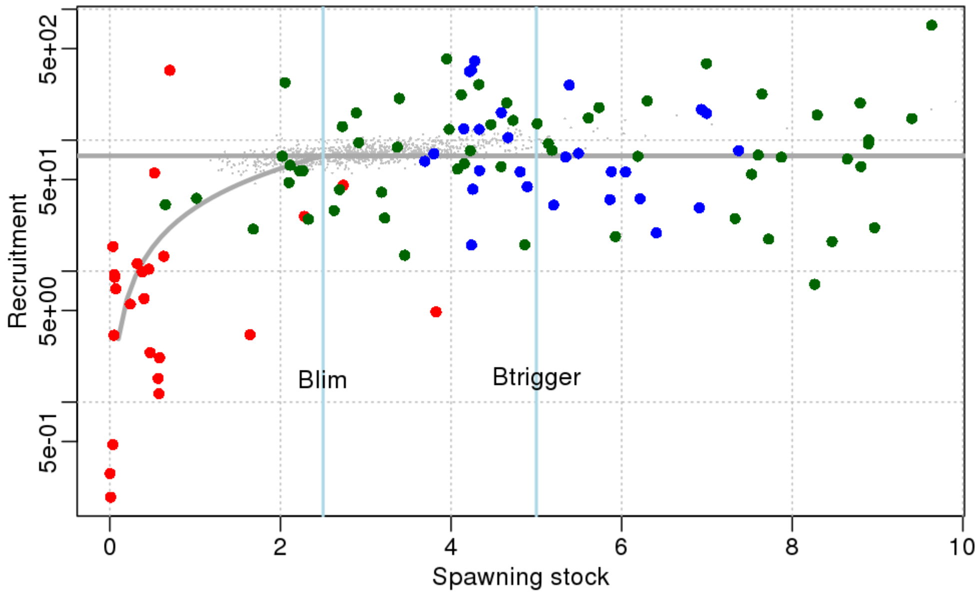

Figure 5.8. SSB-recruitment function with scatter, of R_{max}, SSB_{break} pairs shown. Values before 1965 are dark green, 1966–1988 red and 1989–2014 blue.

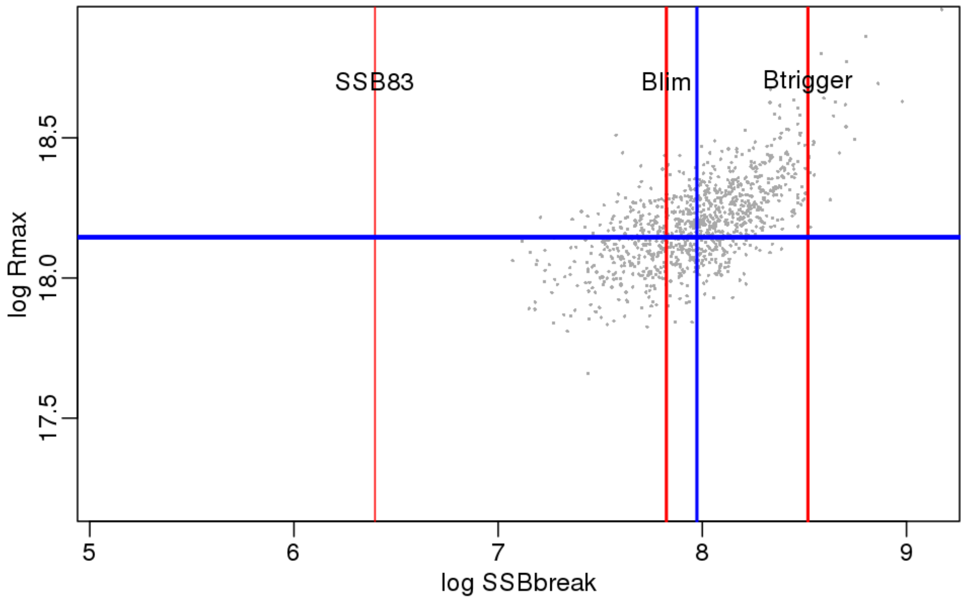

Figure 5.9. Scatter of estimated SSB_{break}, R_{max} pairs. The blue vertical line shows the average

Plotting recruitment on log scale (figure 5.7) shows clearly how small some yearclasses were around the collapse of the stock but yearclasses 5–50 times smaller than "normal small" yearclass were seen in that period.

