ICES IBPBLW REPORT 2016

ICES ADVISORY COMMITTEE

ICES CM 2016/ACOM:36

REF. WGSAM, ACOM

Report of the Inter-Benchmark Protocol for Blue Whiting (IBPBLW)

10 March–10 May 2016

By correspondence

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

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

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

Recommended format for purposes of citation:

ICES. 2016. Report of the Inter-Benchmark Protocol for Blue Whiting (IBPBLW), 10 March–10 May 2016, By correspondence. ICES CM 2016/ACOM:36. 118 pp. https://doi.org/10.17895/ices.pub.19284041

Contents

Executive summary

The Inter-Benchmark Protocol of Blue Whiting (IBPBLW) convened by correspondence from 10 March to 10 May 2016. The meeting was chaired by Patrícia Gonçalves. There were 16 participants from eight countries; 13 scientists and three representing the industry. One independent scientist from outside the ICES community participated in the process and reviewed the conclusions of the inter-benchmark protocol. The participant list is in Annex 1.

The main purpose of the meeting was to evaluate the robustness of the SAM model in situations with clear "year effects" in survey indices as observed in the 2015 International Blue Whiting Spawning Stock Survey (IBWSSS). The use of catches-in-numbers-at-age for the first one or two quarters in the assessment year should be explored both in terms of getting more realistic F in the intermediate year and also on the retrospective pattern seen for the stock. Since the benchmark in 2012, recruitment indices from several surveys have been used qualitatively to predict the year-class size (small, medium, large). Now an attempt should be made to model the indices. Finally Biological Reference Points should be re-evaluated.

The IBPBLW agreed on using a newly developed option for the SAM model as the standard assessment (model 3c). This option takes account to correlation of observations.

The IBPBLW concluded that preliminary age-disaggregated catches from the first two quarters (Q1+Q2) in the assessment year were preferable over only catches from quarter 1. They give a more realistic F in the intermediate year, but do not affect the retrospective pattern in the assessment.

The review of recruitment indices was not finished. More work will be done before WGWIDE in August 2016.

IBPBLW pointed out that there is an ongoing management strategy evaluation (MSE) for blue whiting. Reference point evaluations should preferably be tightly connected to the MSE, because harvest control rules and reference points inherently go together. Therefore reference points were not re-evaluated by IBPBLW.

1 Introduction

The Inter-Benchmark Protocol for Blue Whiting (IBPBLW) considered the assessment of blue whiting (*Micromesistius poutassou*) combined stock (whb-comb; Subareas I–IX, XII and XIV). The reason for this inter-benchmark was the outcome of the assessment of blue whiting made during WGWIDE 2015. By using the same configuration of SAM as in recent years, unrealistically high F was estimated for 2015, based on the IBWSS survey indices from spring 2015. The indices in 2015 were very low compared to the two preceding years, especially for ages 4 and older.

The blue whiting had previously been through a benchmark process in 2012 (ICES, 2012). At that meeting SAM was chosen to be the assessment tool of the blue whiting. Therefore, the focus at IBPBLW was to address issues related to the data and the SAM model.

The ToRs for the IBPBLW are:

a) evaluate the robustness of the SAM model in situations with clear "year effects" in survey indices as observed in the IBWSS 2015, test appropriate model modifications and /or make criteria for posterior discarding of survey indices;

b) estimate recruitment for short-term forecast from the present available survey indices;

c) retrospectively examine the impact of including Q1 catches in the assessment year;

d) estimate Biological Reference Points.

The inter-benchmark took place over two months, from 10th of March to 10th of May of 2016, by correspondence, and a total of six plenary web-meetings (10th March, 18th March, 19th April, 2nd May, 10th May and 13th May) have been conducted (via Skype for Business). A total of 16 participants attended and contributed in the protocol, mostly scientists and also industry representatives. The list of participants is in Annex 1.

One independent scientist from outside the ICES community reviewed and provided comments and inputs during the discussions:

Richard Methot, USA

This report summarizes decisions made at the IBPBLW meetings. Further details on analyses and data considered in making these decisions are detailed in the working documents in the supplementary material (Annex 2) attached to this report.

2 Methods and overview

2.1 The inter-benchmark protocol

In the first meeting, an individual scientist was asked to lead each ToR. The ToR leaders were responsible for the work conducted under their ToR, the investigations and conducted the discussions during the meetings. The ToR leaders were:

ToR a) Morten Vinther (Denmark)

ToR b) Mark Paine (Denmark)

ToR c) Thomas Brunel (Netherlands)

ToR d) Martin Pastoors (Netherlands)

The plenary meetings were held to identify progress, discuss the problems and planning the next steps. The ToR leaders were encouraged to submit their work in working documents at least a day before the meeting.

3 Blue whiting (Subareas I–IX, XII and XIV)

3.1 Catch data

The assessment is run with catch data from 1981 onwards. In the assessment made during WGWIDE 2015, then 2014 was the last catch datapoint. The available catch-at-age data are for the whole year. One of the items was to explore the effect of having the catch-at-age for the first quarter of the year.

Catch-at-age data for quarter 1 (Q1) from the time-series for blue whiting (whb-comb) were made available for the years 2000–2006 and 2010–2015. The data submission for the IBPBLW only included the 2015 quarter 1 catch-at-age data.

The catch-at-age data from 2007 until 2014 were exported from the InterCatch, but the catch-at-age by quarter were not available for the years 2007–2009.

Total catches in 2015, quarter 1, were estimated to 550 807 tonnes based on data provided by WGWIDE members. Total catches by country for the 1st quarter of 2015 and sampling intensity (No. of samples, No. of fish measured and No. of fish aged) are presented in Table 3.1.1. Total catches by area for the 1st quarter of 2015 are presented in Figure 3.1.1.

Table 3.1.1. Blue whiting. ICES estimates, No. of samples, No. of fish measured and No. of fish aged by country for 2015 quarter 1.

Country	Quarter	Catches (t)	No. Samples	No. Fish Measured	No. Fish Aged
Estonia	1	0	0	0	0
Denmark	1	28 601	9	399	399
Faroe Islands	1	84 840	8	863	500
France	1	5994	0	0	0
Germany	1	10 023	0	0	0
Iceland	1	19 766	10	931	249
Ireland	1	20 515	11	3463	1098
Lithuania	1	0	0		0
Netherlands	1	24 783	60	1499	1499
Norway	1	282 070	29	1654	650
Portugal	1	166	13	800	144
Russia	1	47 159	26	4500	399
Spain	1	7392	25	3524	625
Sweden	1	0	0	0	0
UK (England + Wales)	1	0	0	0	0
UK (Northern Ireland)	1	1119	0	0	0
UK (Scotland)	1	18 378	5	666	236
Total		550 807	165	13 145	5274

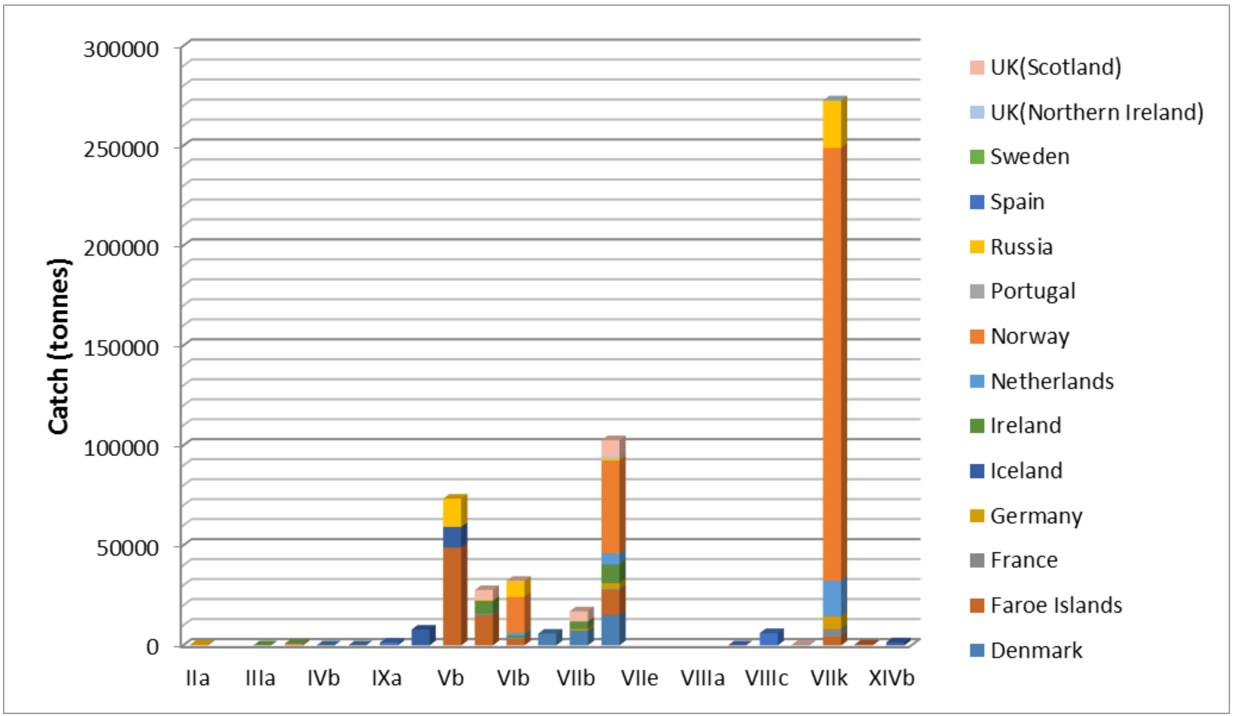


Figure 3.1.1. Blue whiting ICES estimates (tonnes) in 2015 (Quarter 1) presented by ICES area and country.

3.2 Model modifications

The abundance indices from the International Blue Whiting spawning stock survey, IBWSS, in 2015 showed to be unexpectedly low, especially for the older age groups. Survey experts considered the result as "robust" as the survey was conducted as planned, although the weather conditions were not as favourable as in the two previous years, which could bias the result. Preliminary data from the commercial landings in 2015 partly supported the decline in abundance of older fish. The age composition of landings showed a decline in the proportion of old fish for some countries, while other countries had a similar age composition in 2014 and 2015. Since there was no clear justification for exclusion of the 2015 IBWSS data, the data were included in the assessment. However, the assessment model estimated F for 2015 unrealistically high to fit the model. This assessment was done without any information on catches in 2015 (ICES, 2015).

Although the assessment in 2015 was an update assessment, a small change was made to the parameter reflecting the timing of the spawning survey within a year (ICES, 2015) (see WD1 Morten Vinther, Section 1.2.2).

In order, to evaluate ToR a), that is the robustness of the SAM model in situations with clear "year effects" in survey indices as observed in the IBWSS 2015, some model modifications were tested (WD 1 Morten Vinther). They were based on the newly developed options for the SAM model, described in the paper entitled "Accounting for correlated observations in an age-based state-space stock assessment model" by Berg and Nielsen (2016).

Five models were investigated:

1) All observations are independent. No parameters are related to correlation.
2) Regular lattice AR(1) observation correlation structure for all fleets. One correlation parameter per fleet, although surveys may share parameters.
3) Irregular lattice AR(1) observation correlation structure for all fleets. Between 2 and A_f -1 parameters per lattice, and lattice parameters are allowed to be shared among fleets.
4) Unconstrained observation correlation structure for commercial catches and irregular lattice AR(1) observation correlation structure for all surveys.
5) Unconstrained observation correlation structure for all fleets. A_f (A_f -1)/2 parameters per fleet

The main model diagnostics for the final configurations of Model 1–5 are shown in the table below (Table 3.2.1).

Table 3.2.1. Model 1–5 (final configurations) diagnostics.

model	nlogl	nopar	nobs	AIC	AICc
Model 1	310.94	14	406	649.9	651.0
Model 2	275.51	16	406	583.0	584.4
Model 3	274.12	17	406	582.2	583.8
Model 4	251.83	60	406	623.7	644.9
Model 5	240.29	74	406	628.6	662.1

Based on the AIC or AICc criterion, Model 3 is considered the best, followed by Model 2. Model 4 and Model 5 are over-parameterised.

A comparison of the assessment results (Figure 3.2.1) shows similar F_{bar} and SSB for Models 2, 3 and 4. Model 5 has a lower F_{bar} and a lower SSB compared to Models 2–4.

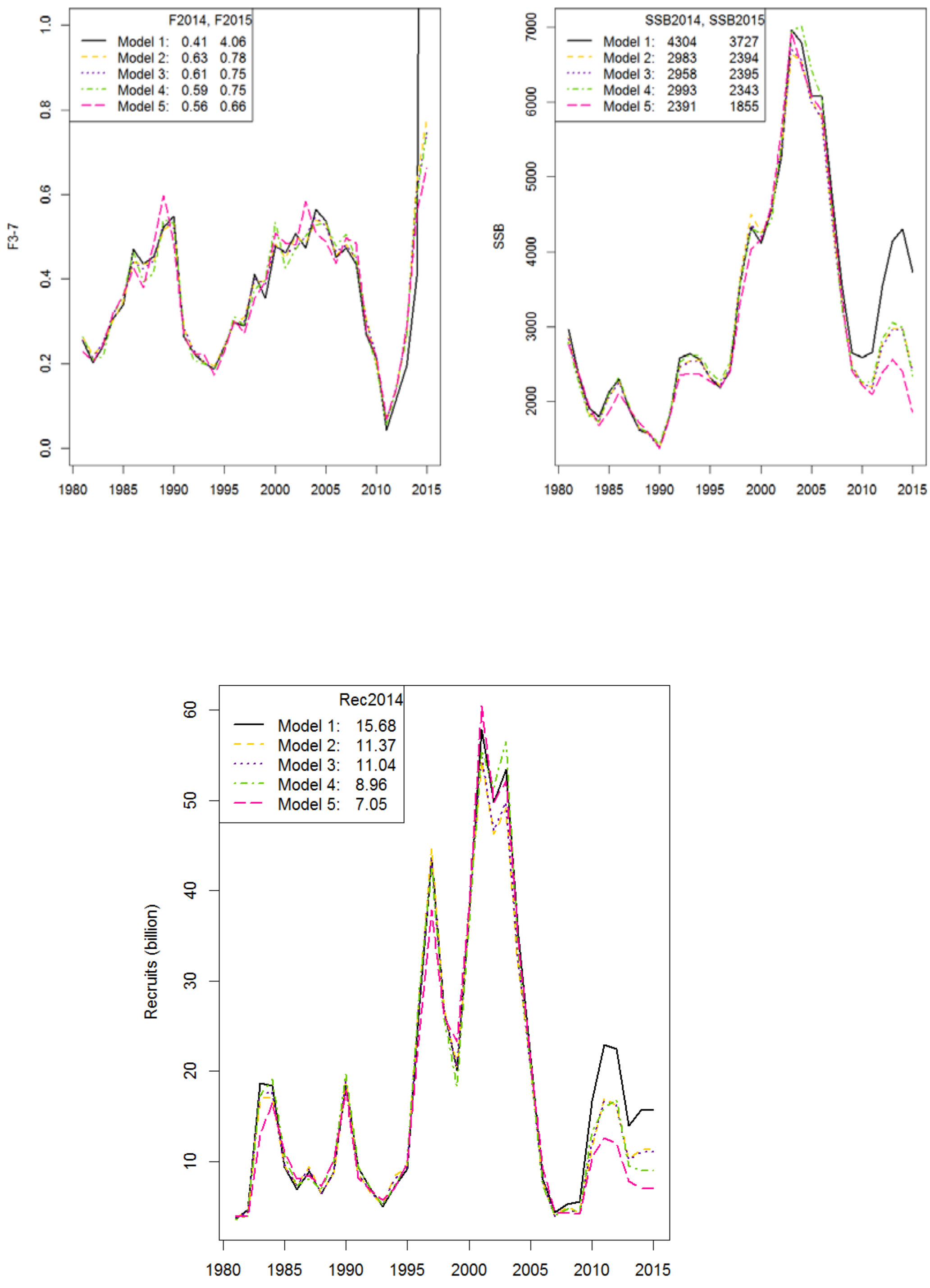


Figure 3.2.1. F_{bar}, SSB and recruitment as estimated by Model 1 to 5 (final configurations).

The "year effects" were considered to be the "too high" IBWSSS indices from 2013 and 2014 followed by the "too low" indices from 2015. Model 3 (configuration 3c) was considered to handle these "year effects" best and was therefore recommended to be used in the assessment of blue whiting (whb-comb).

3.3 Recruitment

Data from the IBWSS survey are used by the stock assessment model, while recruitment indices from several other surveys (Figure 3.3.1) are used qualitatively to adjust the most recent recruitment estimate by the assessment model and to guide the recruitments used in forecast (ICES, 2015).

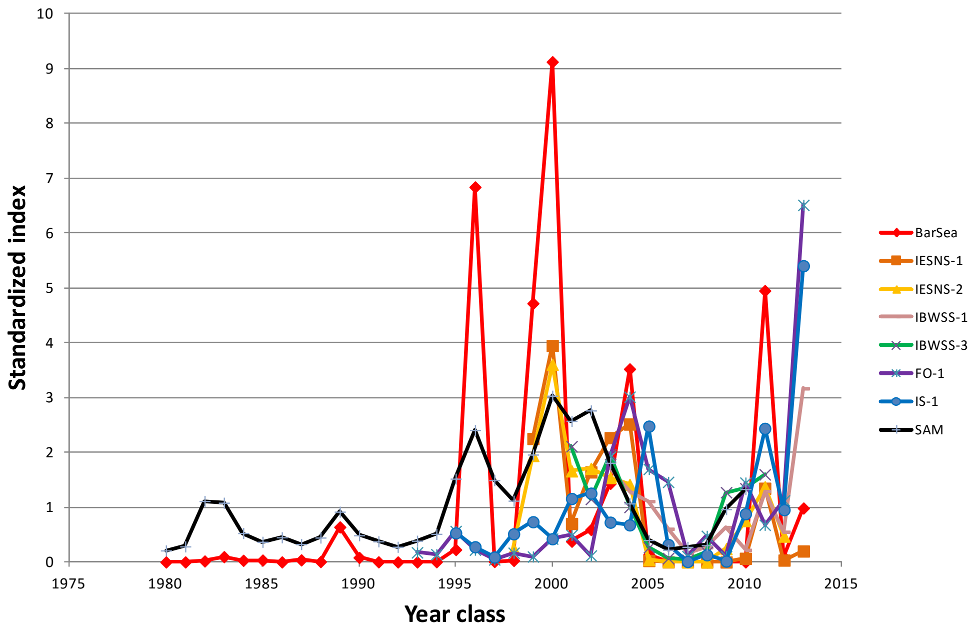


Figure 3.3.1. Blue whiting young fish indices from five different surveys and recruitment index from the assessment, standardized by dividing each series by their mean. BarSea - Norwegian bottom-trawl survey in the Barents Sea, IESNS: International Ecosystem Survey in the Nordic Seas in May (1 and 2 is the age groups), IBWSS: International Blue Whiting Spawning–Stock survey (1 and 3 is the age groups), FO: the Faroese bottom-trawl surveys in spring, IS: the Icelandic bottom-trawl survey in spring, SAM: recruits from the assessment (ICES, 2015).

Time-series with recruitment indicators were available from the following surveys:

- The International Ecosystem Survey in the Nordic Seas (IESNS) only partially covers the known distribution of recruitment from this stock.
- The International Blue Whiting Spawning–Stock Survey (IBWSS) is not designed to give a representative estimate of immature blue whiting. However, the 1-group indices appear to be fairly consistent with corresponding indices from older ages.
- The Norwegian bottom-trawl survey in the Barents Sea (BS-NoRu-Q1(Btr)) in February–March 2015, showed that 1-group blue whiting was present and the index was the third highest in the time-series. This index should be used as a presence/absence index, in the way that when blue whiting is

present in the Barents Sea this is usually a sign of a strong year class, as all known strong year classes have been strong also in the Barents Sea.

- The Icelandic bottom-trawl survey (March) has a time-series from 1996 to present. This survey is aimed at demersal species, but blue whiting juveniles are caught as bycatch. Some signals in recruitment are evident in the time-series. The recruitment index of age 1 fish was obtained by a cut-off length at 22 cm.
- The Faroese Plateau spring (March) bottom-trawl survey has a time-series from 1994 to present. While this survey is not specifically aimed at blue whiting, nor has it been used in any assessments, there are some signals in recruitment evident in the time-series. An index (number per trawl hour) was created based on a length split at 22 cm as an estimate of the abundance of age 1 blue whiting.
- The Portuguese bottom-trawl survey (October–November) has a time-series from 1990 to present. This survey is aimed at demersal species and blue whiting juveniles are caught as bycatch. There are some signals in recruitment evident in the time-series. The recruitment index of age 1 fish was obtained by a cut-off length at 20 cm.

During the IBPBLW process, basic data exploration of the time-series that have previously been used as recruitment indicators in the assessment of this stock (Figure 3.3.1) was prepared (WD 2 Mark Payne and WD 3 Höskuldur Björnsson).

The recruitment indicators show substantial variation in their ability to represent the variations in recruitment, and some of the previous used indicators are not at all correlated with recruitment. There is some evidence however, that the relationships between recruitment and the surveys are non-stationary, and this may need to be accounted for in future modelling endeavours. Strong autocorrelation in the recruitment time-series is present and can potentially be used to our advantage in a forecasting context (see WD 2 Mark Payne).

The indices from the IBWSS age 1 and 2, and the IESNS age 1 and 2 shown a good correlation with the recruitment as estimated by the SAM assessment (WD 2 Mark Payne). Similar results were presented in the WD 3 (Höskuldur Björnsson), being suggested a simple sum of the two survey indices, as a combined recruitment time-series.

The addition of the indices from each one of these surveys (IBWSS or IESNS), from these two surveys (IBWSS and IESNS) and the combined time-series survey (IBWSS + IESNS), using Model 3 were investigated (WD 1- Section 1.6 Morten Vinther). The results from the model without recruitment indices (Model 3) and the model with the summed IBWSS and IESNS were similar (3.3.2).

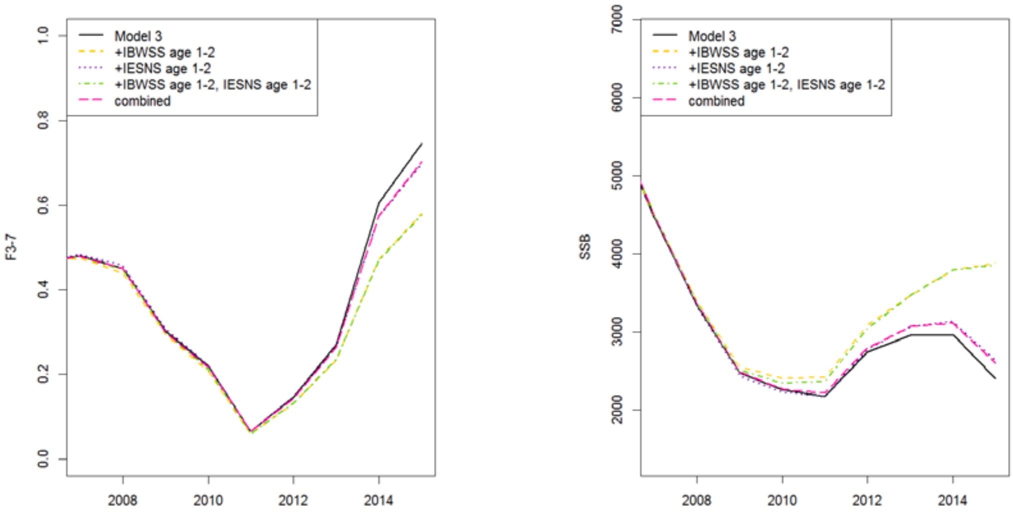


Figure 3.3.2. F_{bar} and SSB as estimated by SAM model with correlated observations and catch data 1983–2014 (Model 3), with additional survey data for estimation of recruits.

The IBWSS indices increases the recruitment indices compared to the model without these indices. F becomes lower and SSB higher in the model with IBWSS indices. This might be because the extra age classes in the IBWSS increases the number of observations to fit and thereby put less weight on the IBWSS age 3–8 indices and maybe because the two youngest age classes constrain the estimate of the older ages in the IBWSS survey. The use of IBWSS age 1 and age 2 indices gives a good fit and can be used to estimate recruitment. The use of the indices also improves the retrospective pattern.

Based on this exercise the IBPBLW concluded that the IBWSS ages 1 and 2 are likely the most suitable recruitment indices. More work is however needed before it can be decided. The work will be done before/at WGWIDE in August 2016.

3.4 Including Q1 catch-at-age data from the assessment year

For blue whiting almost all catches are made in the first half of the year (Figure 3.4.1), around 90% of the annual catches since 2007. In face of this, the WGWIDE proposed as another possible solution for a better evaluation of the survey results is to use the catch data from the current year (ICES, 2015).

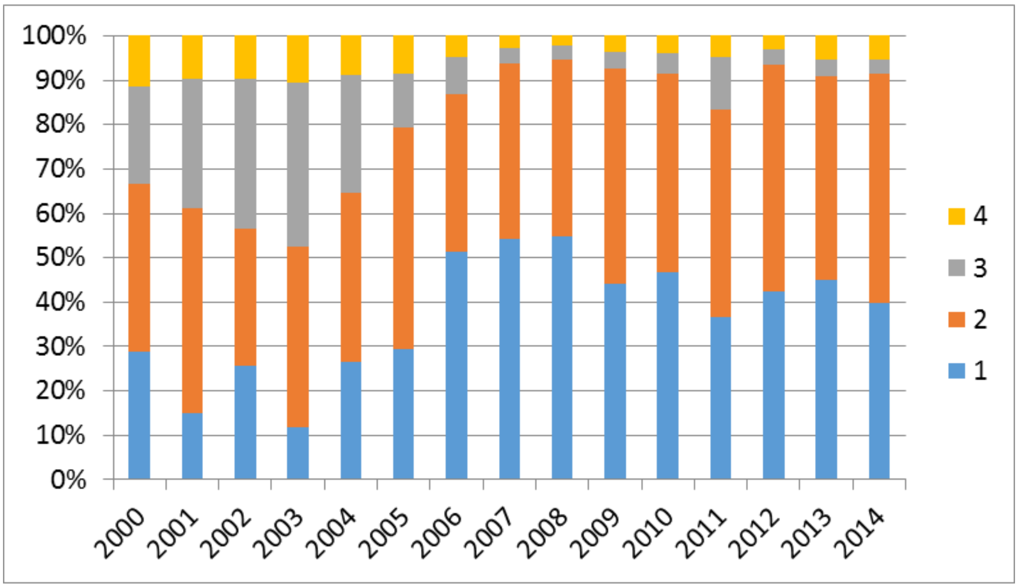


Figure 3.4.1. Distribution of ICES estimates of blue whiting by quarter.

It was therefore suggested that catches during the beginning of the assessment year (either Q1 or Q1+Q2) could be used to predict the annual catches in this year. By incorporating these raised catches in the catch-at-age matrix used for the assessment, the model would have an additional source of information to confront with the most recent survey index, which might result in terminal year estimates being less sensitive to the year effects in the survey.