The series since 1907 gives longer dataseries to infer about autocorrelation of recruitment. The result (figure 5.10) is a first order AR coefficient of around 0.3 on log residuals. This might be a candidate value to use in stochastic long-term predictions but it would still have to be checked against other metrics like 5th percentile of average recruitment over a decade.

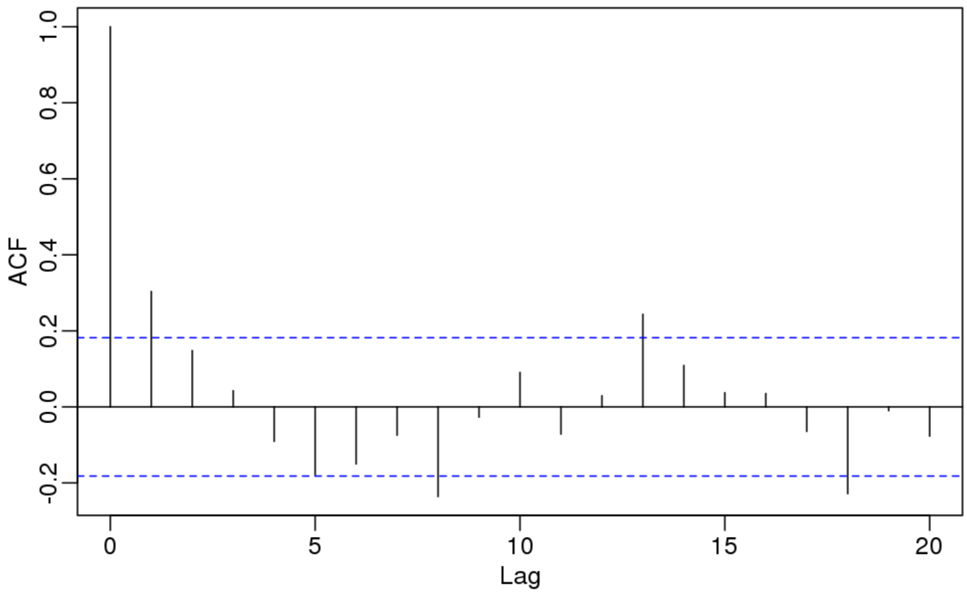

Figure 5.10. Autocorrelation of recruitment residuals 1907–2014

The result of all these exercises is that the value of B_{lim} is appropriate if it should be the point below which recruitment starts to decrease. There is substantial uncertainty in estimation of this point as the stock has not very often been in the range 1–3 million tonnes. The value of B_{lim} is relatively high compared to many other stocks.

6 Evaluation of MSY and $B_{trigger}$

Evaluations of MSY are done with the following settings.

1) MCMC evaluations of an assessment model (see Annex 5) run with parameters saved. Among parameters are the 3 parameters of the Hockey stick SSB-R function R_{max}, SSB_{break} and σ

2) The values saved are used in stochastic simulations.

3) Assessment error $CV = 0.3, \rho = 0.7$ used, based on relatively long periods of over and underestimation. CV=0.12 based on estimated relative uncertainty of biomass in the assessment year 16%. This is usually underestimate so the real CV might be closer to 0.2. There is increase from CV in biomass in the assessment year to CV of F in the advisory year but as fishing mortality is low 0.25 might be a more appropriate value.

4) CV in weights, year factor with CV=0.12 and $\rho = 0.7$. Estimated from data but density-dependence ignored (including it will reduce risk)

5) Autocorrelation of recruitment residuals 0.33 on log scale, estimated from data since 1907. This is the factor having largest effect of the result.

6) Maturity fixed. Selection fixed in each run as the average of last 15 years.

The model used is tuned with surveys 4 and 5 and correlation of survey residuals is not taken into account. The settings are similar to the settings used in TASACS last year and the result today is somewhere in the middle of the values obtained (3.2–5 million tonnes) including runs tuned with survey 1.