The results of raising preliminary annual catches from Q1 or Q1+Q2 catches and the impact of using these raised catches on the assessment are described in the Sections 3.4.1 and 3.4.2. Detailed documentation is in WD 4 (Thomas Brunel and Morten Vinther).

3.4.1 Estimating preliminary annual catches from Q1 or Q1+Q2 data

The use of catches during the beginning of the assessment year (either Q1 or Q1+Q2) to predict the annual catches in this year was investigated.

The results revealed that the proportion of catches occurring in Q1 (Figure 3.4.1.1.) appears to be too variable to be used as a basis for raising the Q1 catches to the whole year. The proportion in Q1+Q2 (Figure 3.4.1.2.), however, is much less variable, especially for ages 3 and older, which represent the bulk of the catches.

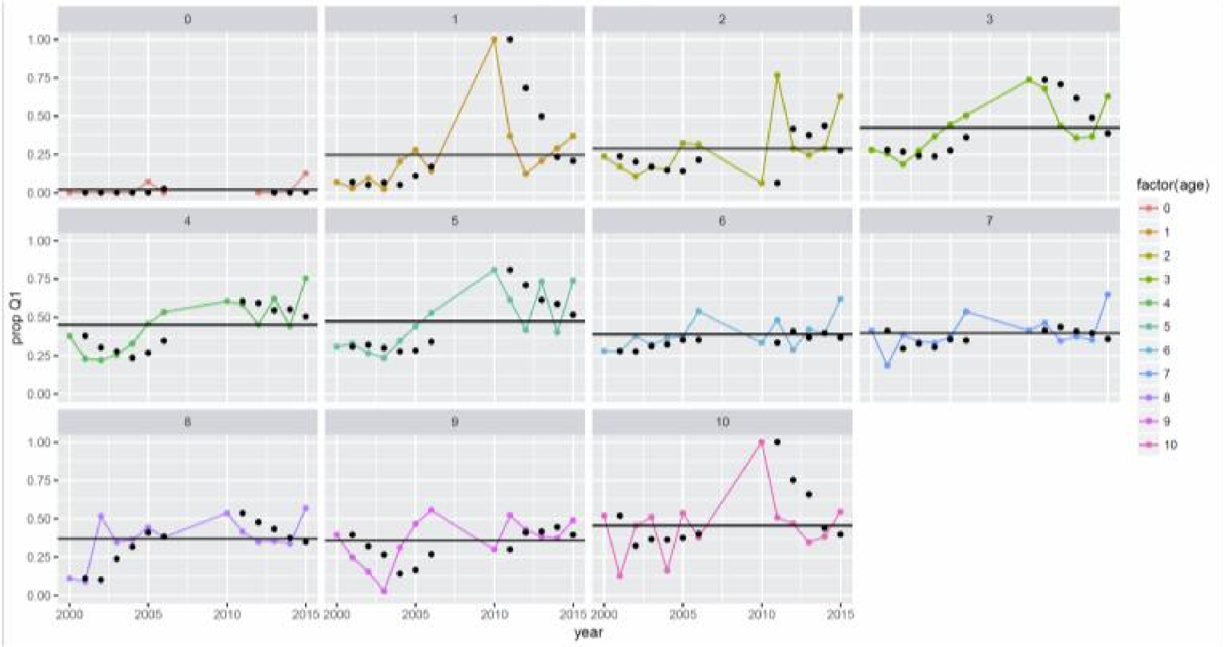


Figure 3.4.1.1. Proportion of the catches-at-age occurring in Q1 (colour lines with dots) and predicted proportion for the coming year based on the mean of the last three years (black dots) and overall mean (horizontal line).

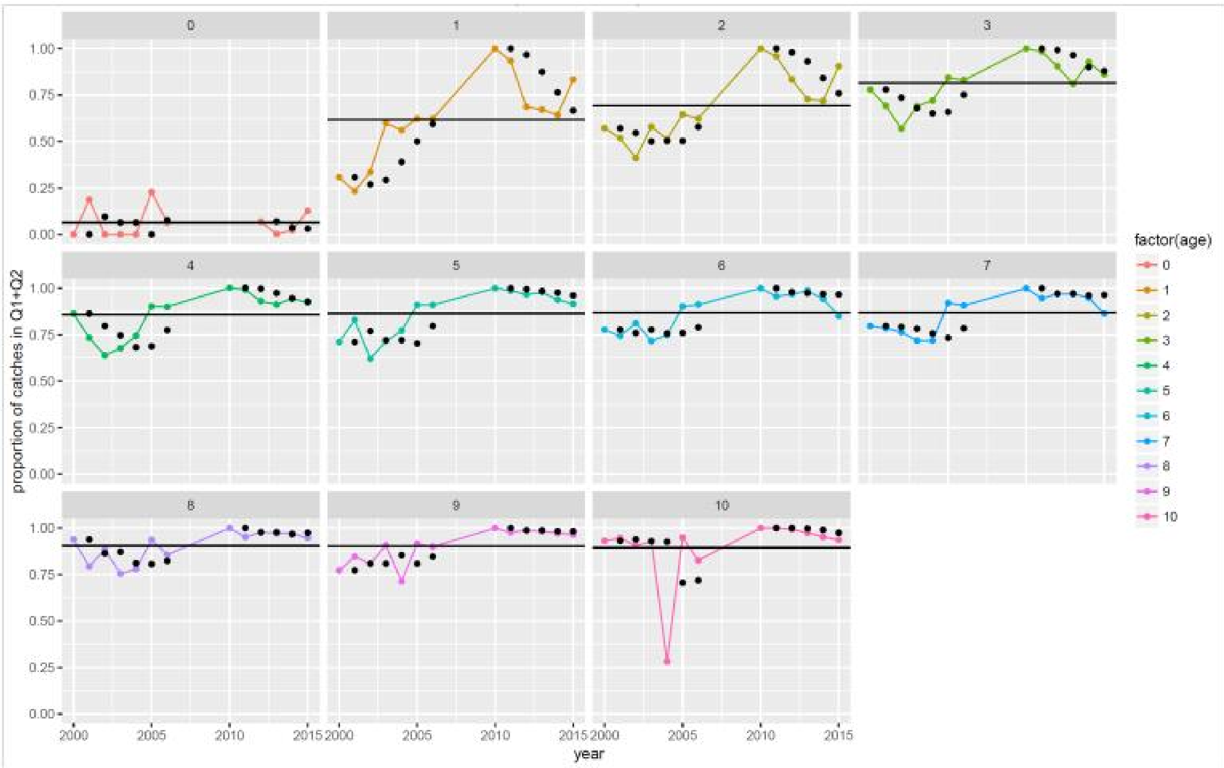


Figure 3.4.1.2. Proportion of the catches-at-age occurring in Q1+Q2 (colour lines with dots) and predicted proportion for the coming year based on the mean of the last three years (black dots) and overall mean (horizontal line).

This exercise showed that the proportion of fish caught in Q1 showed some substantial changes over time. This indicates that the input data "proportion of fishing mortality occurring before spawning", which is currently a constant value across years and across ages, should be made variable and updated regularly.

3.4.2 Improving the assessment by using preliminary annual catch data

The average of the proportion caught during Q1 in 2012–2014 was used as an estimate for 2015, the catches from Q1 in 2015 were raised to the whole year, and added to the catch-at-age data used in the WGWIDE 2015 assessment (which ended in 2014). The catch-at-age matrices corresponding to the 2014 and 2013 assessment were updated by the addition of the in-year catches raised in the same way. Finally, the same catch-at-age matrix was also produced using the proportion caught in Q1+Q2 to raise the in-year catches instead of Q1 alone.

The SAM assessment model for blue whiting was then run for each of these six new catch matrices, in addition to the original 2013, 2014 and 2015 WGWIDE assessments.

The impact of using the preliminary catch data on the assessment results, was investigated by making a series of assessments (using Model 3c configuration, see WD 1 Morten Vinther). Each assessment was extended by one year using the preliminary catch-at-age number (WD 4 Thomas Brunel and Morten Vinther).

The assessment results showed to be highly variable from one year to the next (strong retrospective noise) such that a stable value of F or SSB is only obtained for the period 3–4 years before the last assessment year (Tables 3.4.2.1 and 3.4.2.2).

Table 3.4.2.1. F_{bar} by assessment run and year (upper part) and F in assessment year relative to assessment year+1 (lower part).

RUN	2009	2010	2011	2012	2013	2014	2015
2014_p2015Q1	0.291	0.206	0.06	0.13	0.23	0.463	0.675
2014	0.302	0.219	0.065	0.146	0.270	0.606	0.746
2013_p2014Q1	0.263	0.180	0.052	0.104	0.166	0.278	
2013_p2014Q12	0.264	0.180	0.052	0.104	0.166	0.274	
2013	0.266	0.185	0.055	0.112	0.178	0.178	
2012_p2013Q1	0.283	0.196	0.062	0.119	0.175		
2012_p2013Q12	0.283	0.196	0.062	0.119	0.175		
2012	0.275	0.192	0.059	0.111	0.108		
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	
2013_p2014Q1	87%	82%	80%	71%	61%	46%	
2013_p2014Q12	87%	82%	80%	71%	61%	45%	
2013	88%	84%	85%	77%	66%	29%	
2012_p2013Q1	106%	106%	113%	106%	98%		
2012_p2013Q12	106%	106%	113%	106%	98%		
2012	103%	104%	107%	99%	61%		

Table 3.4.2.2. SSB (1000 tonnes) by assessment run and year (upper part) and F in assessment year relative to assessment year+1 (lower part).

Run	2009	2010	2011	2012	2013	2014	2015
2014_p2015Q1	2548	2365	2333	2999	3315	3468	3199
2014	2469	2253	2173	2741	2955	2958	2395
2013_p2014Q1	2806	2713	2863	3917	4780	5494	
2013_p2014Q12	2797	2703	2855	3908	4776	5490	
2013	2775	2662	2778	3748	4502	5007	
2012_p2013Q1	2660	2510	2566	3382	3748		
2012_p2013Q12	2660	2510	2566	3382	3748		
2012	2668	2536	2634	3586	4342		
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	
2013_p2014Q1	114%	120%	132%	143%	162%	186%	
2013_p2014Q12	113%	120%	131%	143%	162%	186%	
2013	112%	118%	128%	137%	152%	169%	
2012_p2013Q1	96%	94%	92%	90%	83%		
2012_p2013Q12	96%	94%	92%	90%	83%		
2012	96%	95%	95%	96%	96%		

In conclusion, the addition of a new survey year seems to influence the results more than addition of additional catch-at-age data. Preliminary catches do not seem to help much in reducing retrospective pattern. However, the preliminary catches (Q1+Q2) give an extra datapoint and a more realistic F in the intermediate year (a "preliminary F"). It was suggested to explore whether this "preliminary F" and recruitment indices derived with the new model 3c (SAM estimate of N) could be used to reactivate the stochastic forecast module of SAM, which it was abandoned few years ago and instead a deterministic catch constraint short-term program has been used. This will be explored during next WGWIDE in August 2016.

3.5 MSY reference points

The reference points for this stock were last evaluated by ICES in 2013 in connection with a NEAFC request on harvest control rule evaluation (ICES, 2013), and are listed in Table 3.5.1. ToR d) of IBPBLW was to estimate Biological Reference Points. However, ICES is currently in the process of evaluating the management plan for blue whiting. Our previous experience has shown that the estimation methods commonly used often show unwanted instability in their reference point estimates, where for example an annual update of assessment can have unexpected consequences, particularly in cases such as blue whiting that has shown relatively large retrospective patterns in the assessment. Moreover, reference point evaluations should preferably be tightly connected to the MSE, because harvest control rules and reference points inherently go together. IBPBLW therefore decided not to provide an answer to ToR d), but this will be answered in connection with the ongoing management strategy evaluation.

Table 3.5.1. Blue whiting in Subareas I–IX, XII, and XIV. Reference points, values, and their technical basis.

Framework	Reference point	Value	Technical basis	Source
MSY approach	MSY $B_{trigger}$	2.25 million t	B_{pa}	ICES (2013)
	F_{MSY}	0.30	Equilibrium stochastic simulations	ICES (2013)
Precautionary approach	B_{lim}	1.50 million t	Approximately B_{loss}	ICES (2013)
	B_{pa}	2.25 million t	B_{lim} exp(1.645 × σ), with σ = 0.25.	ICES (2013)
	F_{lim}	0.48	Equilibrium stochastic simulations	ICES (2013)
	F_{pa}	0.32	Based on F_{lim} and assessment uncertainties	ICES (2013)

3.6 References

Berg, C. W., and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. – ICES Journal of Marine Science, doi: 10.1093/icesjms/fsw046.

ICES. 2012. Report of the Benchmark Workshop on Pelagic Stocks (WKPELA 2012), 13–17 February 2012, Copenhagen, Denmark. ICES CM 2012/ACOM:47. 572 pp.

ICES. 2013. NEAFC request to ICES to evaluate the harvest control rule element of the long-term management plan for blue whiting. Special request, Advice May 2013. *In* Report of the ICES Advisory Committee, 2013. ICES Advice 2013, Book 9, Section 9.3.3.1.

ICES. 2015. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 25 August–31 August 2015, Pasaia, Spain. ICES CM 2015/ACOM:15. 588 pp.

4 Conclusions

The main outcomes of the inter-benchmark protocol for blue whiting (combined stock) were:

- The correlated SAM version, model configuration 3c, is the appropriate method for the assessment of blue whiting (whb-comb).
- The recruitment indices from IBWSS age 1 and 2 give a good fit to the model and should be considered to be used to estimate recruitment; however, more work is needed.
- The preliminary catches from Q1 and Q2 should be included in the assessment year. They don't diminish the retrospective pattern, but provide a more realistic F in the intermediate year, which possibly can be used in the short-term prediction.

5 The reviewer comments

These are the comments of the independent reviewer Richard Methot:

I participated in most of the Skype calls with the working group and have reviewed the work conducted. My comments on the adequacy with which each of the four Terms of Reference has been addressed is presented below.

ToR a): "evaluate the robustness of the SAM model in situations with clear "year effects" in survey indices as observed in the IBWSS 2015, test appropriate model modifications and /or make criteria for posterior discarding of survey indices;"

A substantial investigation of the impact of the rapid changes in the IBWSS survey on the performance of the SAM model was undertaken, principally by Morten Vinther. The survey was high in 2013 and 2014, then declined to a much lower level in 2015. The model's response to this rapid change was to interpret the 2015 catches as causing a very high F in 2015. The investigation first determined that there was no anomaly in the survey protocol that invalidates the survey result. A second investigation determined that the survey timing (e.g. fraction of catch occurring before the survey) had a substantial effect in that the 2015 low survey observation is fit better with a late survey period. A future model configuration that allows for year-specific survey timing is recommended. Subsequent work used a new, published version of SAM that incorporates correlated observations found that this model provided. The configuration "3" of this model was concluded to provide the best fit to the data. It also achieved a better balance in fit between the survey data and the catches and resulted in a lower F for 2015. This model did not resolve the pattern in the model residuals, but it did provide a more consistent explanation of the pattern through correlated observations. Nevertheless, the existence of residuals at the end of the time-series seems to indicate that data from 2015 and 2016 will have much importance in future model runs.

ToR b): "estimate recruitment for short-term forecast from the present available survey indices;"

The investigations of correlations among the possible recruitment indices and between the indices and model results by Mark Payne and Höskuldur Björnsson finds that several indices have reasonable correlations with recruitment and the model configuration that uses the age 1 and age 2 IBWSS indices seem suitable for recruitment estimation. However, the weaker correlation with the IENSS survey seems mostly due to the low survey observations for 2007 and 2008 year classes. These surveys indicate high recruitment in 2014 and 2015. However, the spatially disjointed aspect of the surveys is disconcerting. The WG considered adding surveys together to provide a combined index that collectively provided more complete spatial coverage. This was not found to be superior and is not recommended at this time. However, further investigation of the spatial aspects of stock distribution, fisheries, and survey coverage seems warranted to provide a more realistic characterization of this population.

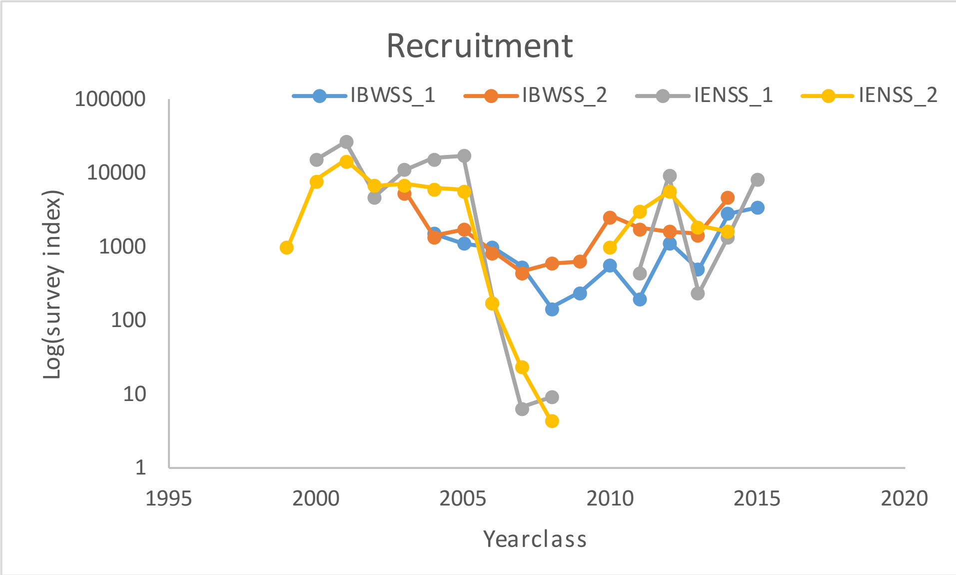


Figure 5.1. Blue whiting young fish log(index) from two different surveys, for ages 1 and 2. IBWSS: International Blue Whiting Spawning–Stock survey; IESNS: International Ecosystem Survey in the Nordic Seas in May.

ToR c): "retrospectively examine the impact of including Q1 catches in the assessment year;"

Examination of the proportion of the catch early in the year by Morten Vinther and Thomas Brunel showed it to be too variable to be used as a robust indicator of the catch for the entire year. However, the variability among years and the fact that younger age groups show up later in the year is a good indication that further model improvements could be made by switching to a quarterly model configuration. This would especially allow for inclusion of early season catches without having to extrapolate for the whole year. At minimum, the proportion of annual catch that occurs before the survey should be age and year specific.

ToR d): "estimate Biological Reference Points."

This investigation was deferred. It will better be conducted as part of the Management Strategy Evaluation that is underway.

5.1 General comments

The members of the working group shared the work well and were very helpful and proactive. The chair kept the conference calls focused on topics being investigated.

It was challenging to meet only by phone/video. At least one face-to-face meeting would much improve common understanding of the investigations being undertaken.

Annex 1: List of participants

NAME	ADDRESS	PHONE/FAX	E-MAIL
Anna Olafsdottir	Marine Research Institute Skùlagata 4 Reykjavik 121 Iceland	+354 5752000	anna@hafro.is
Asta Gudmundsdóttir	Marine Research Institute Skùlagata 4 Reykjavik 121 Iceland	+354 5752000	asta@hafro.is
Casper Berg Invited Speaker	DTU Aqua - National Institute of Aquatic Resources Section for Fisheries Advice Charlottenlund Slot Jægersborg Alle 1 2920 Charlottenlund Denmark		cbe@aqua.dtu.dk
Claus Reedtz Sparrevohn	Danish Pelagic Producers Organisation H.C. Andersens Boulevard 37 1553 Copenhagen V. Denmark	+45 29124549	crs@pelagisk.dk
David Miller ICES Secretariat	International Council for the Exploration of the Sea H.C. Andersens Blvd. 44–46 1553 Copenhagen V Denmark		david.miller@ices.dk
Höskuldur Björnsson	Marine Research Institute Skùlagata 4 Reykjavik 121 Iceland	+354 5752000	hoski@hafro.is
Jan Arge Jacobsen	Faroe Marine Research Institute PO Box 3051 110 Tórshavn Faroe Islands	+298 353900	janarge@hav.fo
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

Name	Address	Phone/Fax	E-mail
Martin Pastoors	Pelagic Freezer-Trawler Association PO Box 72 2280 AB Rijswijk Netherlands	+31 631901027	mpastoors@pelagicfish.eu
Morten Vinther	DTU Aqua - National Institute of Aquatic Resources Section for Fisheries Advice Charlottenlund Slot Jægersborg Alle 1 2920 Charlottenlund Denmark	Phone +45 3588 3350 Fax +45 3588 3333	mv@aqua.dtu.dk
Nikolay Timoshenko	Atlantic Research Institute of Marine Fisheries and Oceanography 5 Dmitry Donskogo Street Kaliningrad 236000 Russia	+7 4012 925554	timoshenko@atlantniro.ru
Patrícia Gonçalves Chair	Instituto Português do Mar e da Atmosfera Av. Brasilia s/n 1449-006 Lisboa Portugal	+351 213027000	patricia@ipma.pt
Richard Methot Invited Expert	National Marine Fisheries Service NOAA Seattle WA 98112 USA		richard.methot@noaa.gov
Søren Anker Pedersen	Marine Ingredients Denmark H.C. Andersens Blvd 37, 1th Denmark	+45 33145800	sap@maring.org
Thomas Brunel	Wageningen IMARES PO Box 68 1970 AB IJmuiden Netherlands	+31 317 487 161	thomas.brunel@wur.nl

Annex 2: Working Documents

WD 1: Vinther, M. 2016. Blue Whiting, ToR A) Appropriate Model Modifications? Working document 1 to ICES. 2016. Report of the Inter-Benchmark Protocol for Blue Whiting (IBPBLW), 10 March–10 May 2016, By correspondence. ICES CM 2016/ACOM:36. 26 pp.

WD 2: Payne, M. 2016. Blue Whiting Recruitment Indicators - Data Exploration. Working document 2 to ICES. 2016. Report of the Inter-Benchmark Protocol for Blue Whiting (IBPBLW), 10 March–10 May 2016, By correspondence. ICES CM 2016/ACOM:36. 26 pp.

WD 3: Björnsson, H. 2016. Blue Whiting Assessment and HCR simulations. (The Icelandic Blue Whiting Commission). Working document 3 to ICES. 2016. Report of the Inter-Benchmark Protocol for Blue Whiting (IBPBLW), 10 March–10 May 2016, By correspondence. ICES CM 2016/ACOM:36. 26 pp.

WD 4: Brunel, T. and Vinther, M. 2016. Blue Whiting, ToR C) Examine the Impact of including Q1 Catches in the Assessment Year? Working document 4 to ICES. 2016. Report of the Inter-Benchmark Protocol for Blue Whiting (IBPBLW), 10 March–10 May 2016, By correspondence. ICES CM 2016/ACOM:36. 26 pp.

The IBPBLW working documents are presented in full in the following pages.

Blue Whiting, TOR a) appropriate model modifications?

WD to the Inter-benchmark for Blue whiting, spring 2016.

Morten Vinther, DTU Aqua

1.1 TOR

a) evaluate the robustness of the SAM model in situations with clear "year effects" in survey indices as observed in the IBWSS 2015, test appropriate model modifications and /or make criteria for posterior discarding of survey indices;
b) estimate recruitment for short-term forecast from the present available survey indices;
c) retrospectively examine the impact of including Q1 catches in the assessment year;
d) estimate Biological Reference Points.

This document is written to document what I have been doing in relation to TOR a) at the WGWIDE 2015 and during the benchmark. There are also some runs using survey indices for estimating recruitment (TOR b).

1.2 TOR a)

1.2.1 Is there a sharply define “breakpoint” in F 2015 due to the IBWSSS index “overall strength”

The presently used SAM configuration estimates an unrealistically high F (3.85) for 2015 based on the IBWSSS survey indices from spring 2015. The last assessment year (last year with catch at age data in the model) is 2014. In previous years, F at age in the year after the last assessment year (intermediate year) were practically set (estimated) to F in the preceding year (random walk).This is the normal result for both blue whiting and other stocks using the SAM model. To investigate the sensitivity of survey indices in 2015 to F2015, a series of SAM assessment was done applying a simple multiplier for all age indices in 2015. This might reflect a general property with an acoustic survey, where a “year effect” in residuals is often seen, probably due to bias in the acoustic measurement (over all stock level) while age samples are unbiased.

Results and conclusion

F in 2015 is highly dependent on the 2015 IBWSSS index (Figure 1) with a gradually decrease in F(2015) with increasing 2015 indices. There is no “flip-flop” where F jumps back to the expected value. Therefore the 2015 indices cannot be discarded, based on the simple criterion, that the SAM model estimates “unreasonably” F values in the intermediate year.

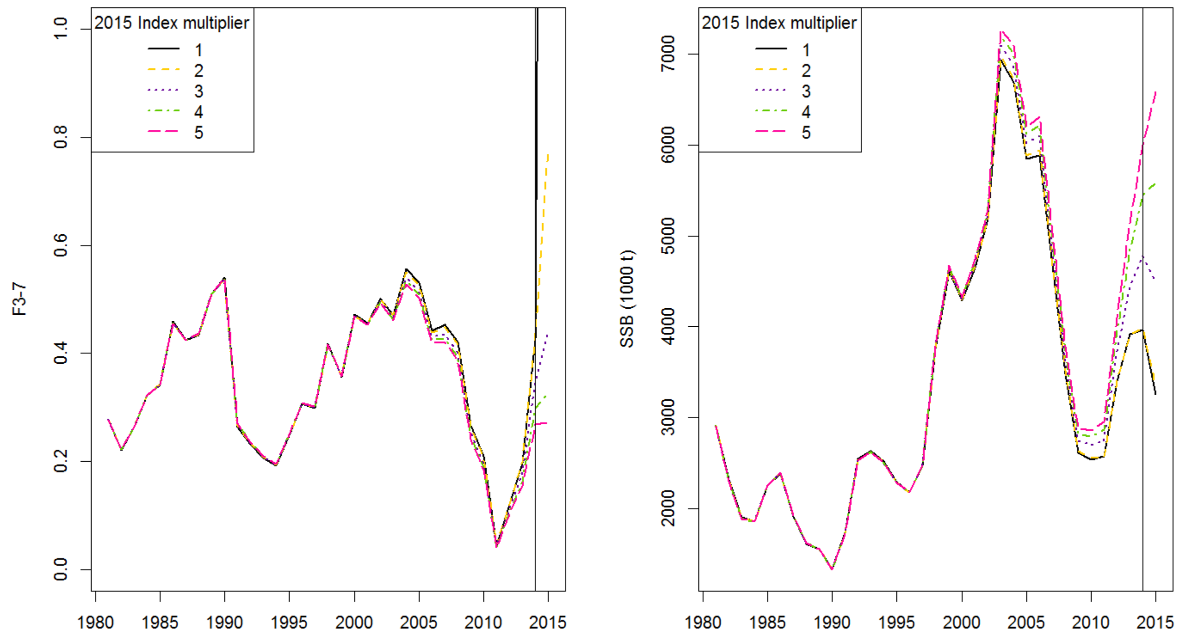


Figure 1. Estimated Fbar and SSB from SAM assessments where the IBWSS 2015 survey indices are multiplied by a factor, 1 to 5.

1.2.2 Alternative SAM settings

During the 2015 WGWIDE several alternative configurations of SAM were tried.

Survey period

The survey period is provided to calculate the average stock numbers within the survey period. It is assumed that the instantaneous mortality rates (F and M) are the same over the full time step (year). At present the survey period reflects the average timing of the survey, in 2015 23/3 to 10/4, equivalent to 0.22 – 0.27 proportion of the year, however this does not correspond to the average proportion of annual M and F before the survey takes place, as up to half of the annual catches are taken in the period January-March (Figure 2). With the present model the survey period must be the same for all survey years. For the survey year range, 2004-2015, between 27% and 53% of the annual yield is taken, which show that a year independent survey period (or more correctly, a proportion of F before the survey takes place) is unsuitable.

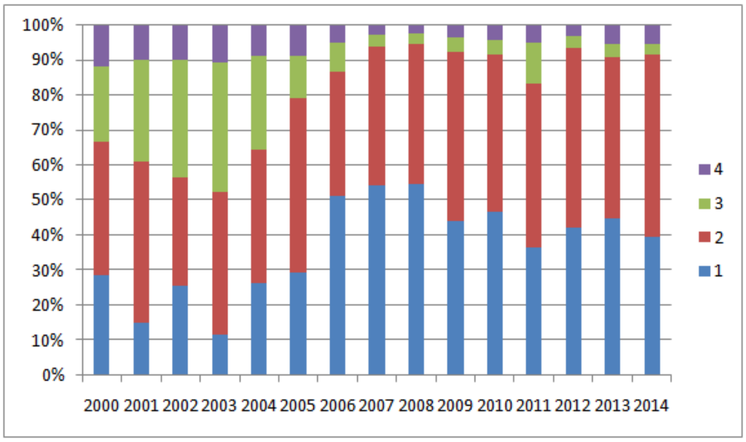


Figure 2. Blue whiting. Distribution of catch of blue whiting by quarter (source WGWIDE 2015)

The model results from the 2015 assessment are highly sensitive to the used period for the survey as shown below for SSB in 2015 (results from WGWIDE 2015 presentation). Even with the latest survey period, F in 2015 is estimated above 2 (or close to, I do not have F from this exact run, but something similar).

survey timing	SSB 2015	neg log-like
0.10-0.20:	1971 kt	262.63
0.22-0.27:	3259 kt	259.93
0.01-0.02	1679 kt	264.40
0.49-0.50:	3375 kt	256.94

This high sensitivity to survey period is not pronounced for the 2014 assessment, as seen below

survey timing	SSB 2014	neg log-like
0.10-0.20:	5472 kt	234.58
0.22-0.27:	5279 kt	234.99
0.01-0.02	5778 kt	234.11
0.49-0.50:	4852 kt	236.34

The model fit (negative log-likelihood) is best with a late survey period in the 2015 assessment, while the best fit is obtained with the early period for the 2014 assessment.

Based in these observations, a quick fix by changing the survey period is not feasibly to get realistic F for 2015. A “survey period” that better reflects the proportion of F before the survey, ideally with a year dependent period, may improve the assessment. Another solution might be smaller model time steps (quarter or half year) to take into account the skewed distribution of fishing mortality over the year. This is a job for Seasonal SAM, however this model is not ready yet, but probably completed before the next WGWIDE.

Back-shifting the IBWSSS

The IBWSS takes place in March-April. By back-shifting (y=y-1, age=age-1) the survey observations to the 31th December, SAM has no observations outside the period for catch at age. Back-shifting does however not take into account the varying F in the beginning of the year (Jan-Mar), before the survey actually takes place.

survey timing	F2014	SSB 2014	neg log-like
0.22-0.27:	0.67	3963 kt	259.82
0.99-1.00, backshifted:	1.18	2395 kt	264.74

A slightly better fit is obtained by the default IBWSSS timing, but there is large difference in F2014 and SSB 2014 for the two models. F2014 in the backshifted model seems too high.

1.3 Updated SAM mode, accounting for correlated observations in an age-based state-space stock assessment model.

This analysis is using the newly developed options for the SAM model, "Accounting for correlated observations in an age-based state-space stock assessment model" by Berg and Nielsen (2016)

The present analysis follows the structure of the case studies presented in Berg and Nielsen (2016). Five models are investigated (see the paper for model details):

1. All observations are independent. No parameters are related to correlation.
2. Regular lattice AR(1) observation correlation structure for all fleets. One correlation parameter per fleet, although surveys may share parameters.
3. Irregular lattice AR(1) observation correlation structure for all fleets. Between 2 and A_f -1 parameters per lattice, and lattice parameters are allowed to be shared among fleets.
4. Unconstrained observation correlation structure for commercial catches and irregular lattice AR(1) observation correlation structure for all surveys.
5. Unconstrained observation correlation structure for all fleets. A_f (A_f -1)/2 parameters per fleet

In addition to the five models, "model 0", the default SAM configuration, is presented for comparison.

Results

1.3.1 Model 0, the default SAM

The residuals from catch at age and survey indices are shown in Figure 3.

1.3.2 Model 1, All observations are independent

Model 1 is the same as the default SAM model, both with an assumption of no correlation of observations as for the default SAM. The default model uses however correlated random walks for the fishing mortalities which is also implemented in Model 1 with two options: 1) compound symmetry and 2) as a AR(1) process. The default Model 0 uses compound symmetry.

The residuals from catch at age and survey indices are shown for Model 1a (compound symmetry) and Model 1b (AR(1)) are shown in Figure 4 and Figure 5. Please note that residuals for model 1-5 are based one-step a head (OSA) prediction errors (see Berg and Nielsen (2016) for details). The two sets of residuals are almost identical, however, the survey "Year effect" for 2013-2015 is more pronounced for the Model 1a.

Model 1a gives exactly the same assessment results as the default SAM model (Figure 7) as it should. Model 1b has a slightly lower F for 2015 and higher F for 2015 compared to Model 1a. All models estimate a high and unrealistic F for 2015. Model diagnostics show that Model 1b fits better than Model 1a. The AR(1) correlation will be used as default for the rest of the analyses, as recommended by Berg and Nielsen (2016).

model	nlogl	nopar	nobs	AIC	AICc
Model 1a	322.3	14	406	672.6	673.6
Model 1b	310.9	14	406	649.9	651.0

The retrospective analysis (Figure 6) for model 1b shows a highly variable SSB estimate, but a constant consistent estimate of F.

1.3.3 Model 2, Regular lattice AR(1) observation correlation structure for catch and survey. One parameter for each.

Residuals are shown in Figure 8. The feature of this options is clearly shown in the correlation plot (Figure 9) where neighbouring ages have the highest correlation.

The retrospective analysis (Figure 10) for model 2 shows a highly variable SSB and F estimate.

1.3.4 Model 3, Irregular lattice AR(1) observation correlation structure for catch and survey.

As a first try, (Model 3a), it is assumed that catch at age correlations are grouped by ages 1-2, ages 3-5 and ages 6-10. IBWSS correlations are grouped by ages 3-4 and ages 5-8. The estimated correlations (Figure 11) show a higher correlation for the 6+ catch at ages. For IBWSS, the correlation seems independent of age.

A test (Model 3b) with only one group for correlation of survey observation gave a practically identical correlation plot and a better fit (see below). A reduction to two correlation groups for catches, ages 1-5 and ages 6-9 (Model 3c) reduced the number of parameter with an almost unchanged likelihood.

model	nlogl	nopar	nobs	AIC	AICc
Model 3a	274.11	19	406	586.2	588.2
Model 3b	274.12	18	406	584.2	586.0
Model 3c	274.12	17	406	582.2	583.8

Model results are practically identical for the three models (Figure 15). Model 3c was chosen as the best configuration for within this model group.

The retrospective analysis (Figure 14) for model 3c shows a highly variable SSB and F estimate.

1.3.5 Model 4, Unconstrained observation correlation for catch and Irregular lattice AR(1) observation correlation structure for survey

As for model 3, nothing is gained by splitting the survey correlation into two groups for Model 4.

model	nlogl	nopar	nobs	AIC	AICc	
Model 4a	251.83	61	406	625.7	647.6	ages 3-4 and ages 5-8
Model 4b	251.83	60	406	623.7	644.9	ages 3-8

Correlations for catches show some negative correlations for age 5 plus (Figure 16). Figure 17 presents residuals. The retrospective analysis (Figure 18) for model 4b shows a highly variable SSB and F estima te.

1.3.6 Model 5, Unconstrained observation correlation for catch and survey

Correlations for catches show some negative correlation for age 5 plus, but survey observations are all positively correlated (Figure 19). Residuals are shown in Figure 20. The retrospective analysis (Figure 2 1Figure 14) for model 5 shows a highly variable SSB and F estimate.

1.3.7 Comparison of results, Model 1– Model 5

The main model diagnostics for the final configurations of Model 1-5 are shown below.

model	nlogl	nopar	nobs	AIC	AICc
Model 1	310.94	14	406	649.9	651.0
Model 2	275.51	16	406	583.0	584.4
Model 3	274.12	17	406	582.2	583.8
Model 4	251.83	60	406	623.7	644.9
Model 5	240.29	74	406	628.6	662.1

Based on the AIC or AICc criterion, Model 3 is the best, followed by Model 2. Model 4 and Model 5 are over-parameterised.

All models except model 1 have a marked retrospective pattern of F (shown for years with catches onl y). Model 1 has however the highest (and most unrealistic) F for 2015. None of the models show a con sisten estimate of SSB.

A comparison of assessment results (Figure 22) shows pretty much the same Fbar and SSB for Models 2, 3 and 4. Model 5 has a lower Fbar and a lower SSB compared to Models 2-4.

Base in these considerations, especially the main model diagnostics, Model 3 is the model is the prefer red model.

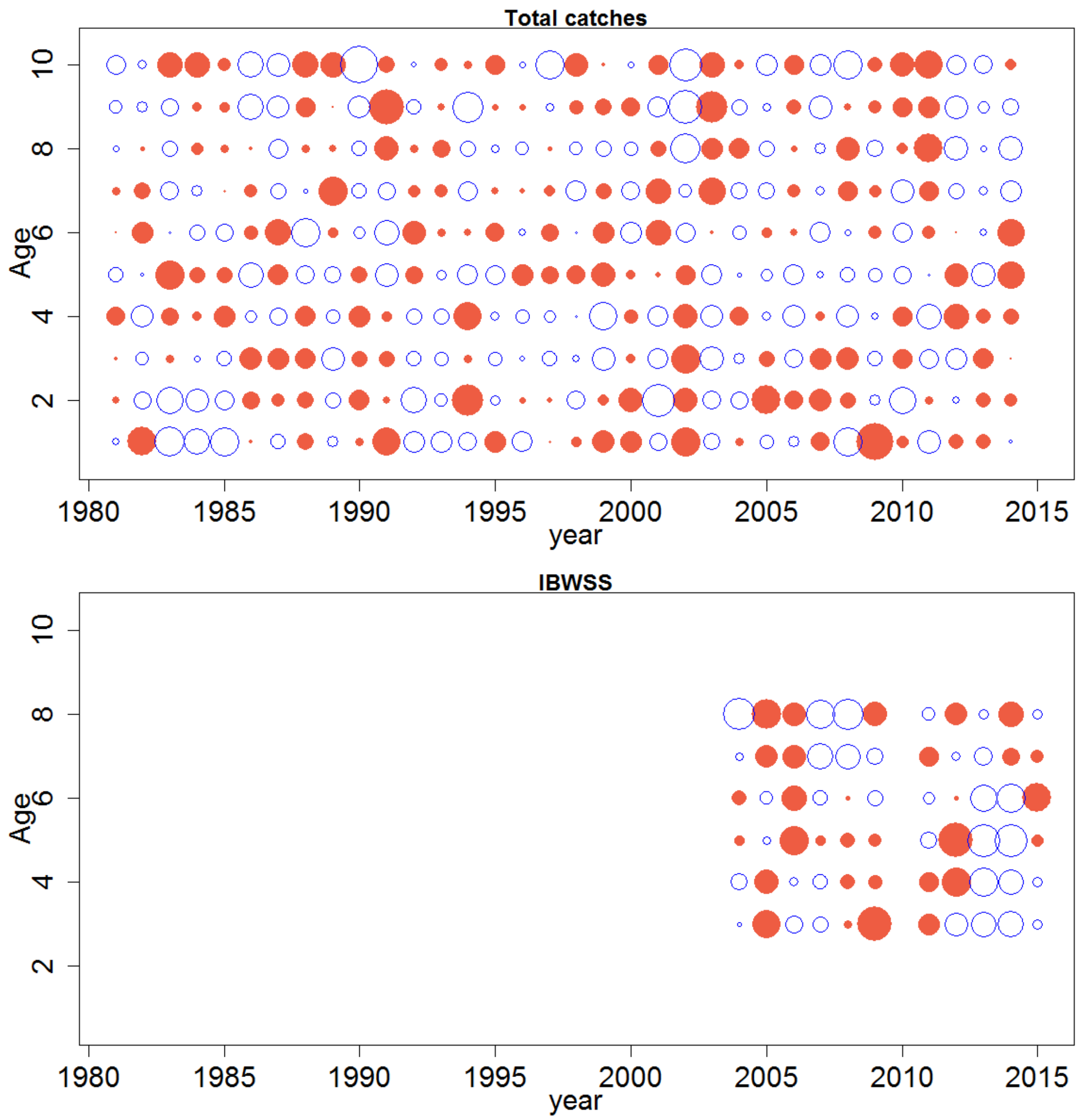


Figure 3. Model 0. Standardized residuals from catch at age and the IBWSS survey. Red (dark) bubbles show that the observed value is less than the expected value.

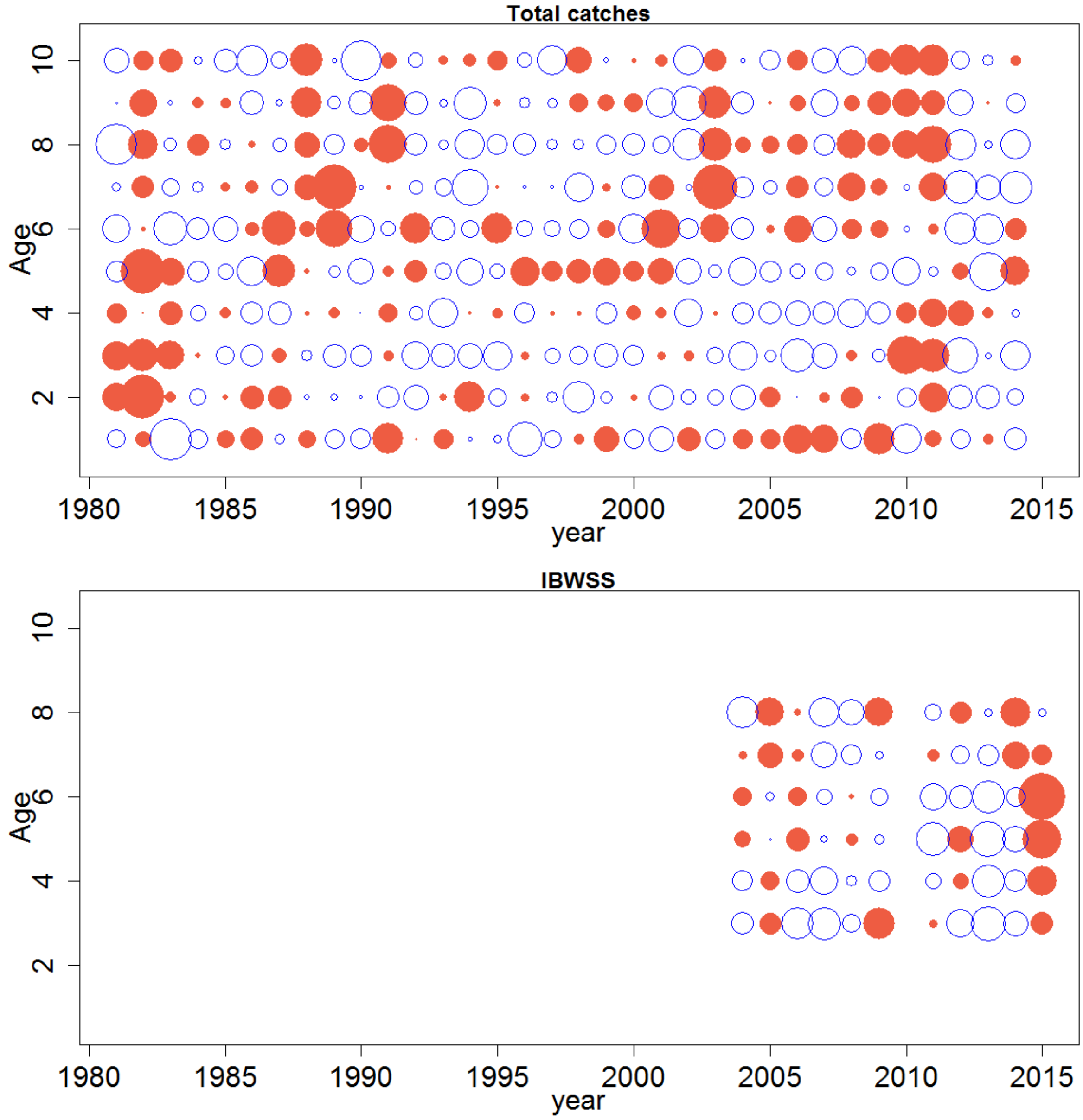
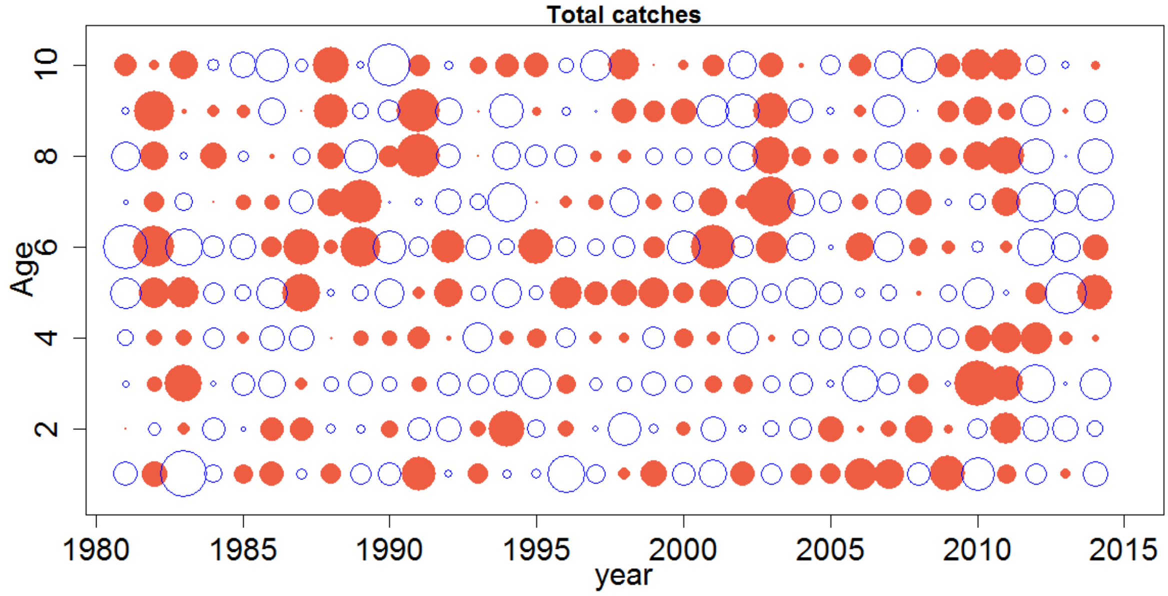


Figure 4. Model 1a. Normalized OSA residuals.



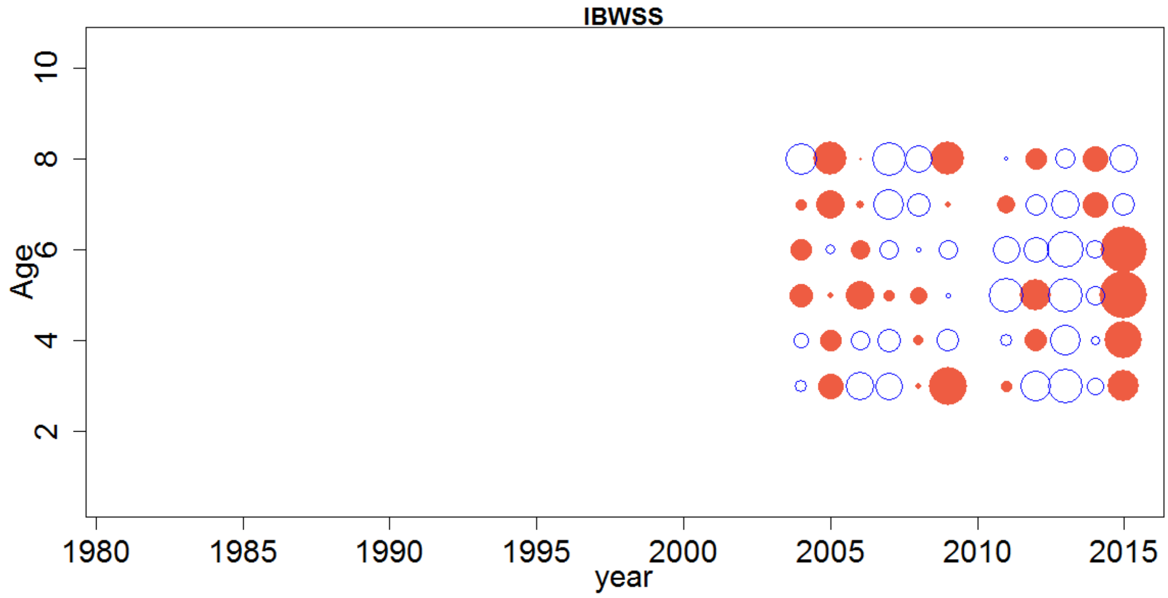


Figure 5. Model 1b. Normalized OSA residuals.

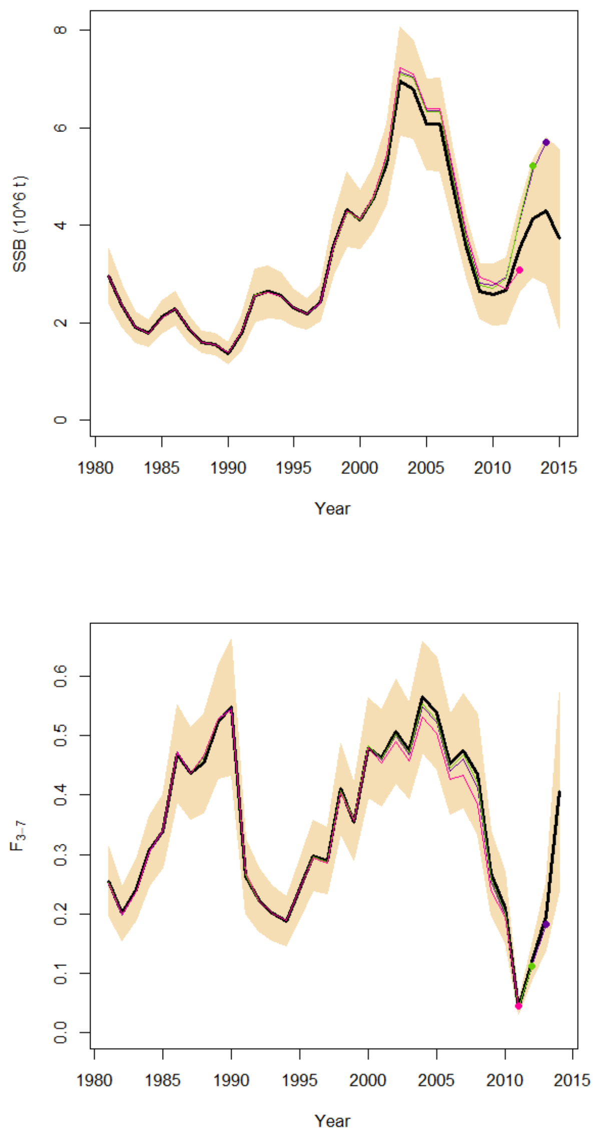


Figure 6. Model 1b. Retrospective analysis.