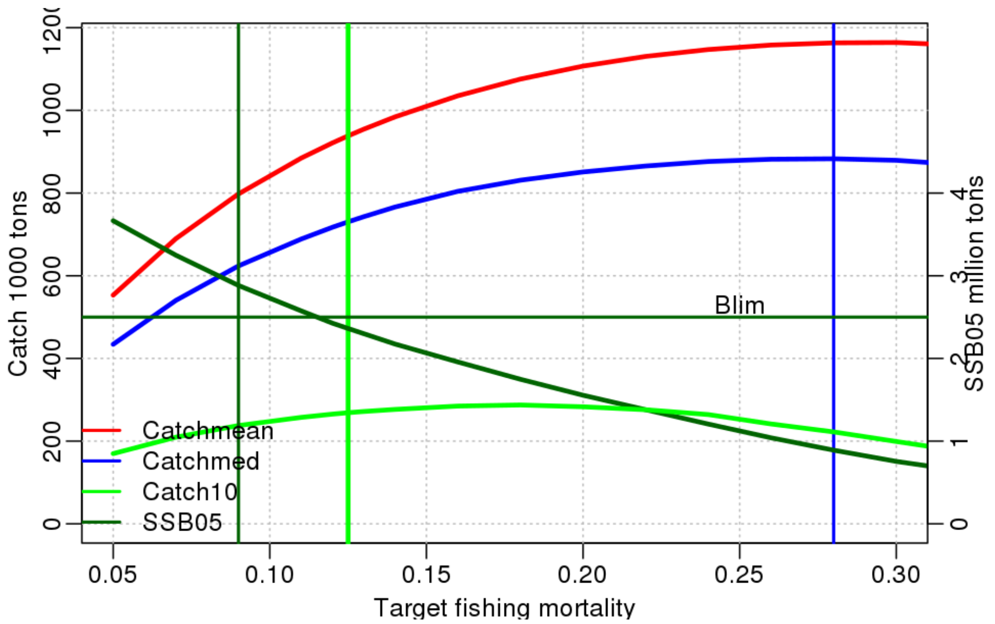

Figure 6.1. Average, median and 10th percentile of catch at "equilibrium". Fifth percentile of SSB is also shown as well as the line that maximizes median of catch, F=0.125 and the line corresponding to P(SSB < B_{lim}=0.05)

The results (figure 6.1) show that long-term catch is maximized around $F = 0.16$ but this is not precautionary. Applying precautionary consideration will result on F leading to more than 5% probability being above B_{lim} is 0.12. There is of course no trigger action here but a trigger will increase F_{pa}. What is seen here is typical for stocks with large recruitment variability and autocorrelation. Fishing mortality giving maximum yield is higher than that fulfilling the precautionary criteria. The F in the management plan (F=0.125) seems like a good compromise, not sacrificing much of average yield.

Autocorrelation of recruitment is the factor having largest effect of the results here, assessment error and variability of weights have much smaller effect. The values used are estimated from the series since 1907(figure 5.10). Increased autocorrelation makes long periods of poor recruitment more likely and does therefore affect precautionary criteria more than median yield. Not including autocorrelation leads to $F_{pa} = 0.135$ (F leading to $SSB_{05} = 2.5\ million\ tonnes$ is called F_{pa} here) but increasing autocorrelation to 0.5 leads to $F_{pa} = 0.07$. The question is really to store parts of the large year classes for periods of poor recruitment. As stated above the current fishing mortality of 0.125 seems like a good compromise.

6.1 Testing the trigger

When a fishing mortality of 0.125 has been selected the same simulations can be used to get the value of the trigger leading to $P(SSB < B_{lim} = 0.05)$. Looking at the results (Figure 6.2) $B_{trigger}$ of 3.2 million tonnes, would be sufficient and $B_{trigger} = B_{lim}$ to reduces the probability of $SSB < B_{lim}$ compared to no $B_{trigger}$. When the stock is underestimated target fishing mortality will be reduced if SSB is estimated to be below $B_{trigger}$ but the action is somewhat one-sided as F will not increase to more than 0.125 if SSB is estimated to be above $B_{trigger}$.

ICES follows a principle that $B_{trigger}$ can not be lower than B_{pa}. that could be estimated as $2.50 \times e^{1.645 \times \sigma}$ or between 3.2–.5 if σ varies from 0.15–0.25. (σ here refers to uncertainty in spawning stock in the assessment year, uncertainty in F one year later for a given catch is considerably higher.)