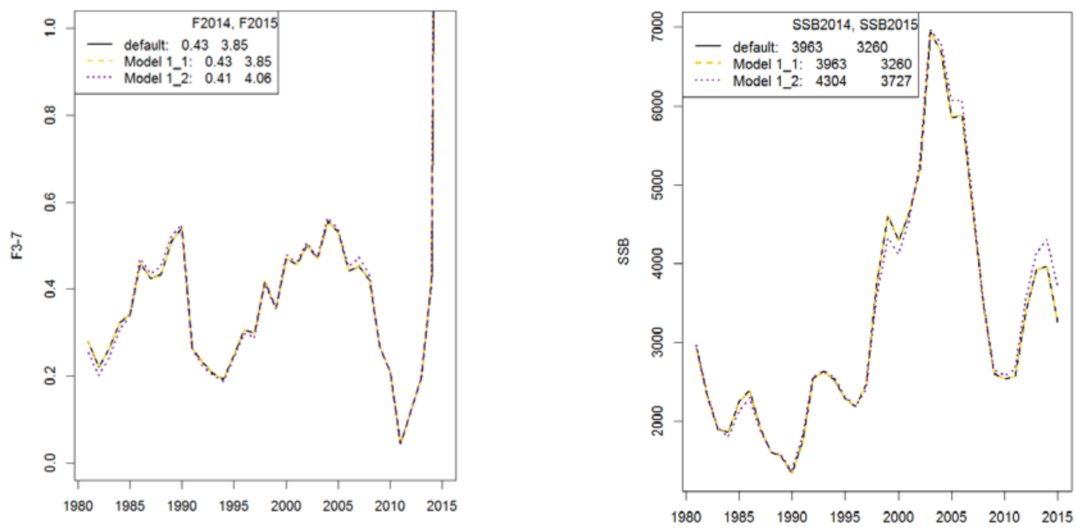


Figure 7. Fbar and SSB as estimated by the default SAM model and the new model implantation (Model 1a and Model 1b, see text for details)

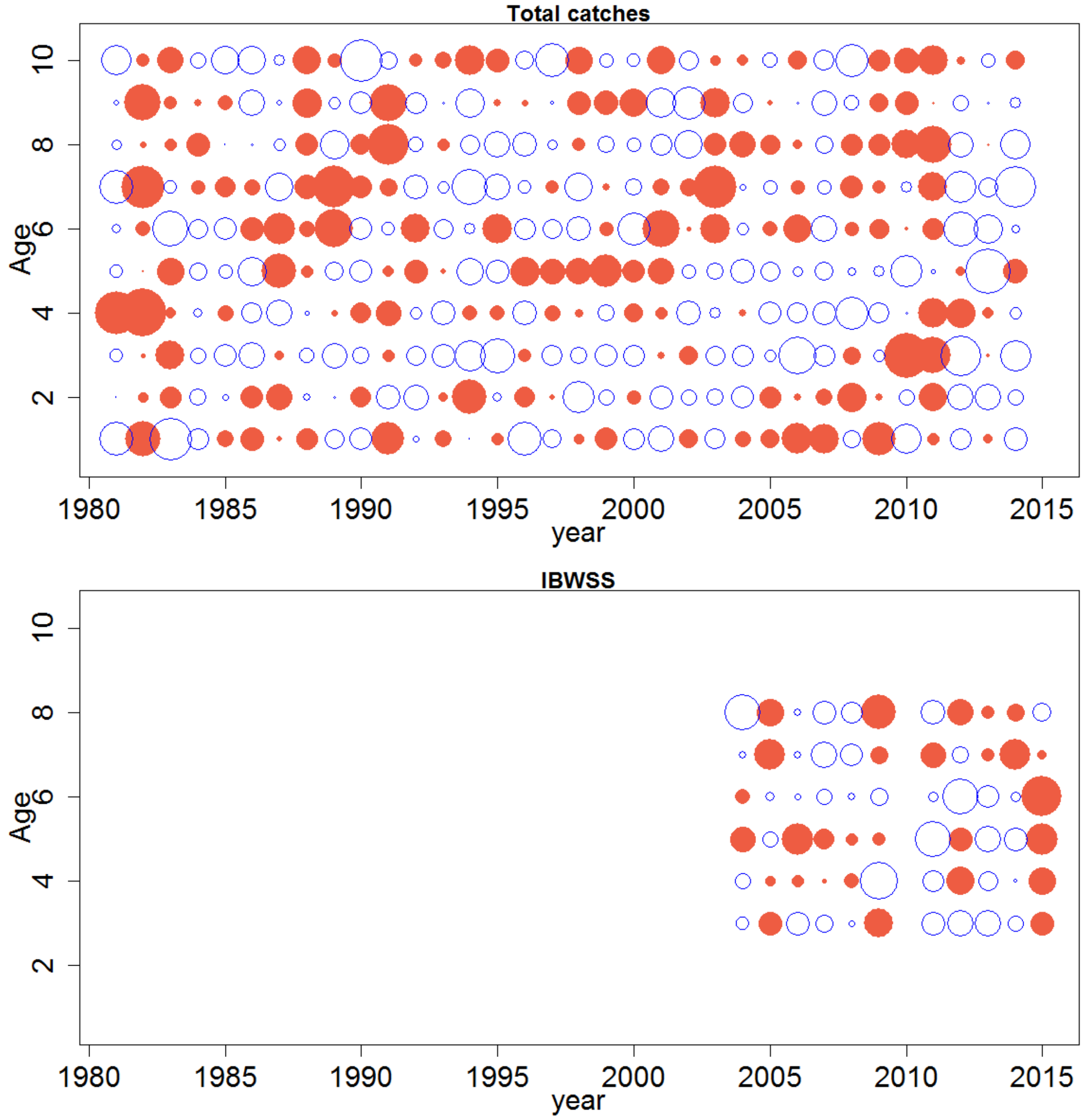


Figure 8. Model 2. Normalized OSA residuals.

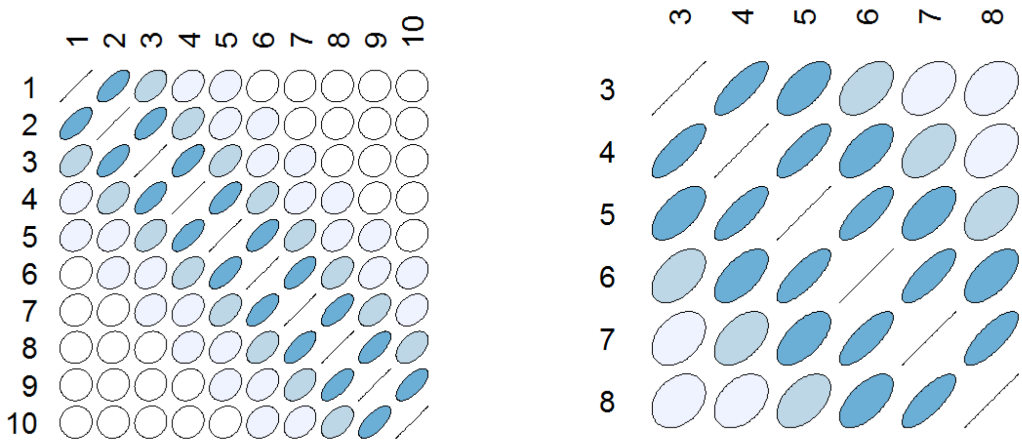


Figure 9. Model 2. Estimated observation correlation structures by catch at age and IBWSSS CPUE. Each ellipse represents correlation. Increasingly darker shading is used for increasingly larger absolute correlations, while uncorrelated pairs of ages are depicted as circles with no shading. Positive correlation in shadings of blue (and negative correlation in shadings of red, not present on this figure).

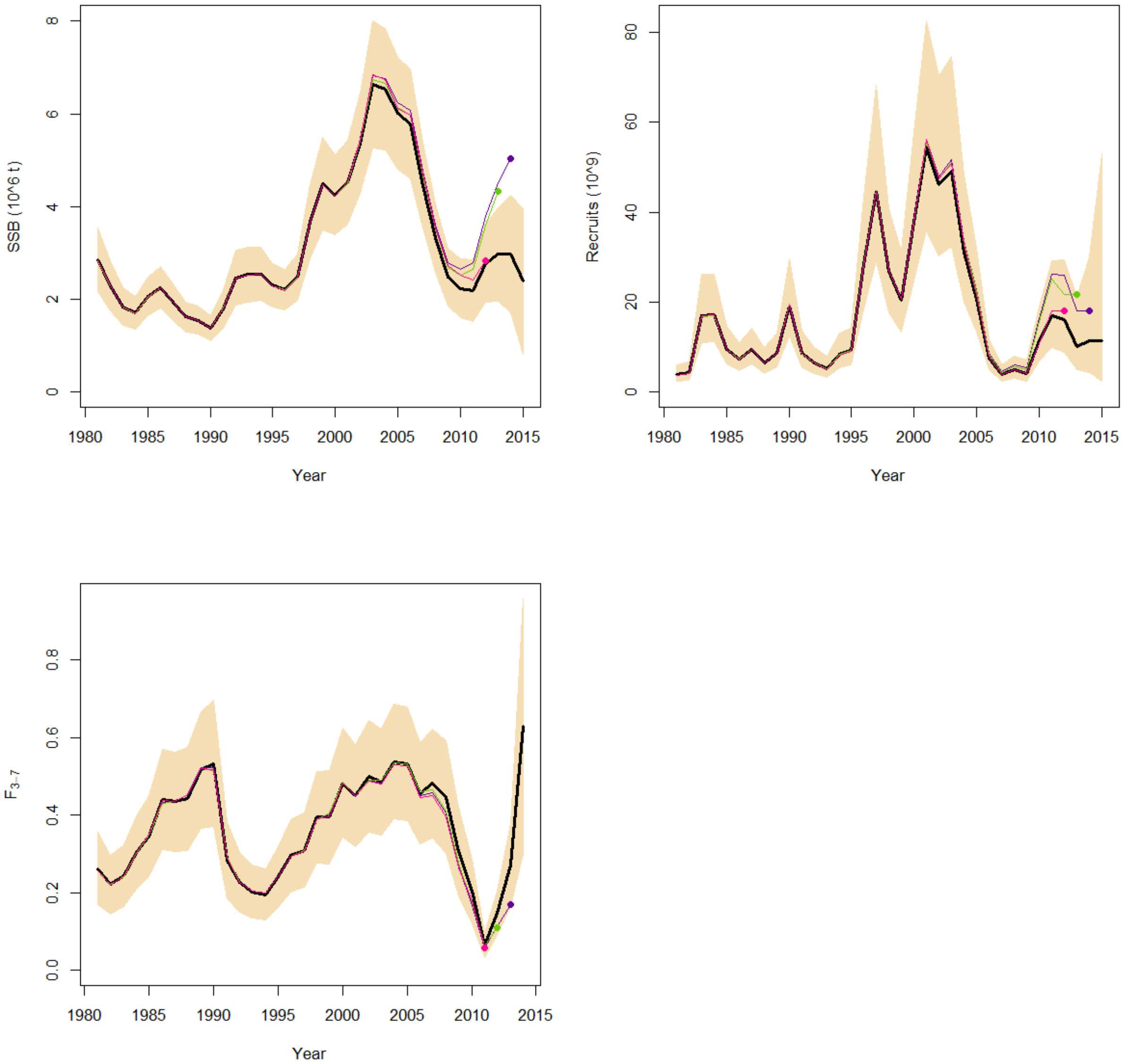


Figure 10. Model 2. Retrospective analysis.

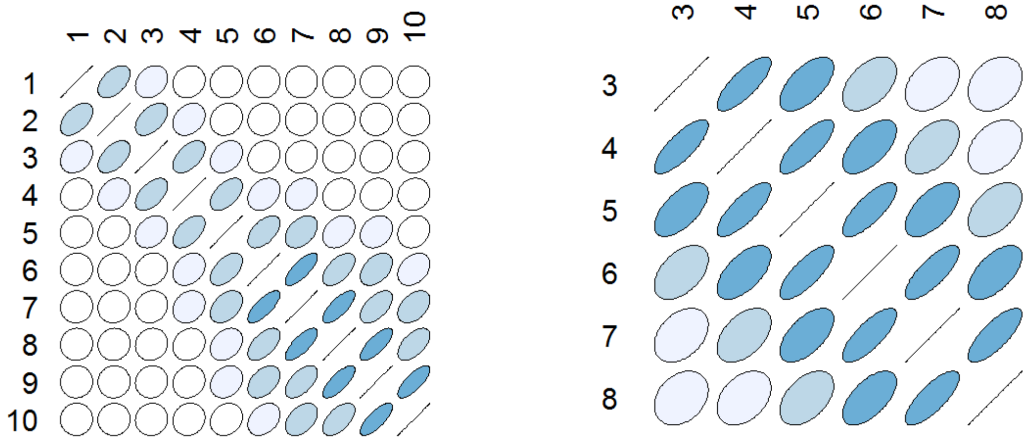


Figure 11. Model 3a. Estimated observation correlation structures by catch at age and IBWSSS CPUE.

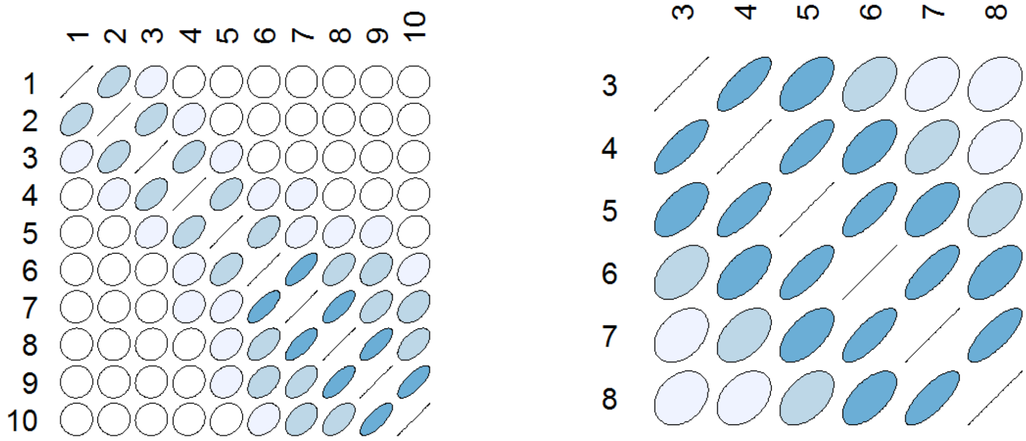


Figure 12. Model 3c. Estimated observation correlation structures by catch at age and IBWSSS CPUE.

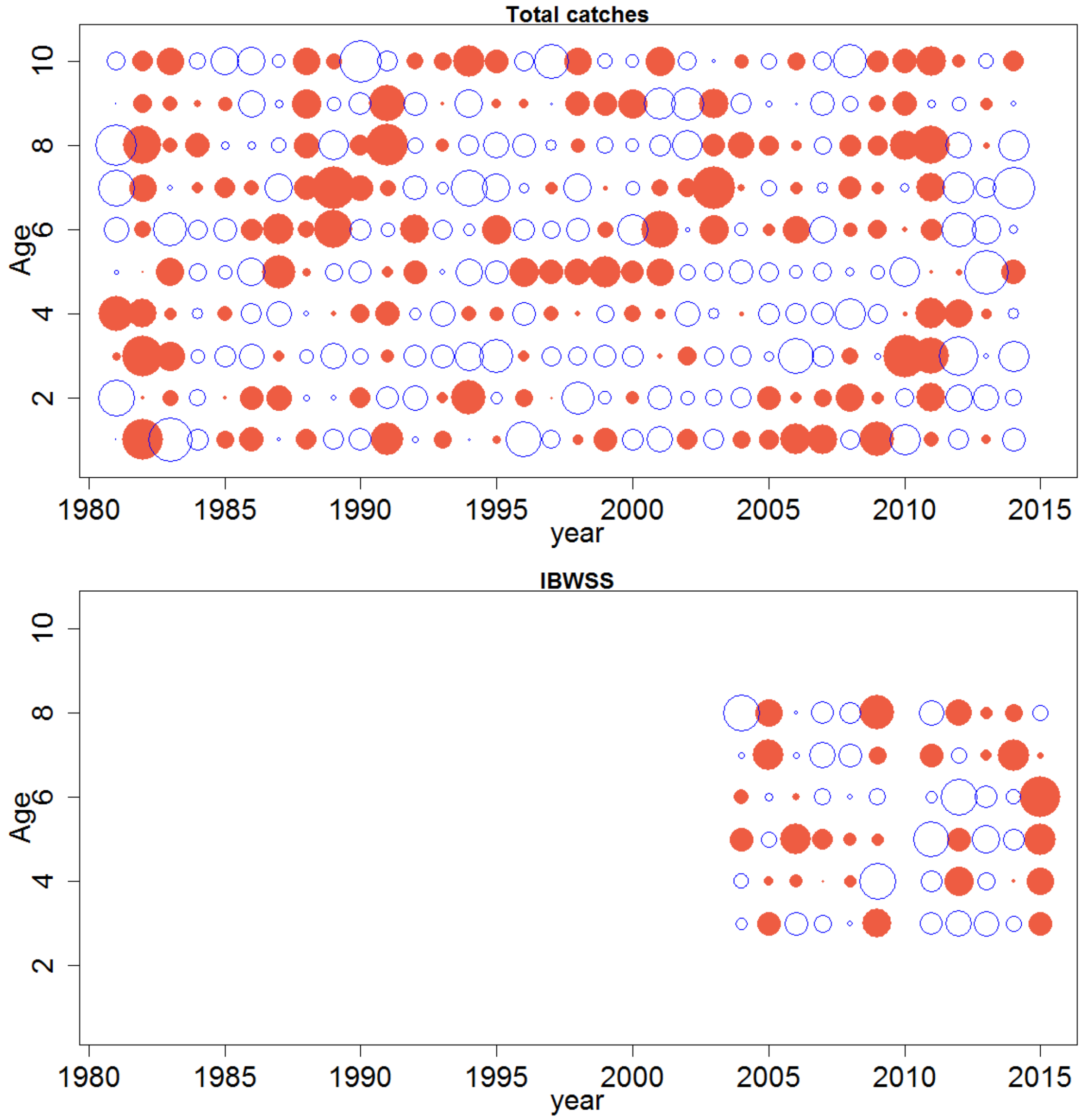


Figure 13. Model 3c. Normalized OSA residuals.

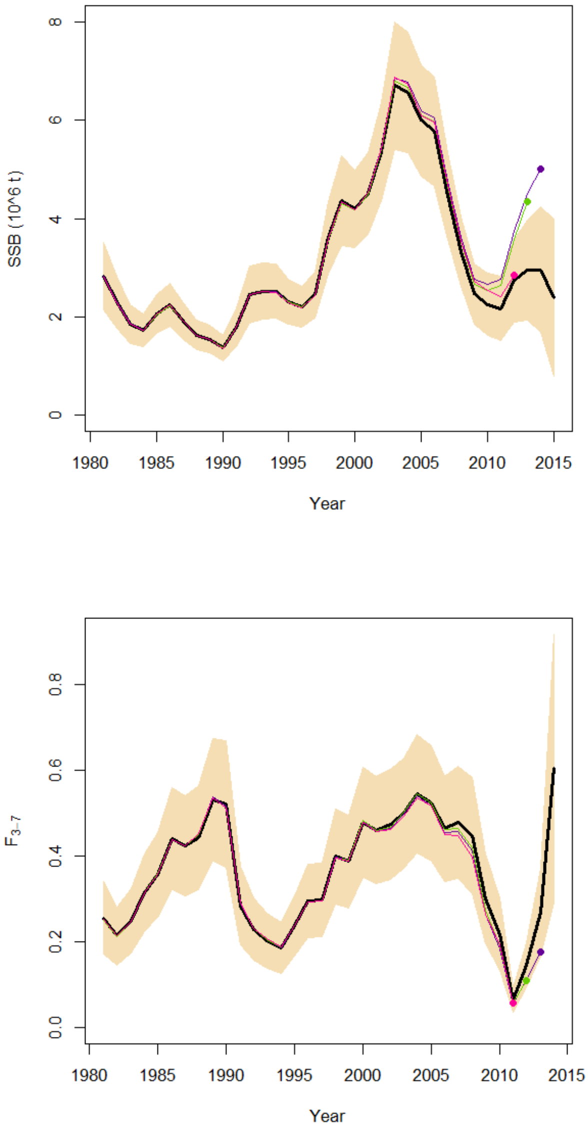


Figure 14. Model 3c. Retrospective analysis.

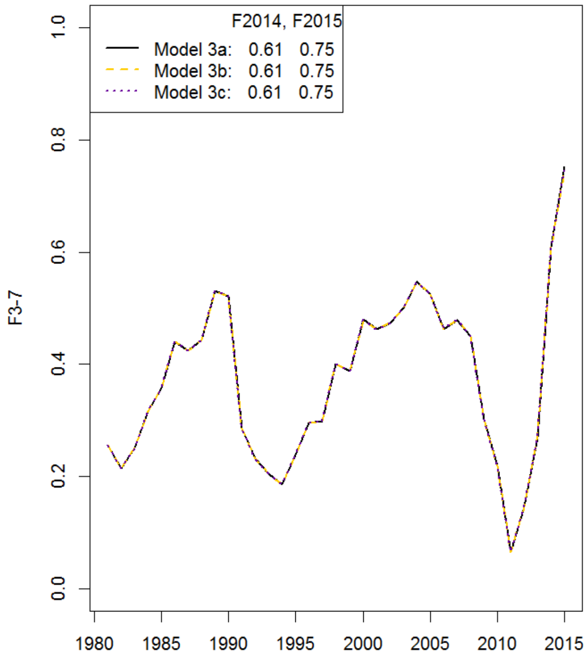


Figure 15. Fbar estimated by the three models (option3)

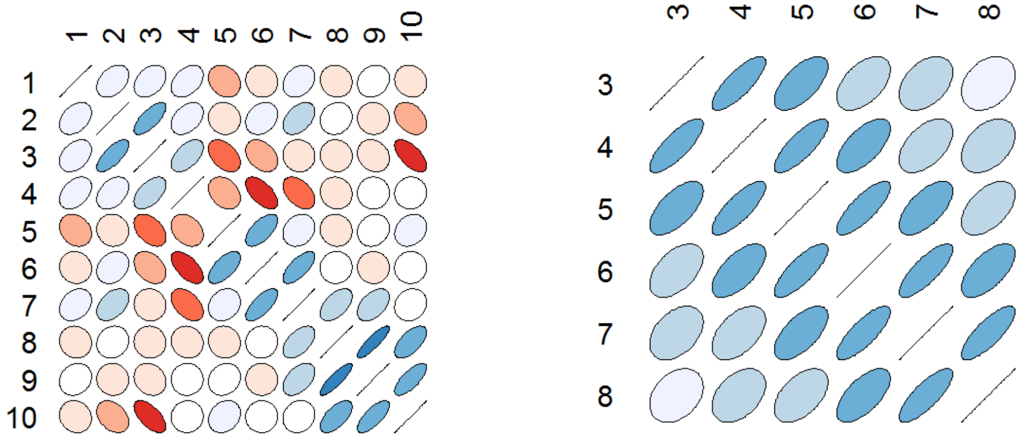


Figure 16. Model 4b. Estimated observation correlation structures by catch at age and IBWSSS CPUE

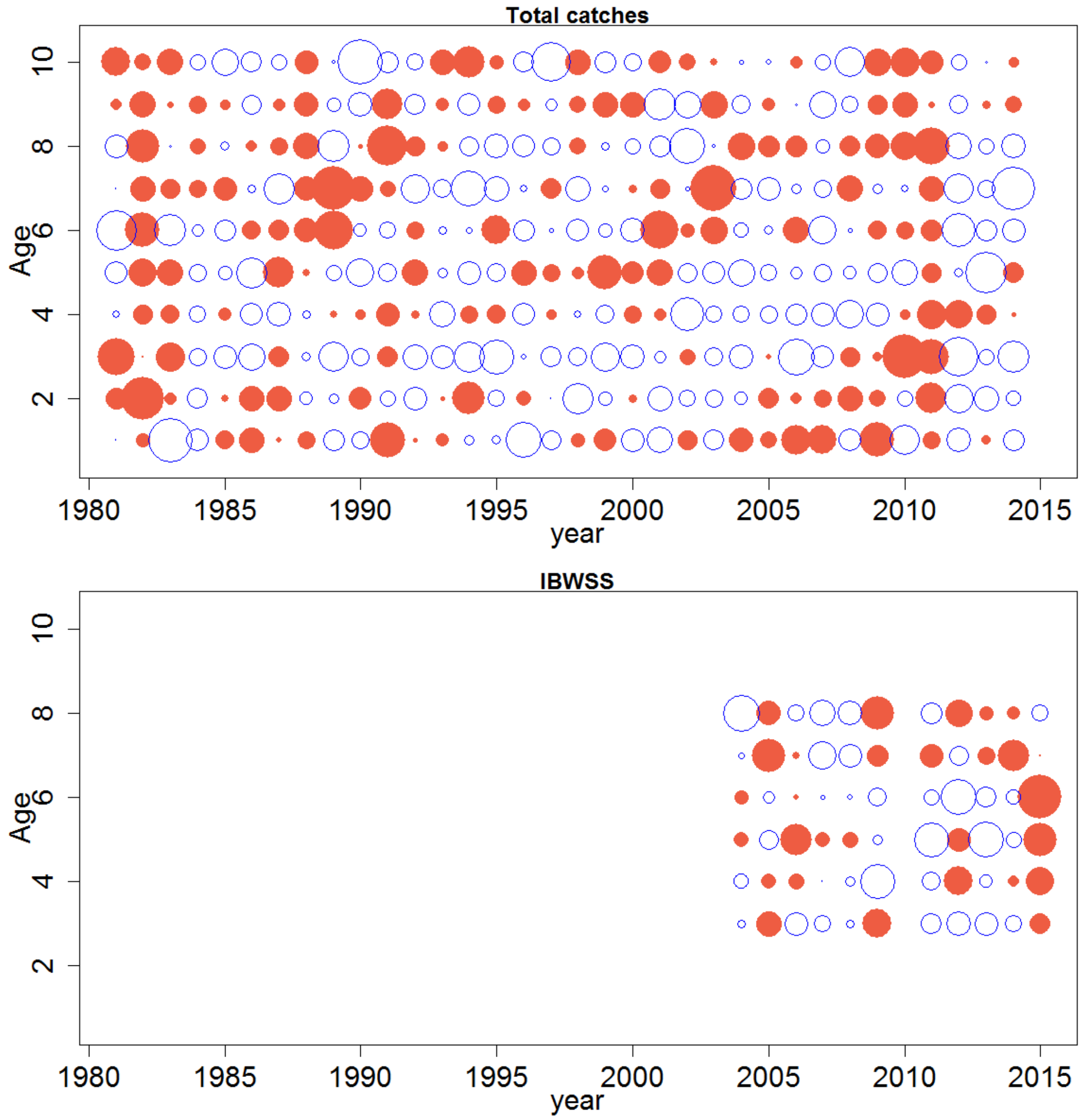


Figure 17. Model 4b. Normalized OSA residuals

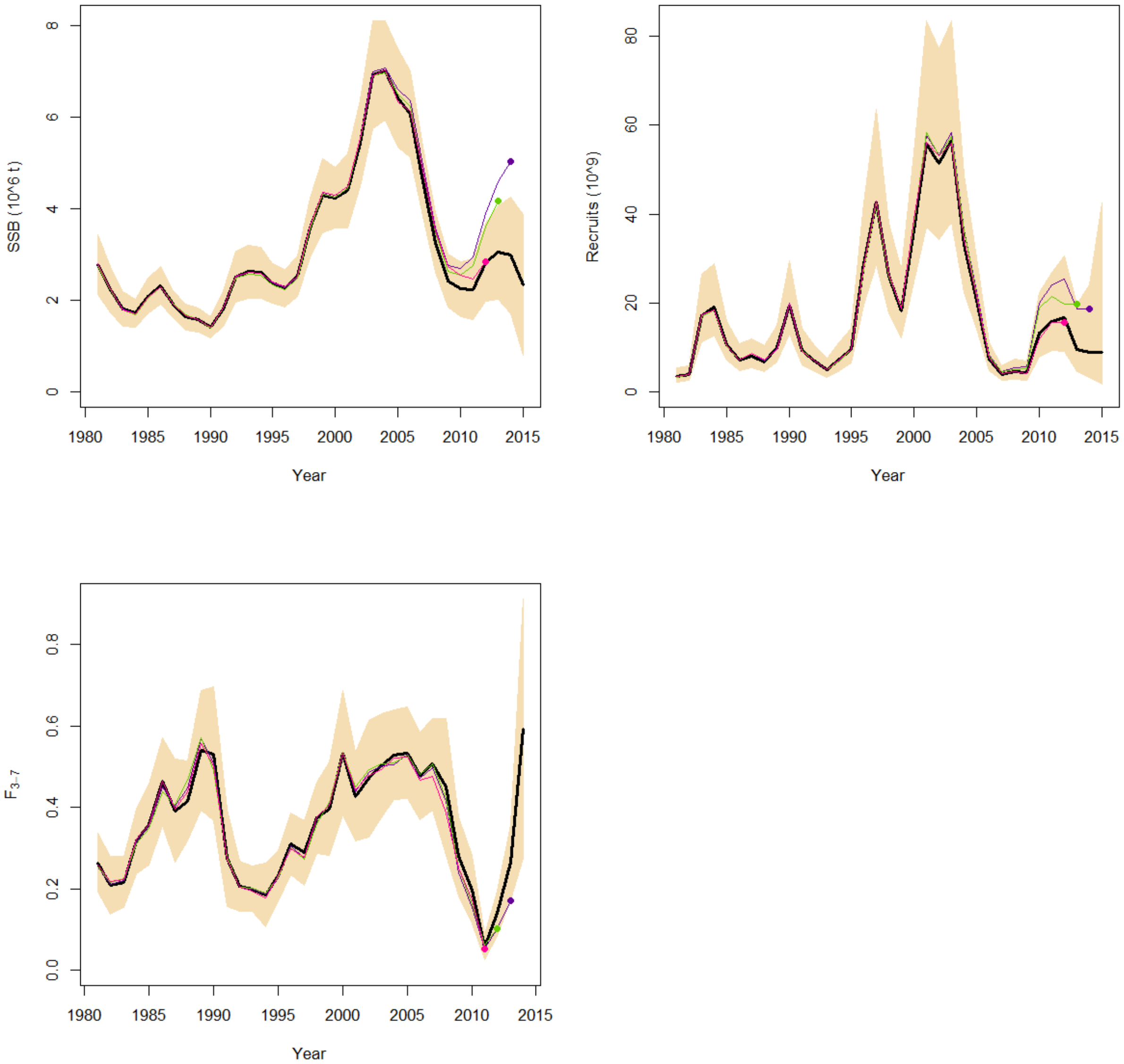


Figure 18. Model 4b. Retrospective analysis.

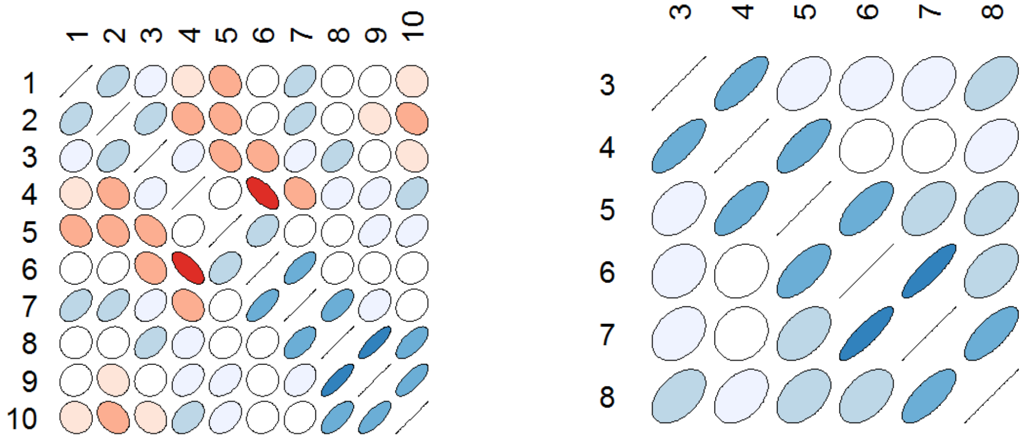


Figure 19. Model 5. Estimated observation correlation structures by catch at age and IBWSSS CPUE

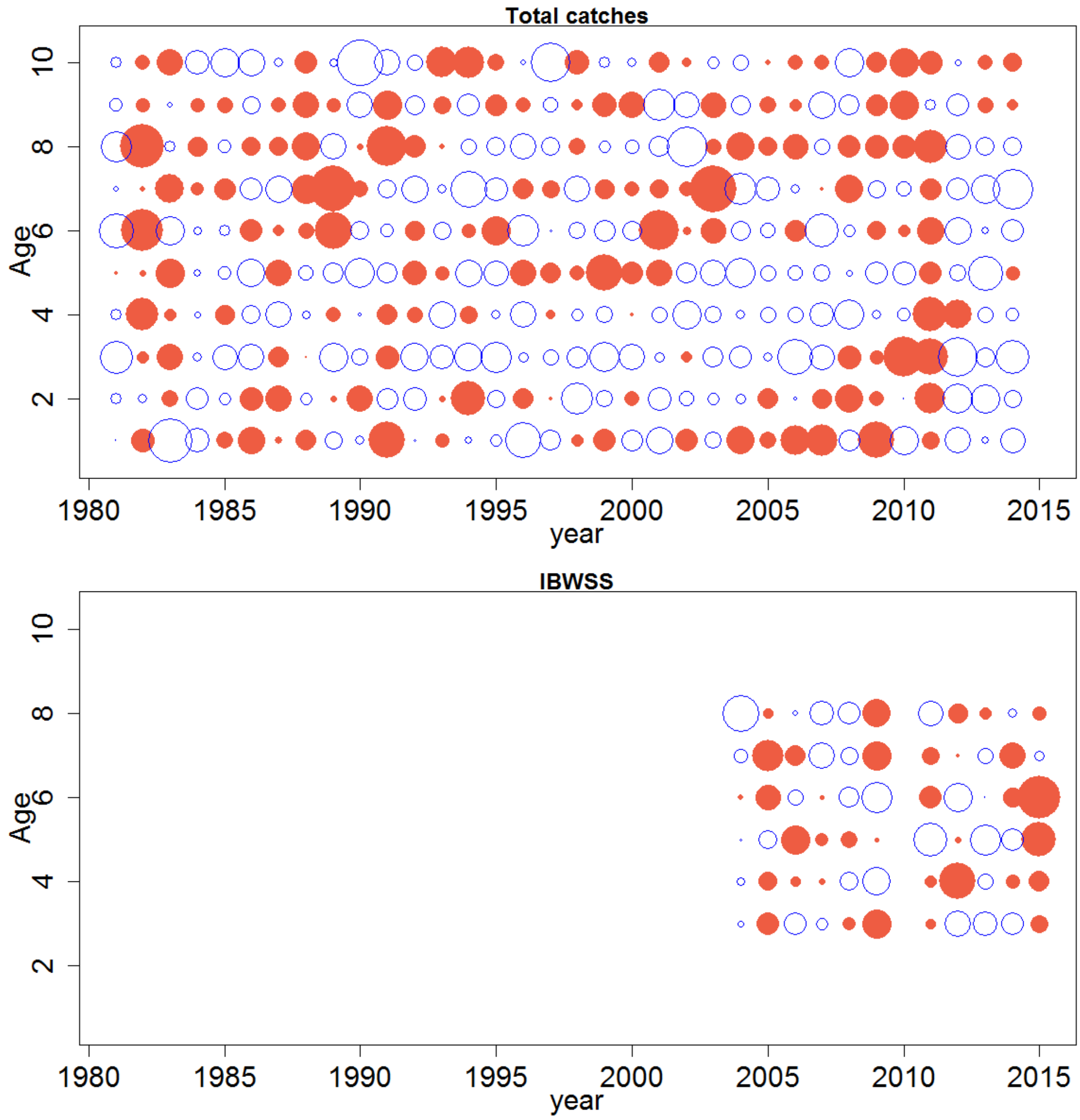


Figure 20. Model 5. Normalized OSA residuals

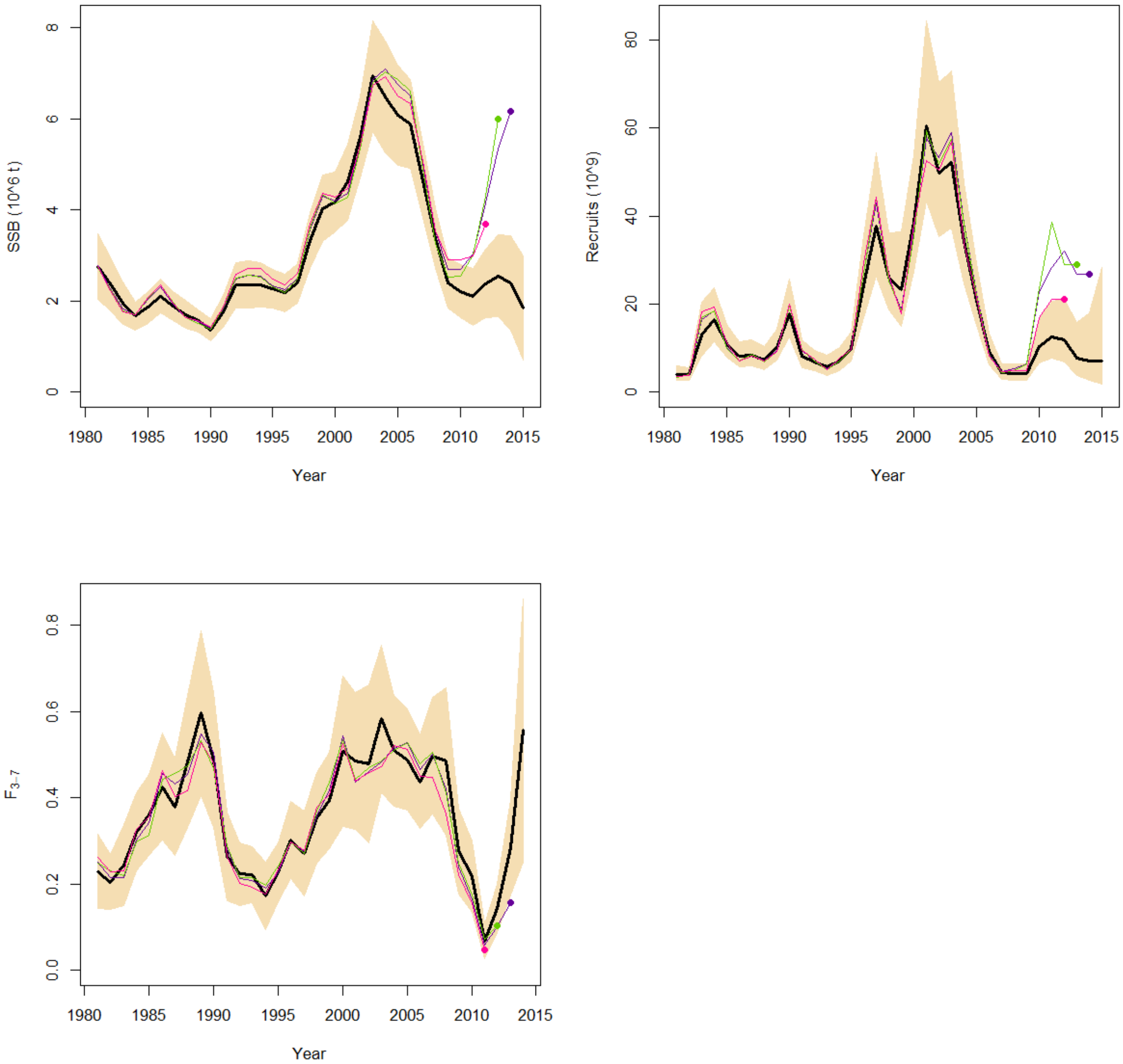


Figure 21. Model 5. Retrospective analysis.

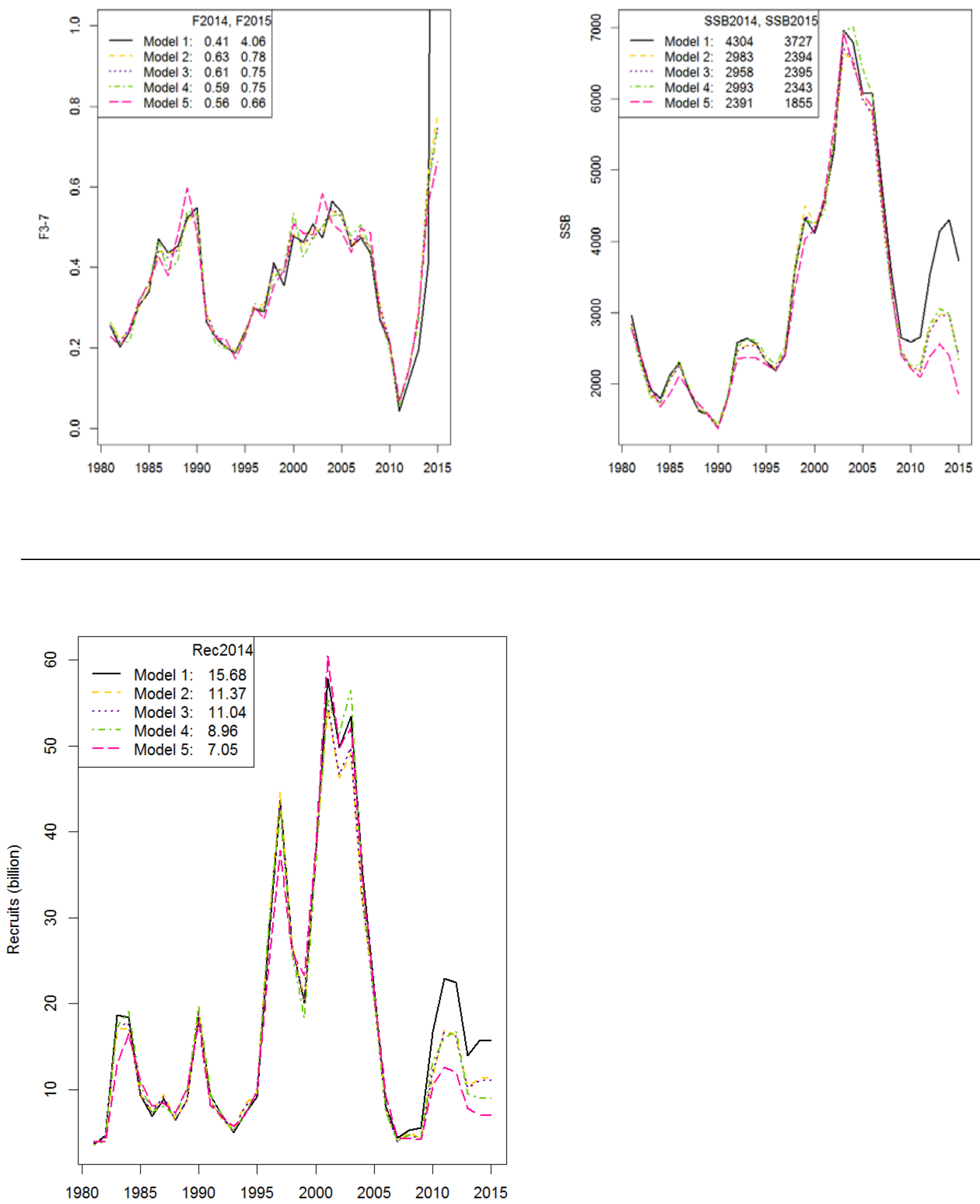


Figure 22. Fbar, SSB and recruitment as estimated by Model option 1 to 5 (final configurations)

1.4 Alternative models

The SMS model was previously used for the assessment of BW. It is a stochastic model with an assumption of separability of F into a fixed age selection and variable year effect. The SAM model estimates highly correlated random walks for the fishing mortalities which indicates that a “separable” F model may be appropriate. In SMS, surveys with data after the last assessment year are assumed to take place the 1. January .

The SMS results (Figure 23) show a steep increase in F in 2014 to a value of 0.80. The 2015 assessment downscales the most recent estimates of SSB and recruitments and upscales F.

Applying the SAM model as default model for BW does not improve the consistency of the assessment.

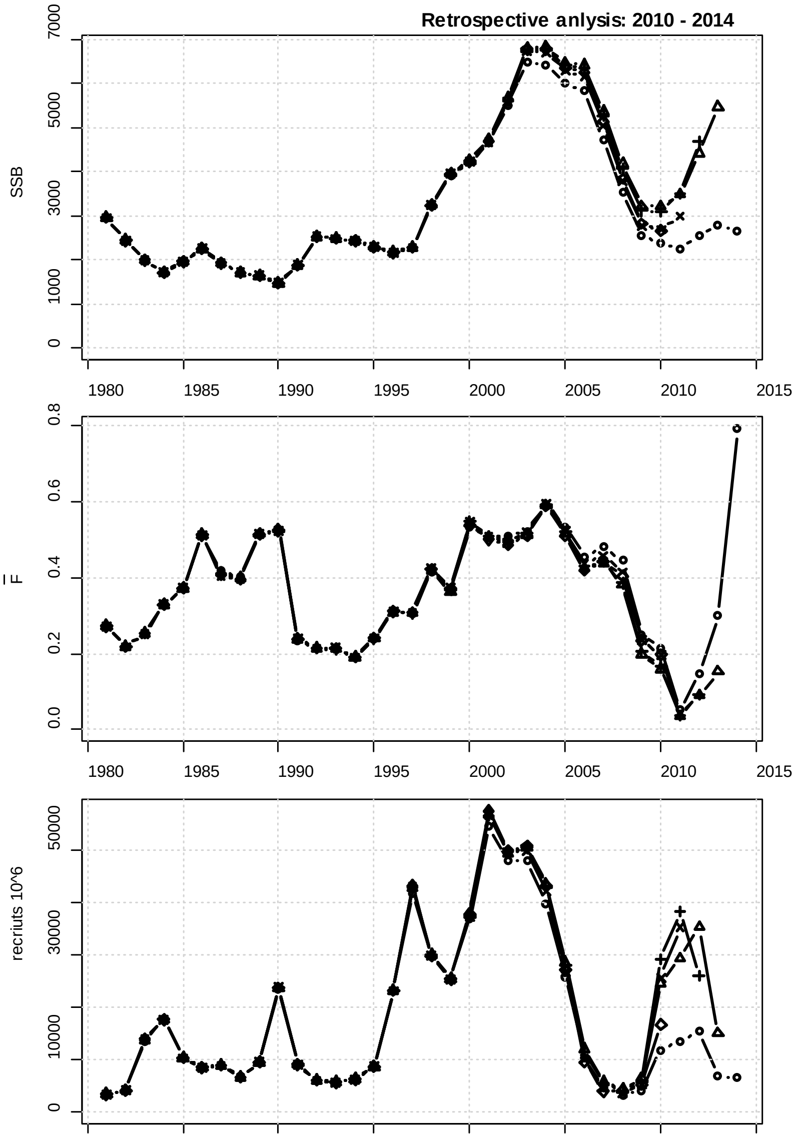


Figure 23. Assessment result and retrospective analysis from the SMS model, assuming that the IBWSSS takes place the 1st January.

1.5 SAM with preliminary catches for 2015

As part of TOR c) "retrospectively examine the impact of including Q1 catches in the assessment year", Thomas Brunel has estimated preliminary catches for 2015 (see WD). These catches are used in the default SAM (without 2016 IBWSSS, not available yet) and in "correlated SAM", Model 3.

1.5.1 Default SAM with preliminary 2015 catches

Catch residuals (Figure 24) have no clear patterns for the most recent years. Survey residuals show a clear positive “year effect” for 2013 and 2014, while CPUE of ages 3-8 in 2015 are much lower than expected.

1.5.1 Correlated SAM (Model 3) with preliminary 2015 catches

Survey residuals (Figure 25) show no “year effect” for 2013 and 2014, while CPUE of ages for 2015 are lower than expected.

1.5.2 Comparison of results

The default SAM model and preliminary 2015 catch data estimate a very high 2015 F (1.51) (Figure 27). When 2015 catch data are used in the “correlated SAM” and F2014 is estimated to 0.46 and F2015 is to 0.68. Without 2015 catch data these F-values becomes slightly higher (0.61 and 0.75).

See WD (Brunel and Vinther) for retrospective performance of use of preliminary catches.

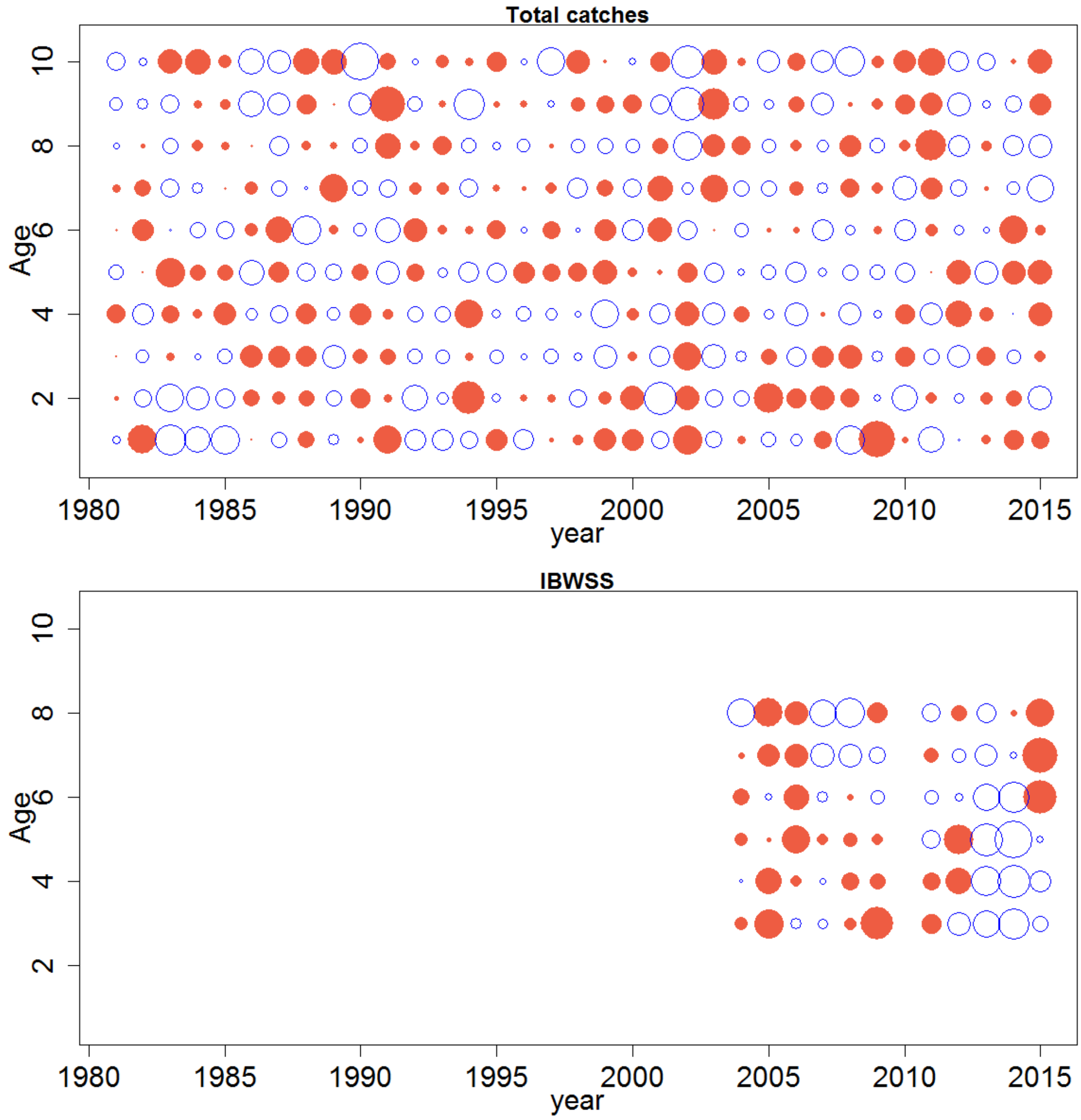


Figure 24. Residuals for default SAM run including preliminary 2015 catches.

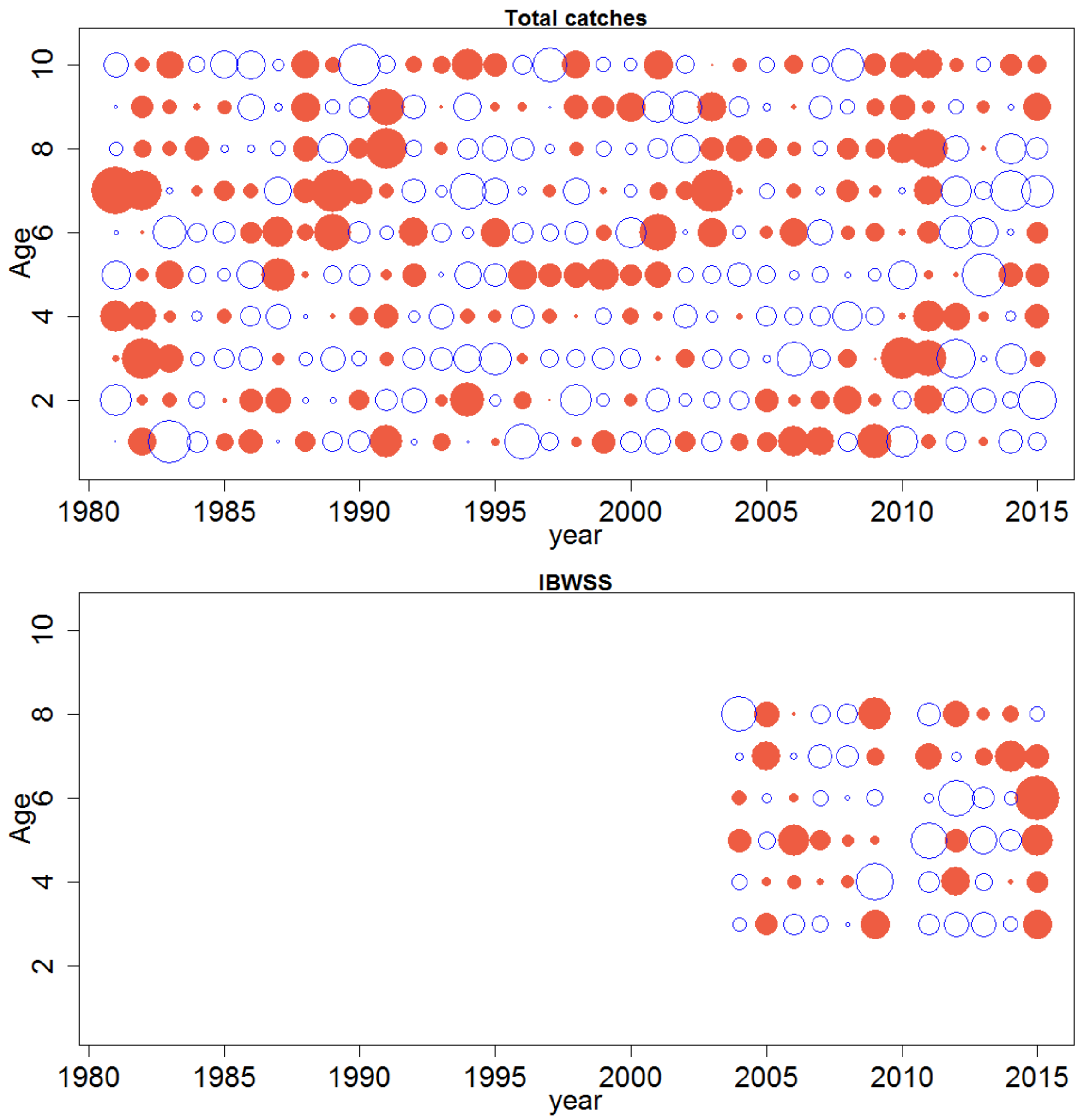


Figure 25. Model 3 with preliminary 2015 catch data. Normalized OSA residuals

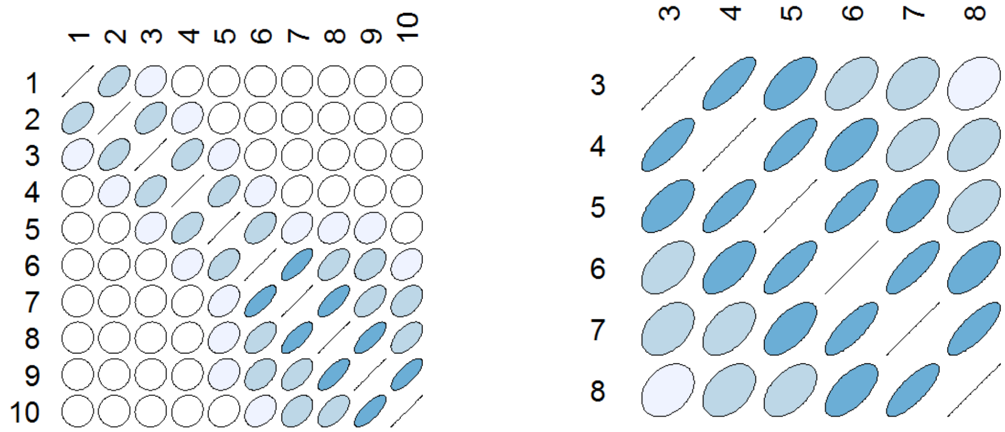


Figure 26. Model 3 with preliminary 2015 catch data Estimated observation correlation structures by catch at age and IBWSSS CPUE.