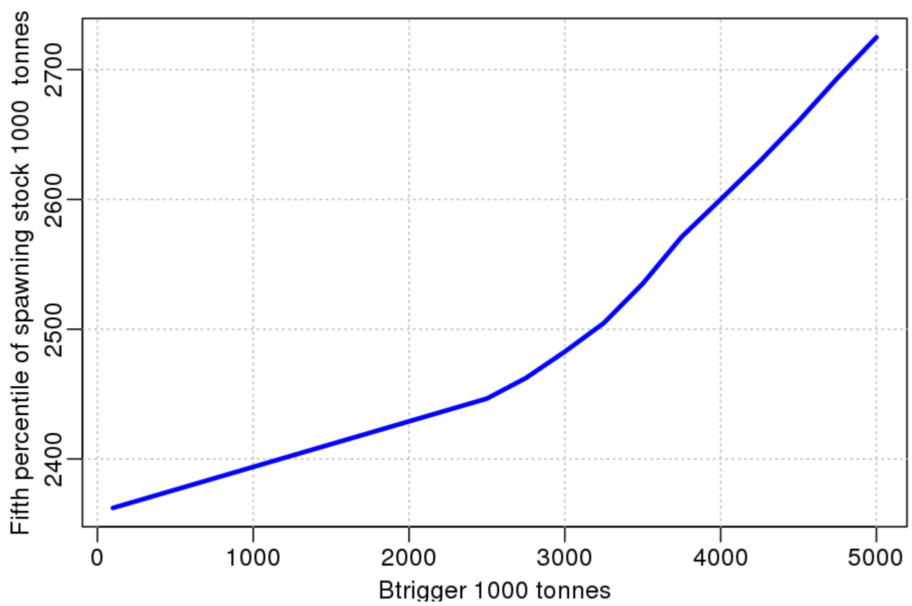

Figure 6.2. Fifth percentile of spawning stock as function of $B_{trigger}$ if HCR is based on F=0.125

7 References

Aanes, S., and Vølstad J.H. 2015. Efficient statistical estimators and sampling strategies for estimating the age composition of fish. Canadian journal of fisheries and aquatic sciences 72 (6): 938–953.

Aanes, S., Engen, S, Sæther, B.-E. and Aanes, R. 2007. Estimation of fish stock dynamical parameters from catch-at-age data and indices of abundance: can natural and fishing mortality be separated? Canadian journal of fisheries and aquatic sciences, 64: 1130–1142.

Engelhard, G.H., Dieckmann, U and Godø, O.R. 2003. Age at maturation predicted from routine scale measurements in Norwegian spring-spawning herring (*Clupea harengus*) using discriminant and neural network analyses. ICES Journal of Marine Science, 60: 304–313.

Engelhard, G.H. and Heino, M. 2004. Maturity changes in Norwegian spring-spawning herring before, during, and after a major population collapse. Fisheries Research, 66: 299-310.

Gavaris, S. 1988. An adaptive framework for the estimation of population size. Can. Atl. Fish. Sci. Adv. Comm. (CAFSAC) Research Doc 88/29.

Gudmundsson, G., 1994. Time-series analysis of catch-at-age observations. Journal of the Royal Statistical Society: Series C (Applied Statistics) 43, 117–126.

ICES 2008: Report of the Working Group on Widely Distributed Stocks (WGWIDE) ICES CM2008/ACOM:13

ICES. 2010a. Report of the Workshop on estimation of maturity ogive in Norwegian spring-spawning herring (WKHERMAT), 1–3 March 2010, Bergen, Norway. ICES CM 2010/ACOM:51. 47 pp.

ICES. 2010b. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 28 August –3 September 2010, Vigo, Spain. ICES CM 2010/ACOM:12.

ICES 2011. Report of the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, 8–10 November 2011, Bilbao, Spain. ICES CM 2011/ACOM:52.

ICES 2012. Report of the Second Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, 6–9 November, Copenhagen. ICES CM 2012/ACOM:54.

ICES. 2014. Report of the third Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, 19–22 November 2013, ICES HQ, Copenhagen, Denmark. ICES CM2013/ACOM:54. 109 pp.

ICES. 2015. Report of the Working Group of International Pelagic Surveys (WGIPS), 19–23 January 2015, ICES HQ, Copenhagen, Denmark. ICES CM 2015/SSGIEOM:05. 279 pp.

ICES. 2016a. Report of the working group of international pelagic surveys (WGIPS). ICES CM 2015/SSGIEOM: (in press).

ICES. 2016b. Report of the Working Group on Commercial Catches (WGCATCH), 9–13 November 2015, Lisbon, Portugal. ICES CM 2015/SSGIEOM:34. 111 pp.

Kristensen, K. 2014. TMB: General random effect model builder tool inspired by ADMB. R package version 1.1.

Nielsen, A., and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158: 96–101.

Óskarsson, G.J., Kjesbu, O.S., Slotte, A. 2002. Predictions of realized fecundity and spawning time in Norwegian spring-spawning herring (*Clupea harengus*). Journal of Sea Research 48: 59–79.

Quinn, T.J., and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York.