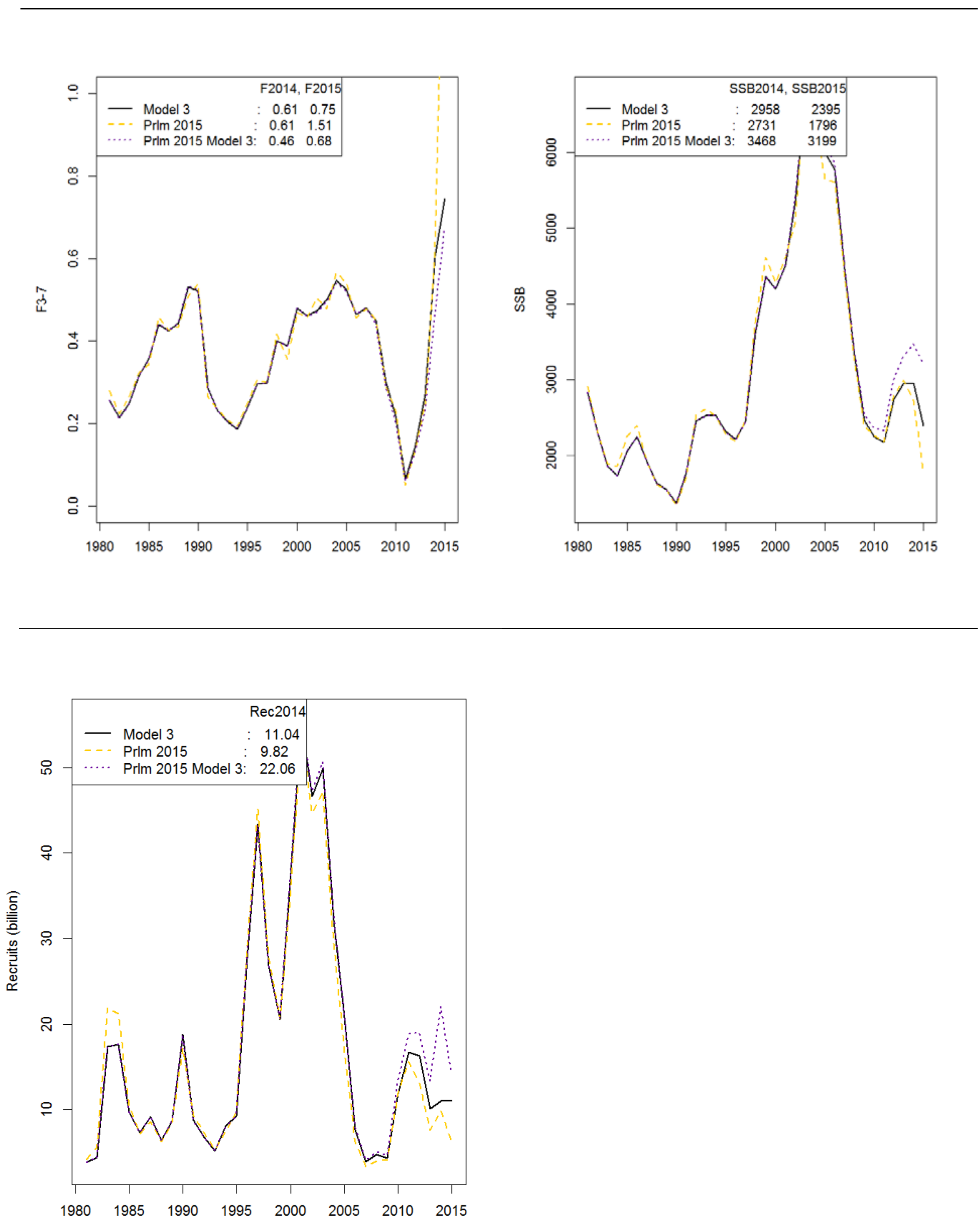


Figure 27. Fbar, SSB and recruitment as estimated by SAM model with correlated observations and catch data 1983-2014 (Model 3), with default SAM model and preliminary catch data for 2015 (Prlm 2015) and with SAM model with correlated observations and preliminary catch data for 2015 (Prlm 2015, Model 3)

1.6 Recruitment for short-term forecast from the present available survey indices

WD "Blue Whiting Recruitment Indicators - Data Exploration" by Mark Payne shows that the IBWSS age 1 and 2, and the IESNS age 1 and 2 both have a good correlation with the recruitment as estimated by the SAM assessment. WD "Blue whiting assessment and HCR simulations" by Höskuldur Bjönsson has the same conclusion, ad suggest a simple sum of the two survey indices, as a combined time series. In this section the addition of the two surveys and the combined survey, using model 3 is investigated.

1.6.1 Recruitment indices, IBWSS age 1 and age 2.

The age 1 and age 2 indices is included in the presently used IBWSS age 3 -8 time series. It is assumed that there is an age dependent catchability for both age 1 and age 2, and observation variance is estimated separately for the two ages. There is an assumed correlation between age 1 and 2, and between age 3-8 indices.

The observation variance is rather high for age 1 and lower for age 2 (Table 1), however both within an acceptable range. The residuals (Figure 28) show no alarming patterns.

The best fit is obtained with an assumption of correlation between age 1-4 and between age 4-8 (Figure 29).

The retrospective analysis (Figure 30) shows a more consistent estimate of SSB compared to the analysis without age 1 and 2 (Figure 14), mainly because the runs with recruitment indices shows an increase in SSB in the most recent years, while the run without recruitment indices shows a decline in SSB. Recruitment is estimated consistently.

1.6.2 Recruitment indices, IESNS age 1 and age 2.

The fit for the IESNS indices is poor with observation variance higher than 1.5 for the two ages (Table 2).

1.6.3 Recruitment indices, IBWSS age 1 and age 2, IESNS age 1 and age 2.

The fit for the IESNS indices is poor with observation variance higher than 1.5 for the two ages (Table 3).

1.6.4 Recruitment indices, IBWSS age 1 and age 2, IESNS age 1 and age 2 summed.

The indices from a simple sum of the IBWSS and the IESNS indices do not give a good fit (Table 4). The observation variance is higher than 1.5 for both ages.

1.6.5 Comparison of results, conclusion

The model without recruitment indices (model 3) and the model with the summed IBWSS and IESNS give pretty much the same results (Figure 31). The recruitments indices have such a high observation variance, that their influence on the overall model results become low, resulting in a similar result as running the model without these indices. Using the IBWSS as the only indices or using the IBWSS together with the IESNS indices give almost the same results. This is because the IESNS has a very high observation variance and becomes down weighted in the model results.

The IBWSS indices increases the recruitment indices compared to the model without these indices. F becomes lower and SSB higher in the model with IBWSS indices. This might be because the extra age classes in the IBWSS increases the number of observations to fit and thereby put less weight on the IBWSS age 3-8 indices and maybe because the two youngest age classes constrain the estimate of the older ages in the IBWSS survey.

Based on this exercise, the IBWSS age 1 and age are the only indices that should be used for recruitment estimate.

Table 1. Parameter estimate, Model 3 plus IBWSS age 1 and age 2 survey indices.

	Value	CV
Random walk variance		
--- F	0.399	0.15
--- log(N@age1)	0.586	0.15
Process error		
--- log(N@age2 to 10+	0.174	0.12
Observation variances		
--- Catch age 1	0.451	0.18
--- Catch age 2	0.301	0.23
--- Catch age 3-8	0.203	0.15
--- Catch age 9-10	0.408	0.13
--- IBWSS age 1	0.695	0.24
--- IBWSS age 2	0.329	0.29
--- IBWSS age 3	0.430	0.24
--- IBWSS age 4-6	0.423	0.21
--- IBWSS age 7-8	0.408	0.23
Survey catchability		
--- IBWSS age 1	0.056	0.26
--- IBWSS age 2	0.143	0.17
--- IBWSS age 3	0.389	0.18
--- IBWSS age 4	0.700	0.17
--- IBWSS age 5-8	0.948	0.17
rho	0.927	

Table 2. Parameter estimate, Model 3 plus IESNS age 1 and age 2 survey indices.

	Value	CV
Random walk variance		
--- F	0.399	0.15
--- log(N@age1)	0.605	0.15
Process error		
--- log(N@age2 to 10+	0.170	0.12
Observation variances		
--- Catch age 1	0.421	0.19
--- Catch age 2	0.331	0.22
--- Catch age 3-8	0.203	0.15
--- Catch age 9-10	0.409	0.13
--- IBWSS age 3	0.490	0.27
--- IBWSS age 4-6	0.419	0.20
--- IBWSS age 7-8	0.385	0.22
--- IESNS age 1	2.407	0.24
--- IESNS age 2	1.524	0.19
Survey catchability		
--- IBWSS age 3	0.429	0.20
--- IBWSS age 4	0.755	0.17
--- IBWSS age 5-8	1.018	0.16
--- IESNS age 1	0.070	0.66
--- IESNS age 2	0.117	0.41
rho	0.930	

Table 3. Parameter estimate, Model 3 plus IBWSS age 1 and age 2, and IESNS age 1 and age 2 survey indices

	Value	CV
Random walk variance		
--- F	0.397	0.15
--- log(N@age1)	0.599	0.15
Process error		
--- log(N@age2 to 10+	0.170	0.13
Observation variances		
--- Catch age 1	0.443	0.18
--- Catch age 2	0.303	0.23
--- Catch age 3-8	0.203	0.15
--- Catch age 9-10	0.409	0.13
--- IBWSS age 1	0.691	0.23
--- IBWSS age 2	0.356	0.28
--- IBWSS age 3	0.423	0.24
--- IBWSS age 4-6	0.428	0.21
--- IBWSS age 7-8	0.408	0.22
--- IESNS age 1	2.350	0.24
--- IESNS age 2	1.543	0.19
Survey catchability		
--- IBWSS age 1	0.056	0.25
--- IBWSS age 2	0.145	0.17
--- IBWSS age 3	0.390	0.18
--- IBWSS age 4	0.704	0.17
--- IBWSS age 5-8	0.955	0.17
--- IESNS age 1	0.059	0.64
--- IESNS age 2	0.104	0.41
rho	0.926	

Table 4. Parameter estimate, Model 3 plus IBWSS age 1 and age 2, and IESNS age 1 and age 2 survey indices SUMMED.

	Value	CV
Random walk variance		
--- F	0.399	0.15
--- log(N@age1)	0.597	0.15
Process error		
--- log(N@age2 to 10+	0.169	0.13
Observation variances		
--- Catch age 1	0.416	0.18
--- Catch age 2	0.340	0.21
--- Catch age 3-8	0.205	0.15
--- Catch age 9-10	0.410	0.13
--- IBWSS age 3	0.500	0.27
--- IBWSS age 4-6	0.418	0.20
--- IBWSS age 7-8	0.384	0.22
--- IESNS+IBWSS age 1	2.899	0.51
--- IESNS+IBWSS age 2	2.016	0.21
Survey catchability		
--- IBWSS age 3	0.429	0.20
--- IBWSS age 4	0.752	0.17
--- IBWSS age 5-8	1.012	0.17
--- IESNS+IBWSS age 1	0.091	0.97
--- IESNS+IBWSS age 2	0.199	0.60
rho	0.929	

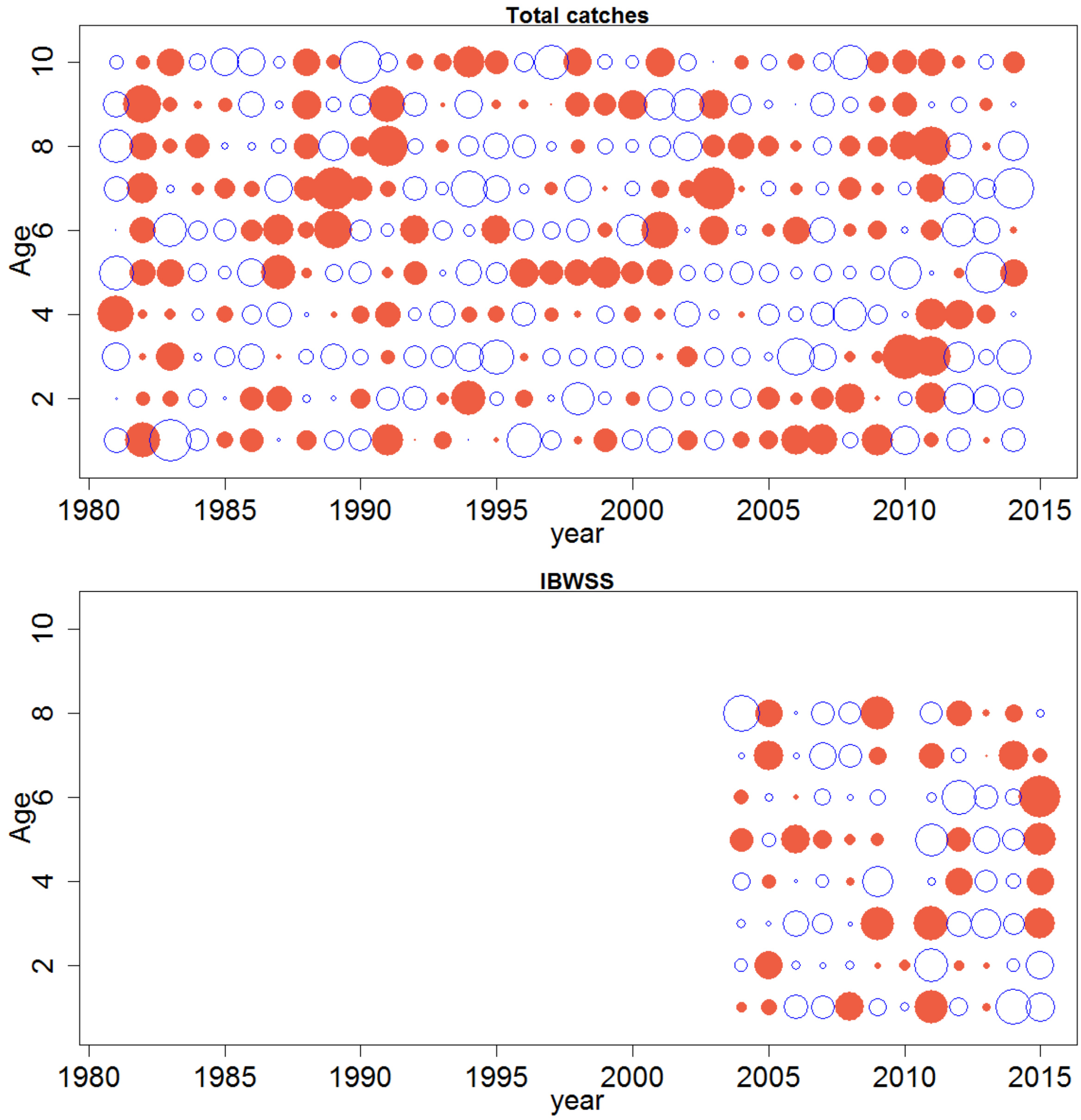


Figure 28. Residuals, Model 3 plus IBWSS age 1 and age 2 survey indices.

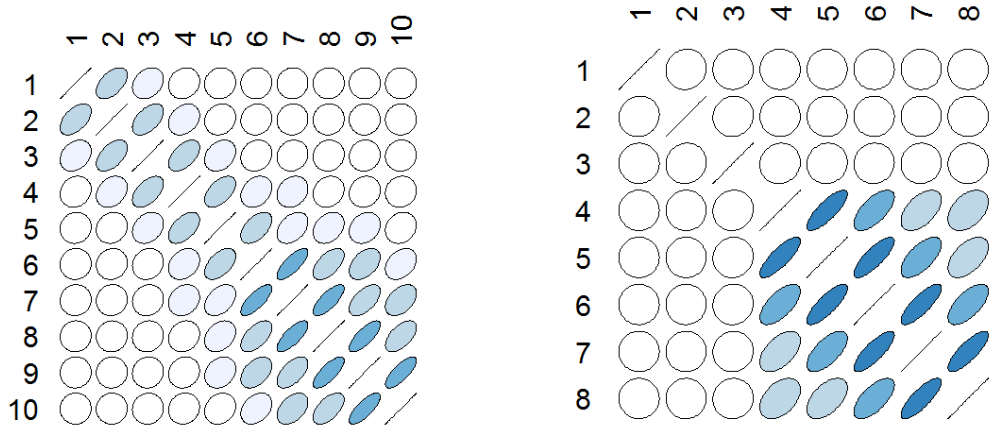


Figure 29. Model 3 plus IBWSS age 1 and age 2 survey indices. Estimated observation correlation structures by catch at age and IBWSSS CPUE.

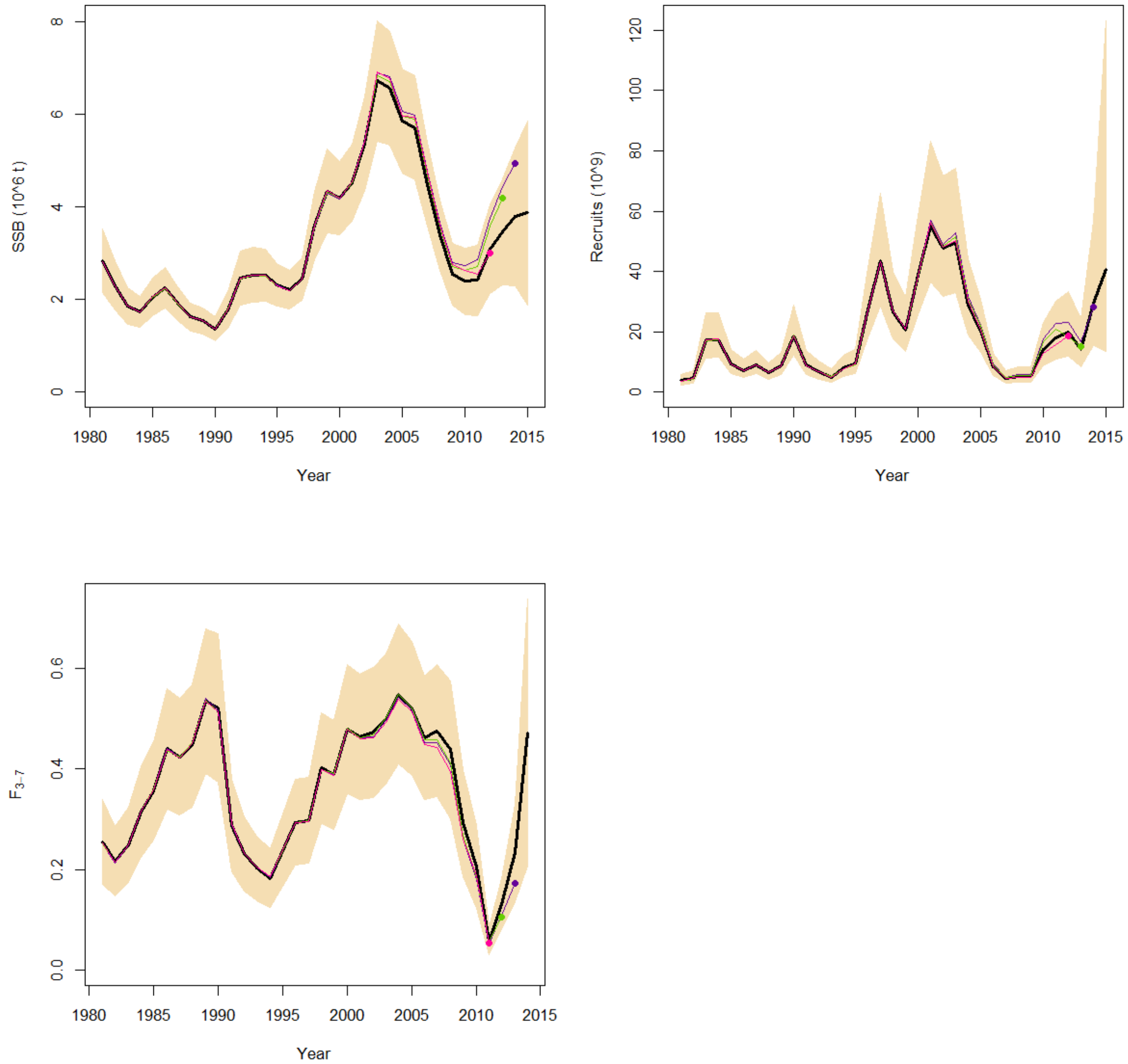


Figure 30. Retrospective analysis. Model 3 plus IBWSS age 1 and age 2 survey indices

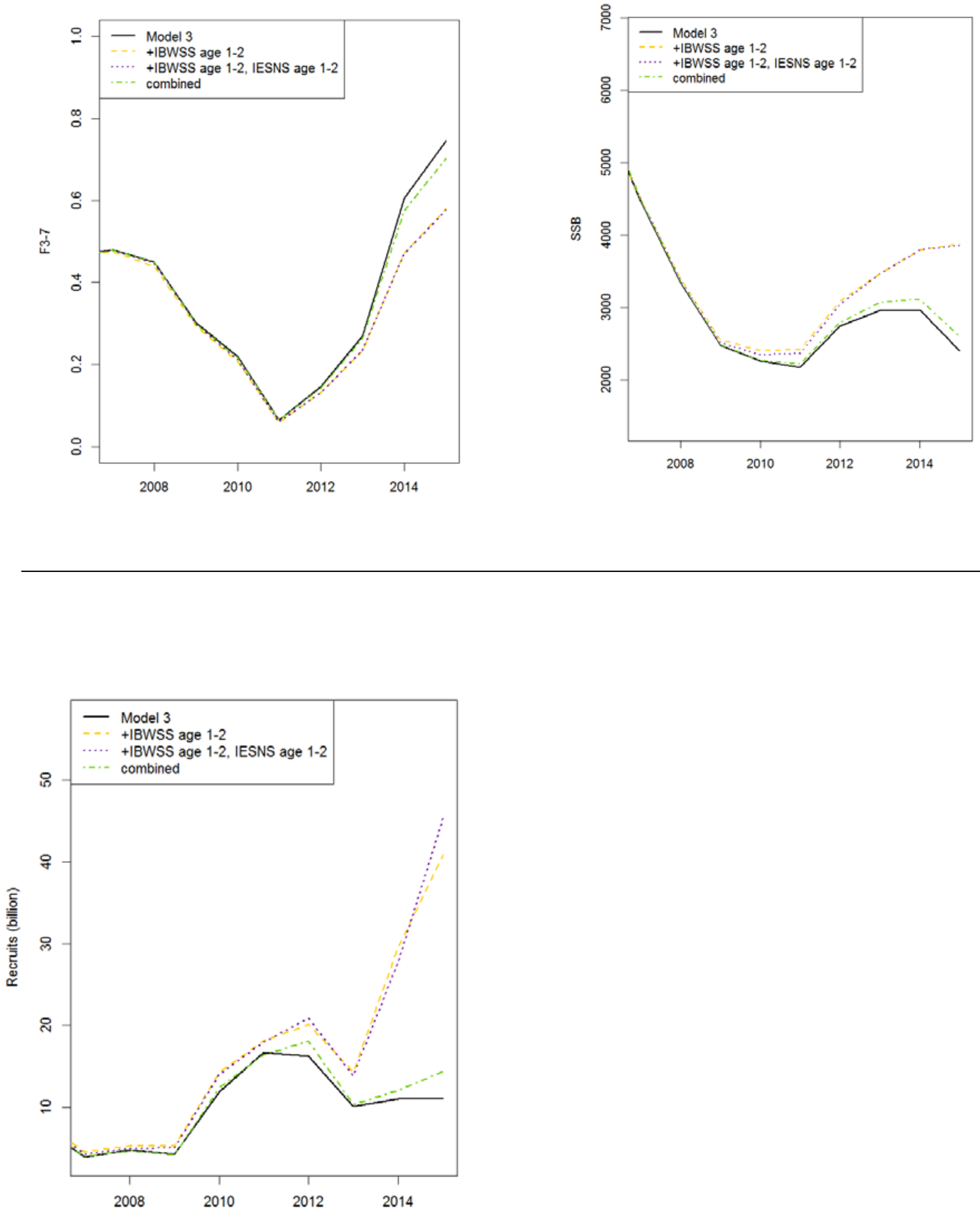


Figure 31. Fbar, SSB and recruitment as estimated by SAM model with correlated observations and catch data 1983-2014 (Model 3), with additional survey data for estimation of recruits.

1.7 Conclusions, TOR a and TOR c.

1. The default SAM configuration gives an unrealistically high F for 2015 and cannot be used with the present IBWSSS time series.
2. No simple fix has been found for criteria for discarding survey indices (the 2015 indices). This may indicate that I am too naïve and that the problem is not just that the 2015 indices are too low. Survey indices for 2013 and 2014 might also be too high!
3. Survey period (The proportion of F and M before the survey):
    a. The default SAM assessment is highly sensitive to the survey period in the 2015 assessment, but not sensitive in the 2014 assessment.
    b. The proportion of annual catch taken before the survey has changed considerably in the IBWSSS year range (2004-2014), which will bias the assessment results.
    c. Back-shifting the survey to 31. December gives a slightly worse model fit (negative log likelihood) and results in a (unrealistic) high (1.18) F2014.
    d. Changing the survey period seems not to solve the problem.
4. The SMS model (a simpler model than SAM) estimates a high F (0.8) for 2014 and shows that the addition of the last year's assessment data creates a substantial revision of previous assessment results. A simpler model may improve the quality of the assessment in this case? , however estimated F for 2014 by SMS might be too high.
5. The use of preliminary catches for 2015 and the default SAM model creates F2015 at 1.15, which seems too high. The main problem seems to be the survey results and preliminary catches do not help much to fix that problem.
6. The use of the SAM model taking correlation of observations into account (Berg and Nielsen, 2016) gives a better model fit than the default SAM and a (more) realistic F2015 (0.75).
7. The use of preliminary catches for 2015 and the "correlated SAM" model creates F2015 at 0.68, slightly lower than the model run without 2015 catch data.
8. The age 1 and age 2 IBWSS indices seem suitable for recruitment estimation.
9. The IESNS indices seem not suitable for recruitment estimation, neither a simple sum of IESNS and IWBSS indices.
10. The use of IBWSS recruitment indices in a "correlated SAM" run seems to give a more consistent retrospective estimate of SSB and F, compared to a run without these recruitment indices.
11. Whatever we do, F in 2014 and 2015 seem higher than intended.

Main conclusion: The issue seems to be too high IBWSSS indices for 2013 and 2014 followed by too low indices for 2015. To handle these "year effects" we should use the SAM model taking correlation of observations into account (Berg and Nielsen, 2016), Model 3 configuration. The use of preliminary (raised) Q1 catches, or even better, raised catches from the first half year (Q1 and Q2 which include ~90% of annual catches) seems to give a more realistic F for 2015, however (see WD Brunel and Vinther) retrospectively examination of the impact of including Q1 catches in the assessment year seems not to improve the quality of the assessment much. The use of IBWSS age 1 and age 2 indices gives a good fit and can be used to estimate recruitment. The use of the indices also improves the retrospective pattern.

Blue Whiting Recruitment Indicators - Data Exploration

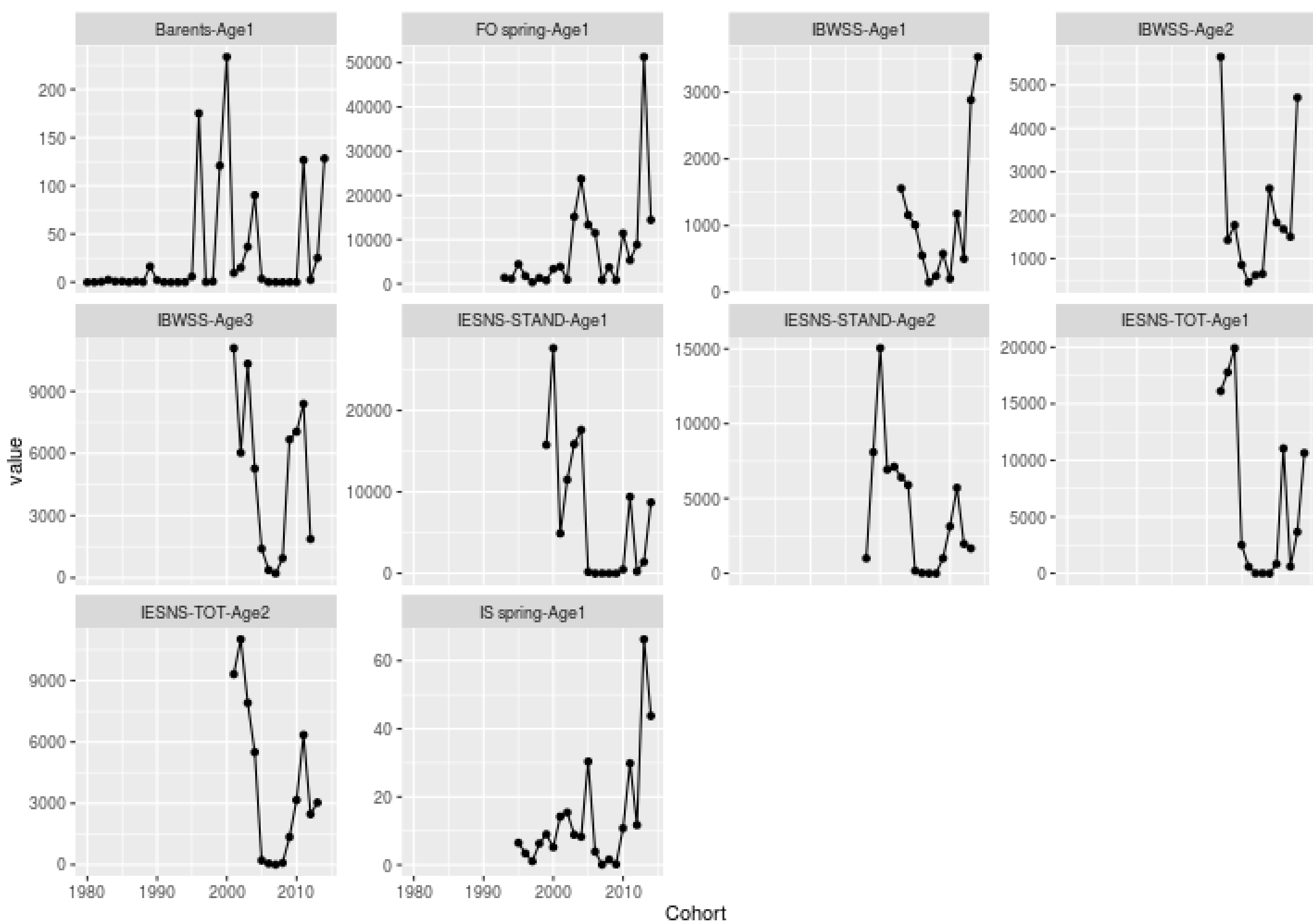
by Mark R Payne
DTU-Aqua, Charlottenlund, Denmark
http://www.staff.dtu.dk/mpay

Fri Mar 18 13:01:03 2016

Summary

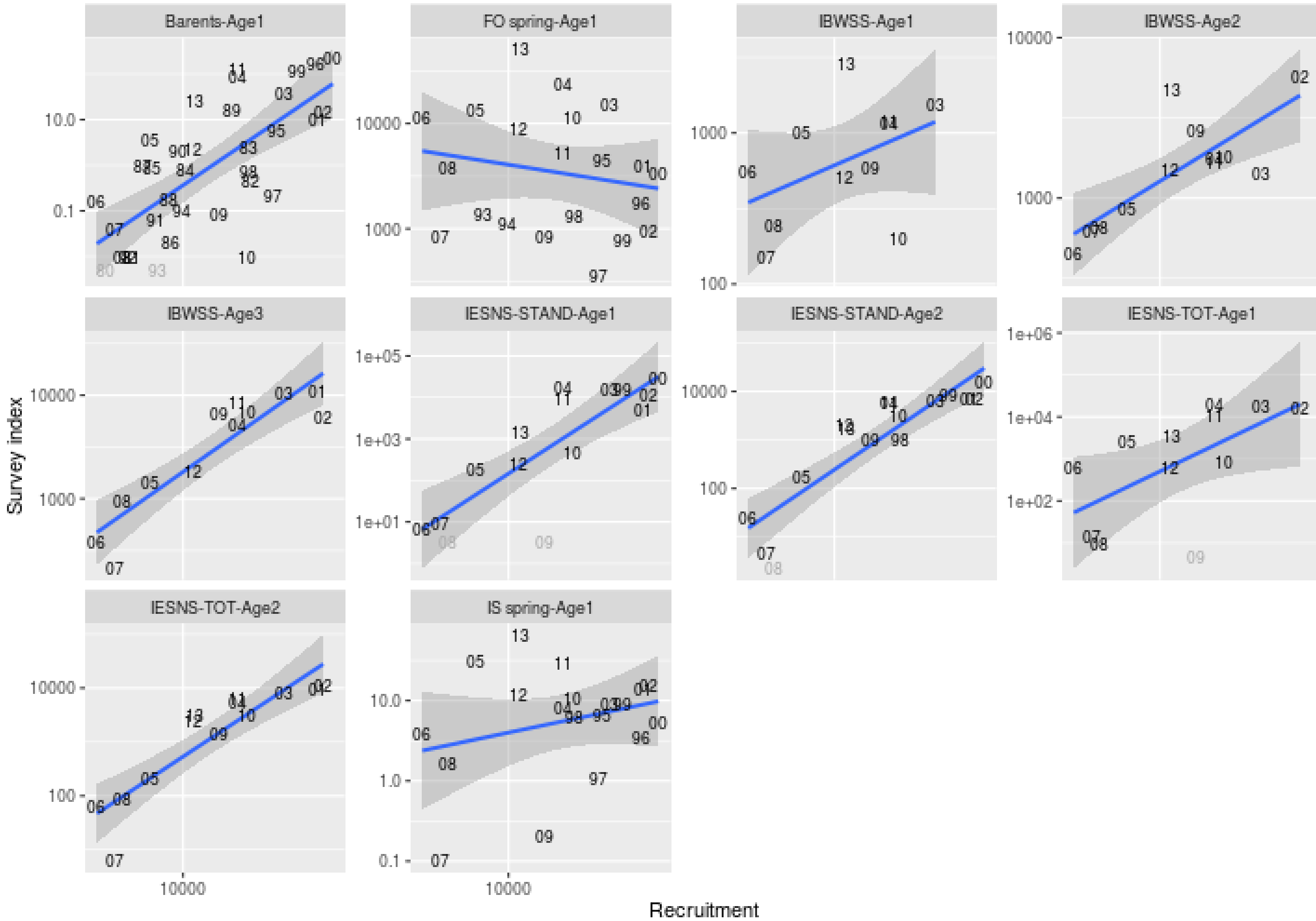
As part of the ICES Interbenchmark Protocol on Blue Whiting (Micromesistius poutassou), I have performed as basic data exploration of the time-series that have previously been used as recruitment indicators in the assessment of this stock. I find that the recruitment indicators show substantial variation in their ability to represent the variations in recruitment, and that some of the previous used indicators are not at all correlated with recruitment. There is some evidence however, that the relationships between recruitment and the surveys are non-stationary, and this may need to be accounted for in future modelling endeavours. Strong auto-correlation in the recruitment time-series is present and can potentially be used to our advantage in a forcasting context.

Survey Time Series



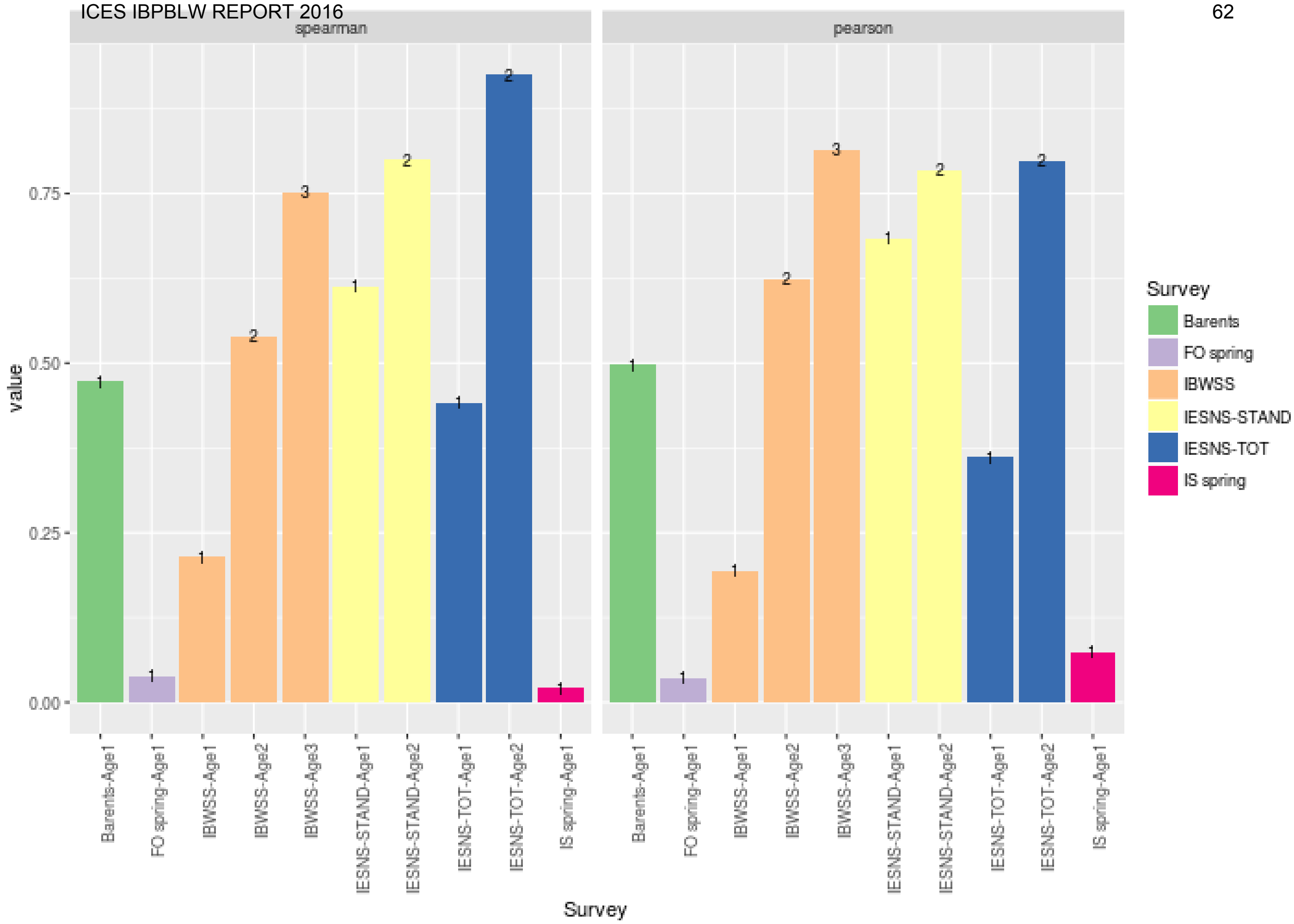
The individual time series available are plotted as a function of the cohort strength that they are measuring. Note the varying vertical scales in each panel.

X-Y Scatter Plots



Recruitment estimated by the stock assessment is plotted against the corresponding values in each individual time series. Points are labelled according to the last two digits of their cohort. Note the logarithmic transformations on both scales - survey values of zero are replaced by a "Limit of Detection" value corresponding to 0.5 of the lowest non-zero value and are plotted in grey

Correlations between time-series

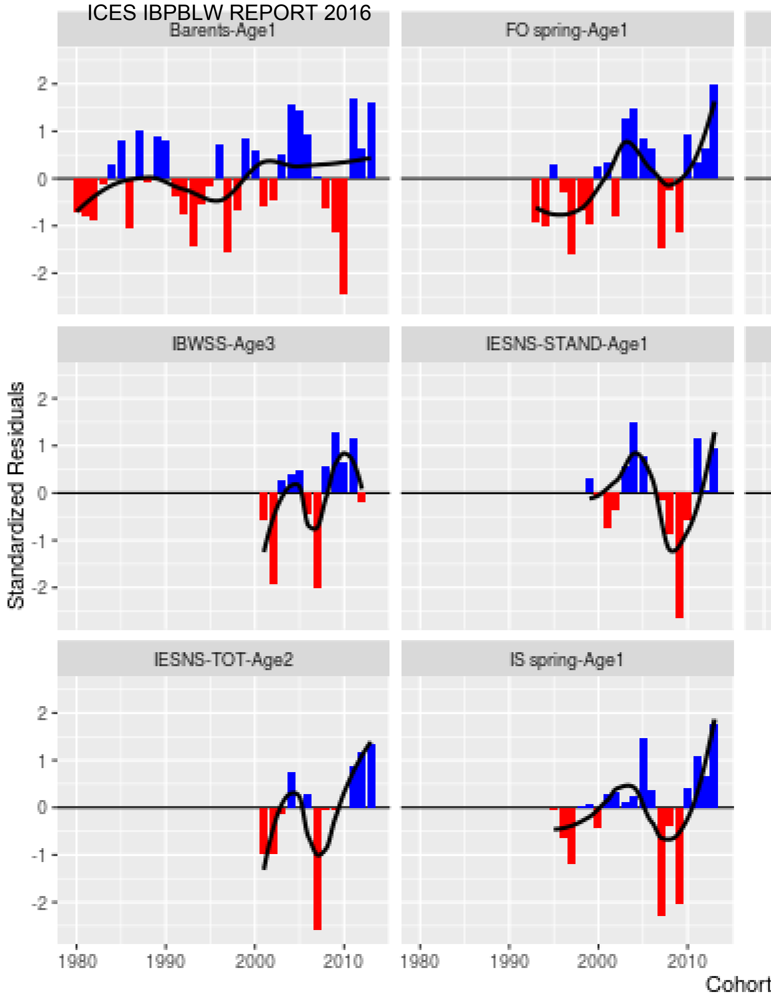


Correlation between the log-transformed time-series and the log-transformed recruitment (with limit-of-detection adjustments made as appropriate). The separate panels correspond to spearman (rank) correlation coefficients and pearson (r^2) correlation coefficients. Bars are coloured according to the survey and labelled with their age, for easy reference.

There appears to be good agreement between the pearson and spearman results. Most surveys, with the notable exceptions of the Farorese and Icelandic surveys, appear to contain some information about year class strength.

Time-varying relationships

One potential concern in a widely distributed stock like blue whiting is that the proportion of the juveniles that end up in one area may not be the same every year due to shifts in migration patterns, environmental conditions or oceanic circulation. Such changes can potentially cause a previously-strong relationship between an individual survey index and recruitment to suddenly break down. We can text for such changes quite simply, however, by examining a time-series of the residuals of the fitted relationship between the survey index (Limit-of- detection adjusted, log transformed) and the recruitment (log-transformed).

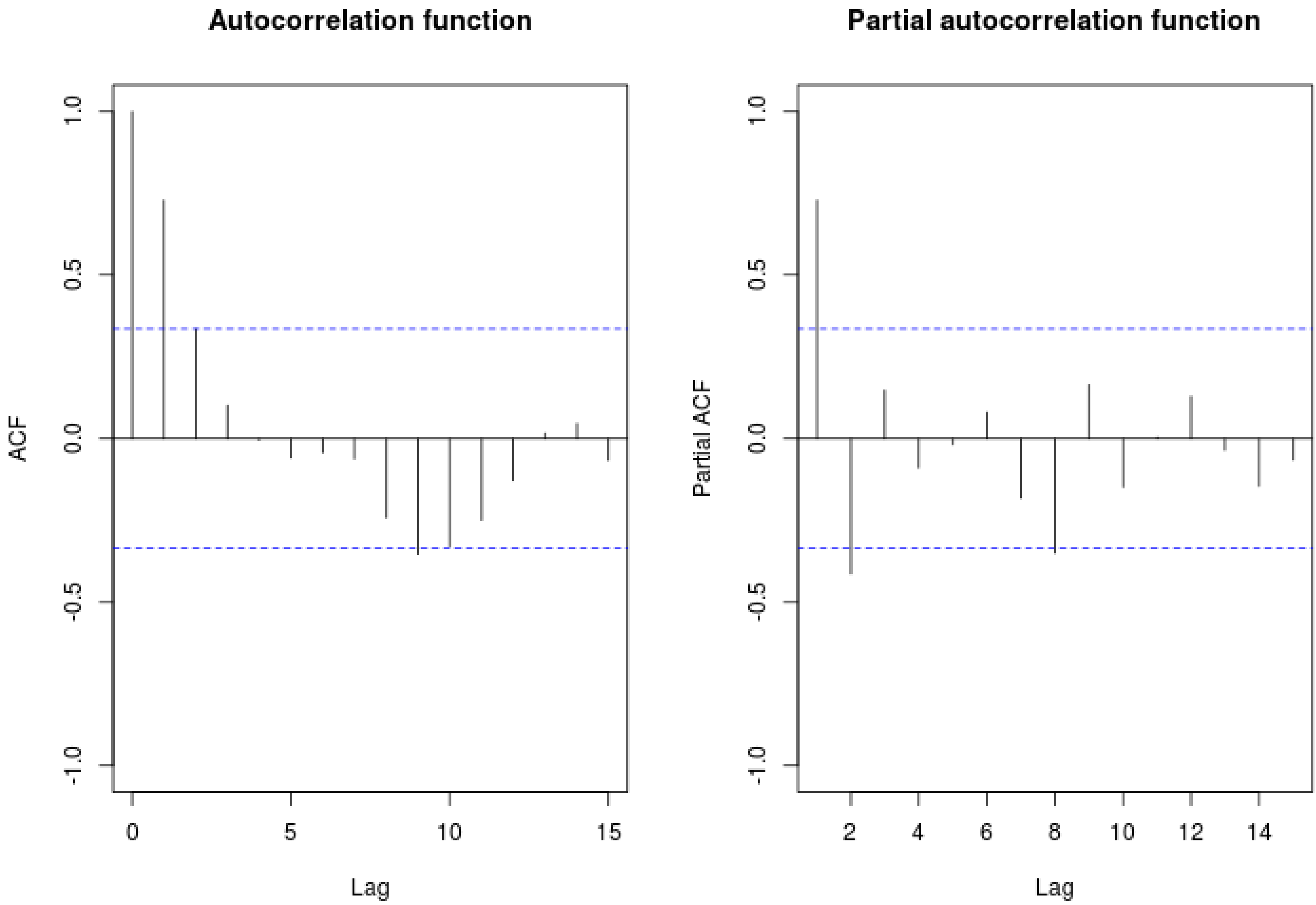


Timeseries of residuals in the survey-recruitment relationship for each survey index. Standardised residuals in individual-years are plotted as bars (red for negative residuals, blue for positive) while a loess smoother (black line) is added to guide the eye.

There is some evidence here of non-stationary relationships between the individual surveys and the recruitment time-series. However, part of the problem appears to arise due to the presence of outliers. Modelling approaches may need to account for this by using robust regression approaches and allowing time-varying catchability.

Recruitment autocorrelation

Finally, we examine the degree of auto-correlation present in the recruitment time series. This auto-correlation can potentially be used to our advantage in a forecasting context.



Autocorrelation and partial-autocorrelation functions for the (log-transformed) recruitment time series. Dotted blue lines correspond to 95% significance levels for this time series.

The presence of significant autocorrelation at lag 1 and partial autocorrelation at lag 2 suggests that a time-series model with both autoregressive and moving average elements (i.e. ARMA) may be appropriate for this time series.

Blue whiting assessment and HCR simulations.

The Icelandic Blue Whiting Commission.

March 22, 2016

1 Recruitment estimates

Recruitment indices are not used directly for tuning in the Blue Whiting assessment. Still a number of recruitment indices are available, both from acoustic and bottom trawl surveys. The most important recruitment survey is probably the acoustic survey in the Norwegian sea in May (IESNS). Information are also available from the acoustic survey on the spawning ground (IBWSS). That series is rather short or from 2004 but can possibly be extended back by use of the Norwegian survey in the spawning grounds that was conducted in most years from 1991-2003.

The acoustic survey in the Norwegian sea (IESNS) has been conducted in May since 1995. Prior to 2004 only areas in the Norwegian Sea outside the Icelandic EEZ were covered. In 2004 and 2005 areas in the east and south east of the Icelandic zone were partly covered. Since 2006 the shelf areas west and south of Iceland and areas in the east and southeast of Iceland have been covered.

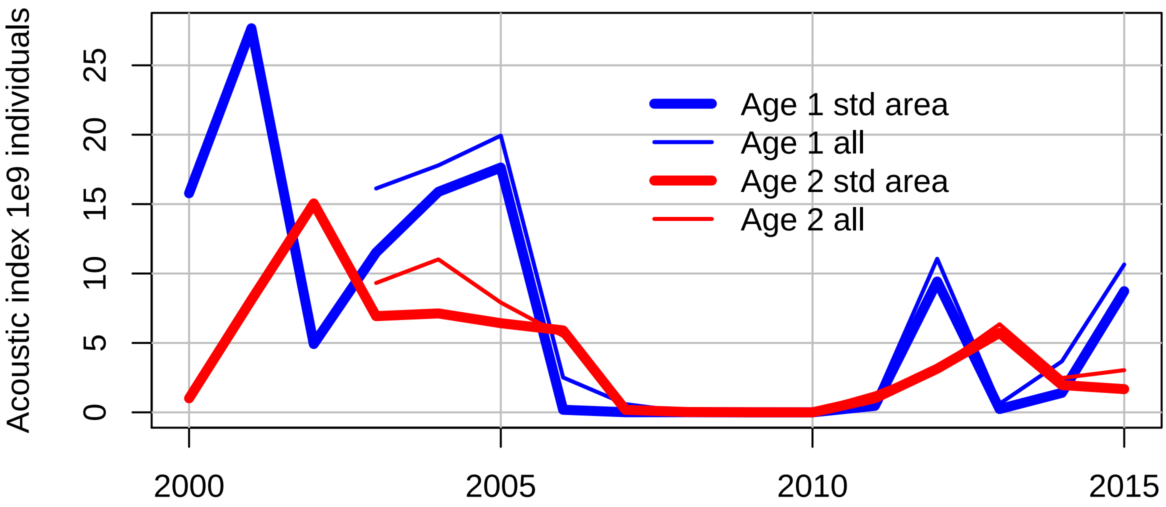


Figure 1: Indices of age 1 and 2 from the acoustic survey in the Norwegian sea

Large yearclasses are usually detected in the survey (IESNS) while the smaller ones are not observed figure (1). The correlation between age 1 and age 2 one year later is high or 0.9 on "ordinary scale", driven mostly by few large yearclasses (figure 2). Correlations on log scale is lower as the small cohorts are hardly seen.

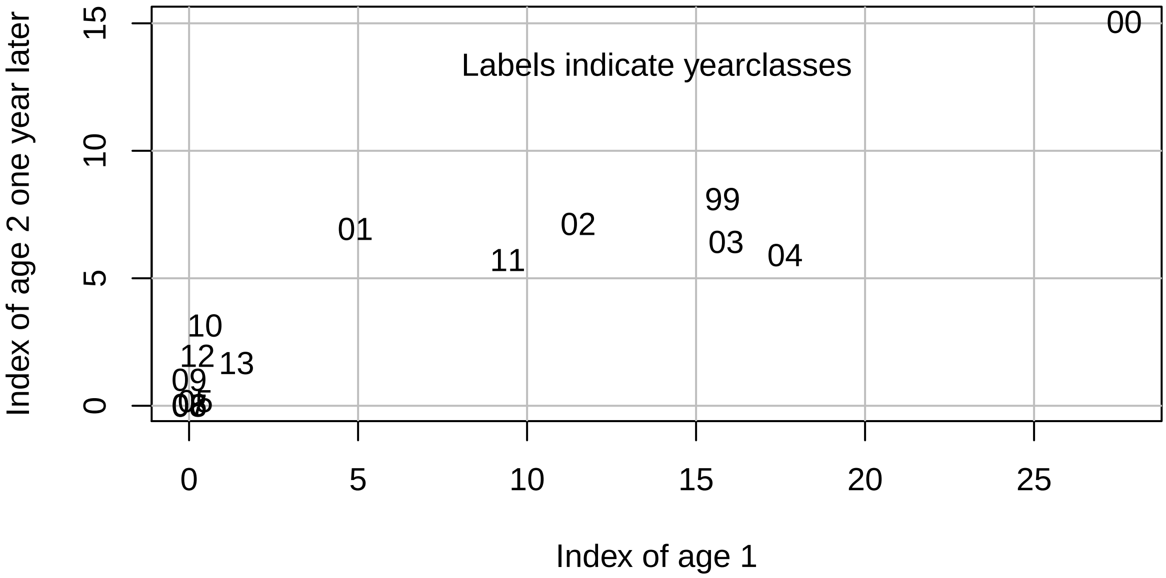


Figure 2: Age 2 indices from the acoustic survey in the Norwegian sea (IESNS) plotted against index of age 1 indices a year earlier. The indices shown apply to the standard area, surveyed since 2000.