Salthaug, A. and Johnsen, E. 2014. Validation of Norwegian spring-spawning herring surveys. Working document in the Report of the working group on widely distributed stocks (WGWIDE). ICES CM 2014 / ACOM:15.

Salthaug, A. and Aanes, S. 2015. Estimating the Norwegian catch-at-age of blue whiting, mackerel, North Sea herring and Norwegian spring-spawning herring with the ECA model. Working document in the Report of the working group on widely distributed stocks (WGWIDE). ICES CM 2015 / ACOM:15.

Skagen, D.W. and Skålevik, Å (2009). A Toolbox for Age-structured Stock Assessment using Catch and Survey data (TASACS) Institute of Marine Research: Fisken og Havet FH2009-1, http://www.imr.no/publikasjoner/andre_publikasjoner/fisken_og_havet/2009/nb-no

Stenevik, E., Vølstad, J.H, Høines, Å., Aanes, S., Óskarsson, G., Jacobsen, J., and Tangen, Ø. 2015. Precision in estimates of density and biomass of Norwegian spring-spawning herring based on acoustic surveys. Marine Biology Research 11 (5): 441–461.

Working documents

WD01. Björnsson, H. 2004. Use of an offset parameter in log ratios.

WD02. Skagen, D.W. 2016. On the retrospective error in the NSSH assessment with TASACS.

WD03. Óskarsson, G.J., Slotte, A. and Hömrum, E. 2016. How well does the biological sampling of Norwegian spring-spawning herring represent the acoustic registrations with respect to depth in IESNS?

WD04. Aanes, S. 2016a. A statistical model for estimating fish stock parameters accounting for errors in data: Applications to data for Norwegian Spring-spawning herring. WD for ICES WKPELA 2016. (see Annex 3)

WD05. Stenevik, E.K. 2016. Updating the maturity ogive for Norwegian spring-spawning herring (NSSH).

WD06. Stenevik, E.K. 2016. Re-estimating of the herring larval index (SSB fleet) using the new open source software StoX

WD07. Homrum, E. í, Óskarsson, G.J., Slotte, A. 2016. Working Document to WKPELA 2016. Spatial, seasonal and interannual variations in growth and condition of Norwegian spring-spawning herring during 1994–2015. 41 pp.

WD08. Erling Kåre Stenevik, E.K. 2016. Re-estimating of the herring larval index (SSB fleet) using the new open source software StoX.

WD09. Björnsson, H. 2016. Investigation of data and assessment methods for Norwegian spring-spawning herring (NSSH)

WD10. Salthaug, A. and Slotte, A. 2016. Re-estimation of abundance from the spawning survey of Norwegian spring-spawning herring using the new open source software StoX.

WD11. Same as WD07 including additional work during WKPELA.

WD12. Aanes, S. 2016b. Diagnostics of models fits by XSAM to herring data. WD for ICES WKPELA 2016. (see Annex 4)

WD13. Björnsson, H. 2016. Working document on assessment model for Norwegian Spring-spawning Herring (see Annex 5).

Annex 1: List of participants

Name	Address	Phone/Fax	E-mail
Asta Gudmundsdóttir Chair	Marine Research Institute Skùlagata 4 Reykjavik 121 Iceland	+354 5752000	asta@hafro.is
Åge Høines	Institute of Marine Research Nordnes 5817 Bergen Norway		
Alexander Krysov	Knipovich Polar Research Institute of Marine Fisheries and Oceanography(PINRO) 6 Knipovitch Street 183038 Murmansk Russian Federation	+79113266589	a_krysov@pinro.ru
Are Salthaug	Institute of Marine Research P.O. Box 1870 Nordnes 5817 Bergen Norway		are.salthaug@imr.no
Aril Slotte	Institute of Marine Research PO Box 1870 Nordnes 5817 Bergen Norway	+4799519782	aril.slotte@imr.no
Audun Maråk	Fiskebåt Røysegata 15 6001 Ålesund Norway		audun@fiskebat.no
Carey McGilliard	NOAA Fisheries 1315 East West Highway Silver Spring, MD 20910-6233 United States		carey.mcgilliard@noaa.org
Claus Reedtz Sparrevohn	Danish Pelagic Producersorganisation HC Andersens Boulevard 37, DK-1553 Copenhagen V. Denmark	Phone: (+45) 29124549	crs@pelagisk.dk

Name	Address	Phone/Fax	E-mail
Dankert Skagen	Fjellveien 96 5019 Bergen Norway	+47 93257452	dankert@dwsk.net
Erling Kåre Stenevik	Institute of Marine Research PO Box 1870 Nordnes NO-5817 Bergen Norway	+47 915 98 748	erling.stenevik@imr.no
Eydna í Homrum	Faroe Marine Research Institute PO Box 3051 FO-110 Tórshavn Faroe Islands	+298 353922	eydnap@hav.fo
Gudmundur J. Oskarsson	Marine Research Institute Skùlagata 4 Reykjavik 121 Iceland	+354 5752000	gjos@hafro.is
Höskuldur Björnsson	Marine Research Institute Skùlagata 4 Reykjavik 121 Iceland	+354 5752000	hoski@hafro.is
Jan Arge Jacobsen	Faroe Marine Research Institute Nóatún 1 Tórshavn 110 Faroe Islands	+298 353900	janarge@hav.fo
Jan Birger Jørgensen	Norwegian Fishermen's Association Havnegt 9 oppgang C 5.etg i Pirsenteret NO-7462 Trondheim Norway	+47 930 44 346	jan.birger.jorgensen@fiskarlaget.no
Jan Horbowy	National Marine Fisheries Research Institute ul. Kollataja 1 81-332 Gdynia Poland	+48 587-356-267	horbowy@mir.gdynia.pl
Jason Cope (External Chair)	NOAA Fisheries 1315 East West Highway Silver Spring, MD 20910-6233 United States		jason.cope@noaa.gov
Jens Christian Holst	'Ecosystembased Advisors and Developers AS'		jens@ecosystembased.com

Name	Address	Phone/Fax	E-mail
Jonny Berfjord	Norwegian Fishermen's Association Havnegt 9 oppgang C 5.etg i Pirsenteret NO-7462 Trondheim Norway		jonny@bnord.no
Katja Enberg	Institute of Marine Research PO Box 1870 Nordnes 5817 Bergen Norway	+47 410 06368	katja.enberg@imr.no
Mark R Payne	Centre for Ocean Life National Institute of Aquatic Resources Technical University of Denmark (DTU-Aqua) Kavalergaarden 6 2920 Charlottenlund Denmark		mpay@aqua.dtu.dk
Samuel Subbey	Institute of Marine Research Nordnes 5817 Bergen Norway	+47 5523 8687	samuel.subbey@imr.no
Sondre Aanes	Norwegian computing Center PO Box 114 NO-0314 Oslo Norway		sondre.aanes@nr.no

Annex 2: WKPELA Terms of Reference

2014/2/ACOM35 A **Benchmark Workshop on Pelagic stocks** (WKPELA), chaired by ICES Chair Asta Gudmundsdóttir, and attended by invited external experts Jason Cope, (USA), Jan Horbowy, (Poland) and Carey McGilliard, (USA) will be established and will meet in ICES HQ, 24–26 November 2015 for a data evaluation meeting and 29 February–4 March 2016 for a Benchmark meeting to:

a) Evaluate the appropriateness of data and methods to determine stock status and investigate methods for short-term outlook taking agreed or proposed management plans into account for the stocks listed in the text table below. The evaluation shall include consideration of: i. Stock identity and migration issues;

ii. Life-history data;

iii. Fishery-dependent and fishery-independent data;

iv. Further inclusion of environmental drivers, multispecies information, and ecosystem affects for stock dynamics in the assessments and outlook

b) Agree and document the preferred method for evaluating stock status and (where applicable) short-term forecast and update the stock annex as appropriate. Knowledge of environmental drivers, including multispecies interactions, and ecosystem affects should be integrated in the methodology. If no analytical assessment method can be agreed, then an alternative method (the former method, or following the ICES data-limited stock approach) should be put forward;

c) Evaluate the possible implications for biological reference points, when new standard analyses methods are proposed. Propose new MSY reference points taking into account the WKFRAME2, results and the introduction to the ICES advice (section 1.2), WKMSYREF3 and WKMSYREF4.

d) Develop recommendations for future improving of the assessment methodology and data collection;

e) As part of the evaluation:

i. Conduct a 3 day data evaluation workshop. Stakeholders are invited to contribute data (including data from non-traditional sources) and to contribute to data preparation and evaluation of data quality. As part of the data compilation workshop consider the quality of data including discard and estimates of misreporting of landings;

ii. Following the Data Evaluation, produce working documents to be reviewed during the Benchmark meeting at least 7 days prior to the meeting.

STOCKS	STOCK LEADER
Her-noss	Erling Stenevik erling.stenevik@imr.no