Comparison of age 1 with estimated recruitment shows some major discrepancies, escpecially for yearclasses 2000 - 2004. There are indications that expansion of the area in 2004 did help, and the Icelandic groundfish survey in March indicates that yearclasses 2001 and 2002 were abundant in Icelandic waters (figure 5). Also, yearclasses 2009 and 2010 are not observed in the survey and small yearclasses are usually not noticed (might be caused by the nature of acoustic measurements??). It is interesting to see that the 2011 yearclass is rather large in the survey at age 1 but of similar size in assessment as yearclasses 2009-2010 that are not seen in the survey. The 2011 yearclass also appeared as strong as age 2 in 2013. A reservation must made in the discussions here that the estimated number of age 1 from assessment in last 5 years is of rather uncertain.

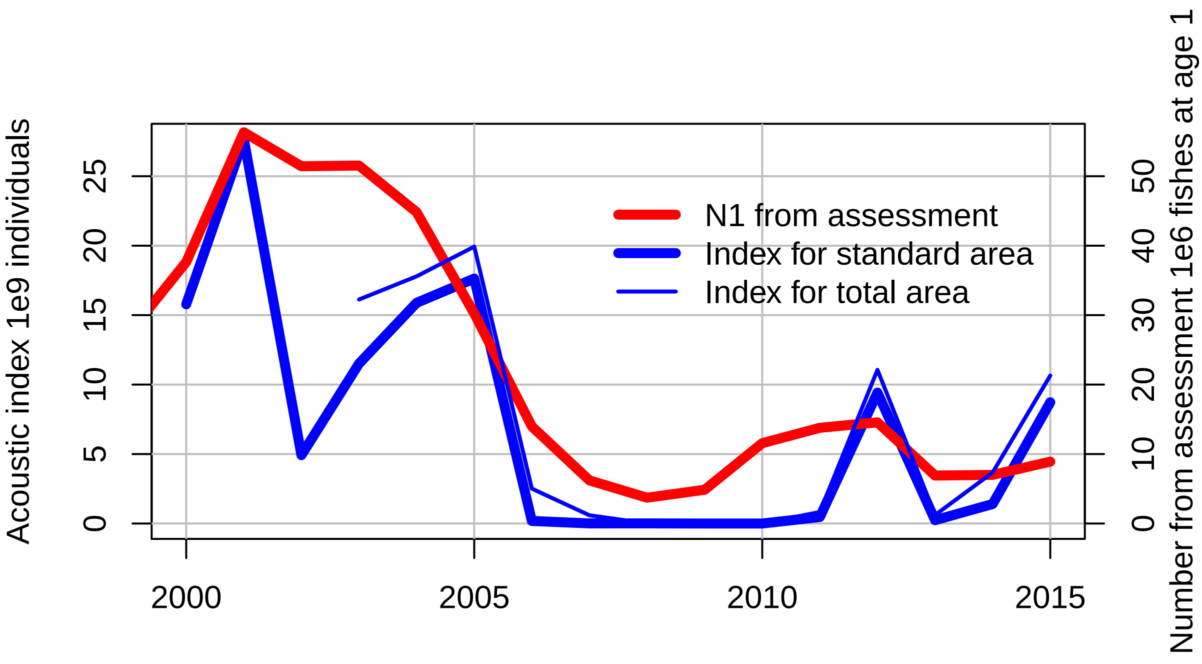


Figure 3: Indices of age 1 from IESNS against estimated number of age 1 from assessment.

Recruitment indices from the acoustic survey on the spawning grounds (IBWSS) show more fish in 2014 and 2015 compared IESNS (figure 4 and the values seen in 2014 and 2015 are the highest since the series started in 2004 (the series is of course rather short). Could be an indication of earlier maturation? or just change in distribution.

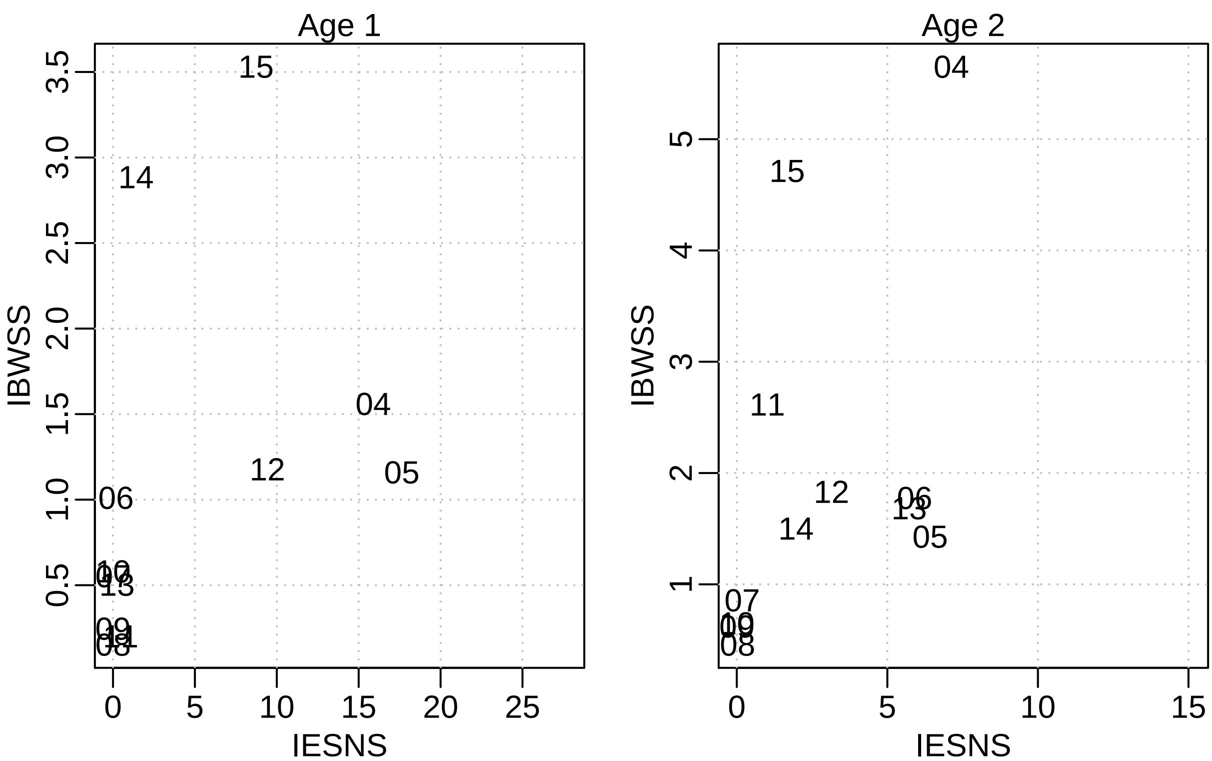


Figure 4: Comparison of recruitment indices from IBWSS and IESNS. Labels indicate years

The acoustic surveys are not the only data on recruitment as information are available from the groundfish surveys in Iceland, Barentssea and the Faeroe Islands. Indices from those sureys are not age disaggregated so fish below certain length (20-22cm) is interpreted as age 1. Autocorrelation of recruitment of blue whiting is very high (1st order AR model fitted has $\rho \approx 0.8$) so age 1 and 2 are often abundant in the same years. The fish are also mostly pelagic so occurrence on the bottom is rather sporatic, still the indices seem to give indication about recruitment. Indices of age 1 scaled to their own mean are shown in figure 5. The acoustic survey shows less variability than the demersal surveys, not an unexpected result as schools of fish might occasionally be close to bottom.

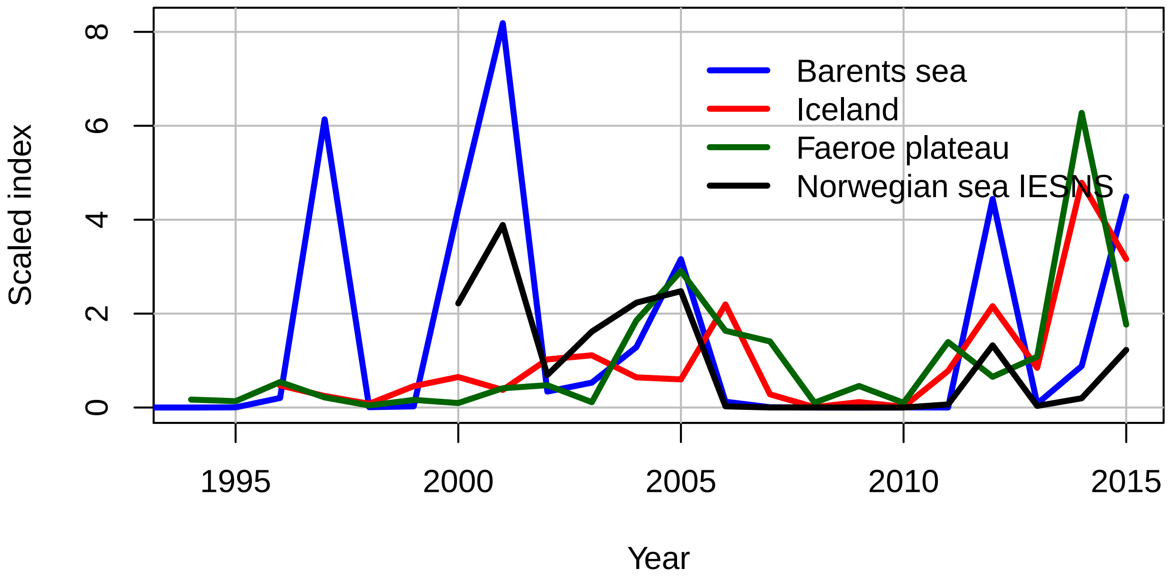


Figure 5: Indices of age 1 from different bottom trawl surveys, each scaled to its mean. Results from IESNS are shown for comparison.

How indices from different surveys are correlated is of interest (figure 6). Iceland and the Faeroe Island correlate rather well as do the acoustic indices (IBWNS) and indices from the Barents sea. Correlation between Iceland/Faeroes and IBWNS is on the other hand low. The results do therefore indicate that the recruitment in east and west is not correlated and indices from east and west can not be used indepenently as index of the total stock, they should rather be added (possibly including a weighting factor) making a total index.

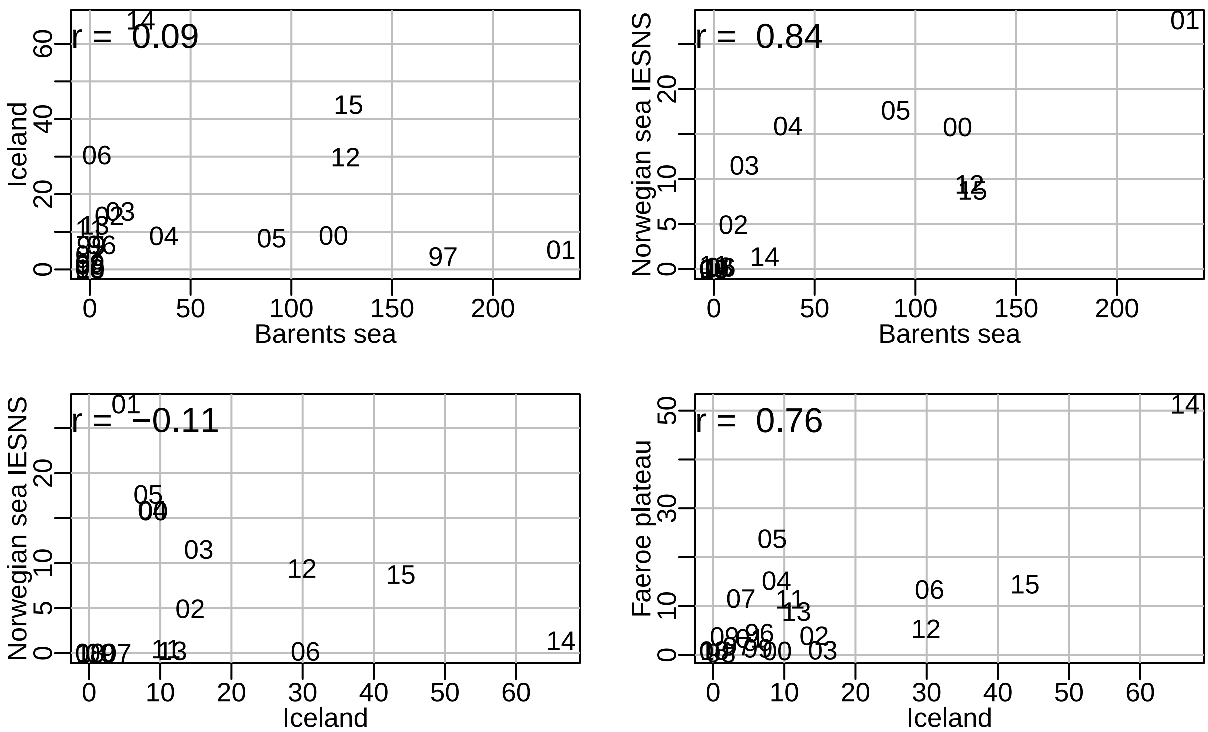


Figure 6: Correlation between indices of age 1 from different surveys

The results above (figure 6) are for the "standard" survey area in IESNS. The survey was expanded in 2006 including the area south and west of Iceland. Acoustic indices from that part correlate better with data from the Icelandic groundfish survey ($\rho = 0.46$), as expected as the area surveyed is partly the same.

One possible way to model recruitment is to use a combined index $I = q_0 + q_1 \times I_1 + q_2 \times I_2$ with q_0, q_1 and q_2 estimated parameters. This equation could be incorportated in an assessment model but was first used here based on results from assessment until 2012 i.e. for years when age 1 should be reasonably estimated. Only surveys with low correlation beween them will be tested. The first combination used is

$I = q_0 + q_1 \times I_{ice} + q_2 \times I_{bar}$. This model can be based on data since 1996 but the results are not too convincing (see figure 7)as mean (q_0) takes over in the forecast. Based on data from 1996-2012 the mean square residual is 218 but 238 based on data 2003-2012 (worse fit!!)

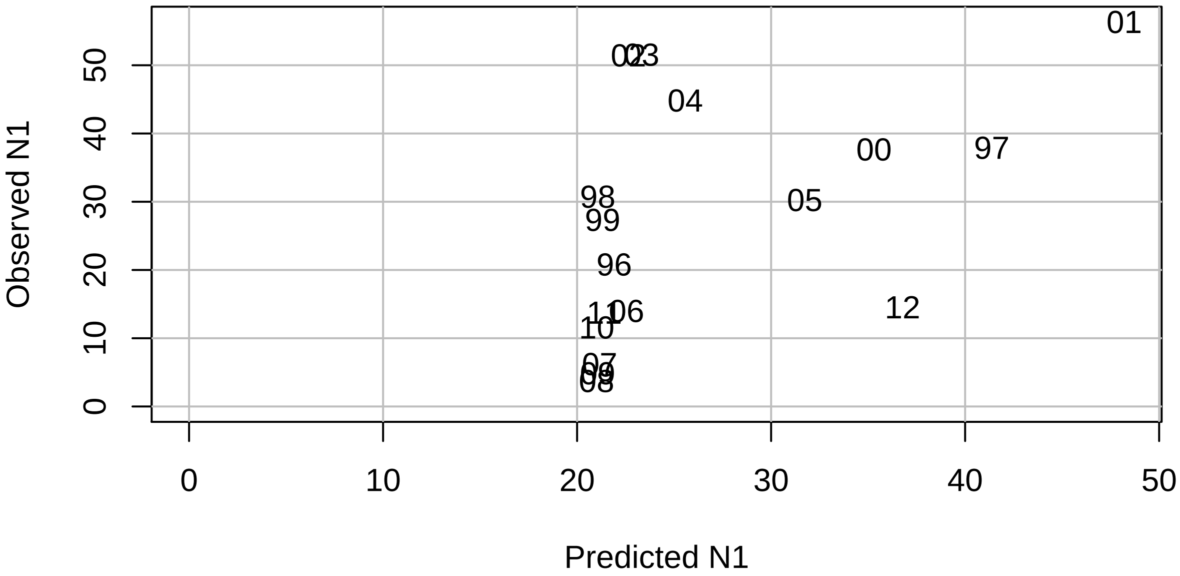


Figure 7: Number of age 1 1996-2012 derived from indices from indices from bottom trawl survey in the Barents sea and Iceland

Using the acoustic survey (IESNS) for the same period (2003-2012), gives a mean square residual of **65** (compared to 238 using 2 bottom trawl surveys) and gives a reasonable fit (figure 8). The intercept is 7.5 that is then minimum yearclass that would be predicted from the model.

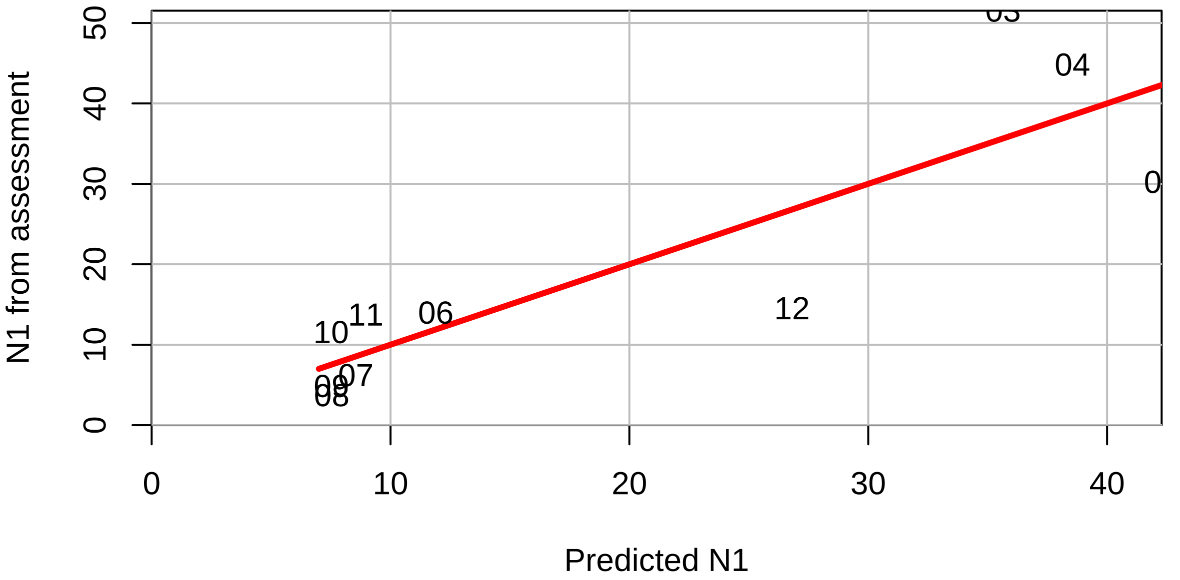


Figure 8: Predicted number of age 1 based on IESNS sea (full area) since 2003 using the equation $I = q_0 + q_1 \times I_{nor}$

Basing the prediction on data from the **standard area** of IESNS leads to somewhat worse fit or mean square residual of 85 compared to 65 for the total area.

The use of intercept term in the equation can be justified but looking at the figure below (figure 9) low index can mean anything from 2 -15 million recruits at age 1. If the regression was implemented in a stock assessment model regression through the origin might be the appropriate form as the stock-recruitment model/shrinkage to the mean (with autocorrelated residuals) has some weight.

Fitting with a model based on IESNS going through the origin leads to mean square of 97 and bias in the estimate of 4.3 milliard fishes. (figure 9). This model might still be appropriate with "shrinkage/ssb-rec" but too high weight of low values (as obtained by using logs) has to be avoided.

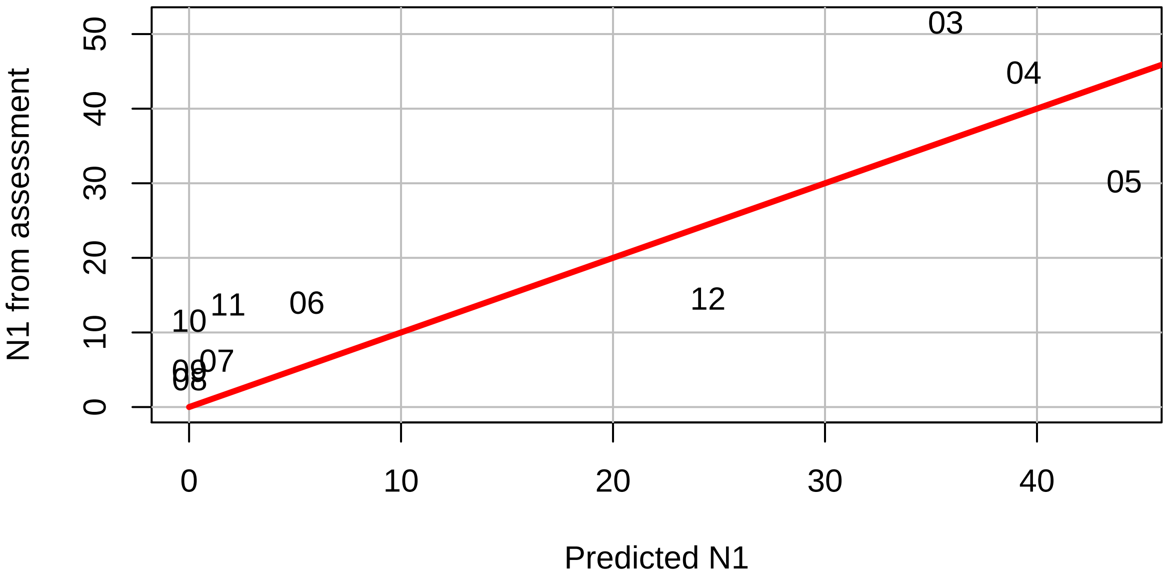


Figure 9: Predicted number of age 1 from the survey in the Norwegian Sea 2003-2012 based on a line going through the origin. Estimated numbers from assessment shown for comparison.

The international blue whiting spawning stock survey(IBWSS) started in 2004. Data from IBWSS in 2010 were are not included in the assessment as results from that year considered "invalid" i.e. too low. (needs to be checked). The correlation between age 1 from IBWSS and age 1 from IESNS is 0.45 but correlation with N_1 from assessment for the period 2004-2012 is 0.82(figure 10) so IBWSS looks like an useful for recruitment survey, even though it is a spawning stock survey.

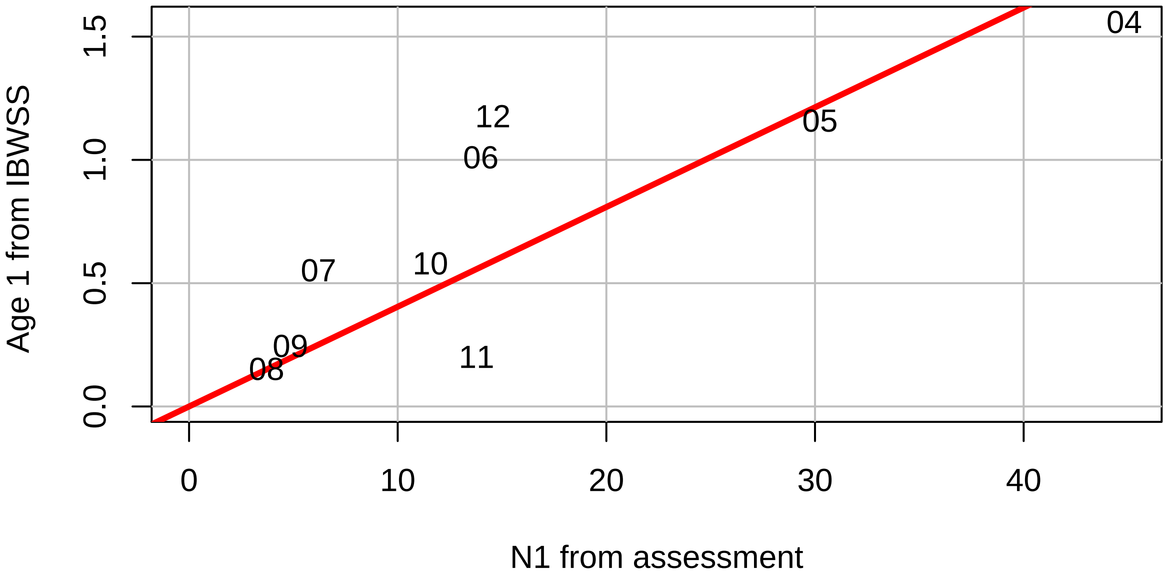


Figure 10: Number of age 1 from IBWSS against estimated number from assessment 2004 - 2012

Adding the two most recent data points reduces the correlation to 0.12, but as shown below (figure 11) the indices for 2014 and 2015 exceed what has been seen since 2004. Earlier it was shown that the index of age 1 from IBWSS in 2014 and 2015 was relatively much higher than what was observed IESNS (figure 4). The Norwegian acoustic survey on the spawning grounds 1991-2003 mightalso be examined for possible recruitment signals.

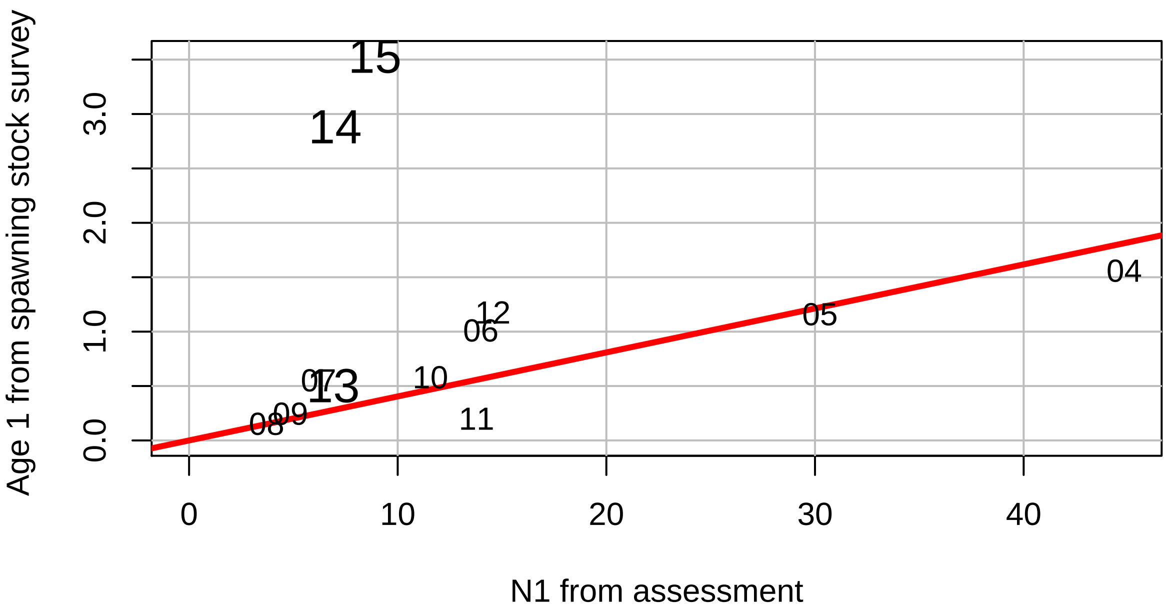


Figure 11: Number of age 1 from the spawning stock survey against estimated number from assessment 2004 - 2015

Looking at age 1 and 2 from IBWSS the index for age 2 in 2015 is in line with what was observed for age 1 in 2014. (figure 12).

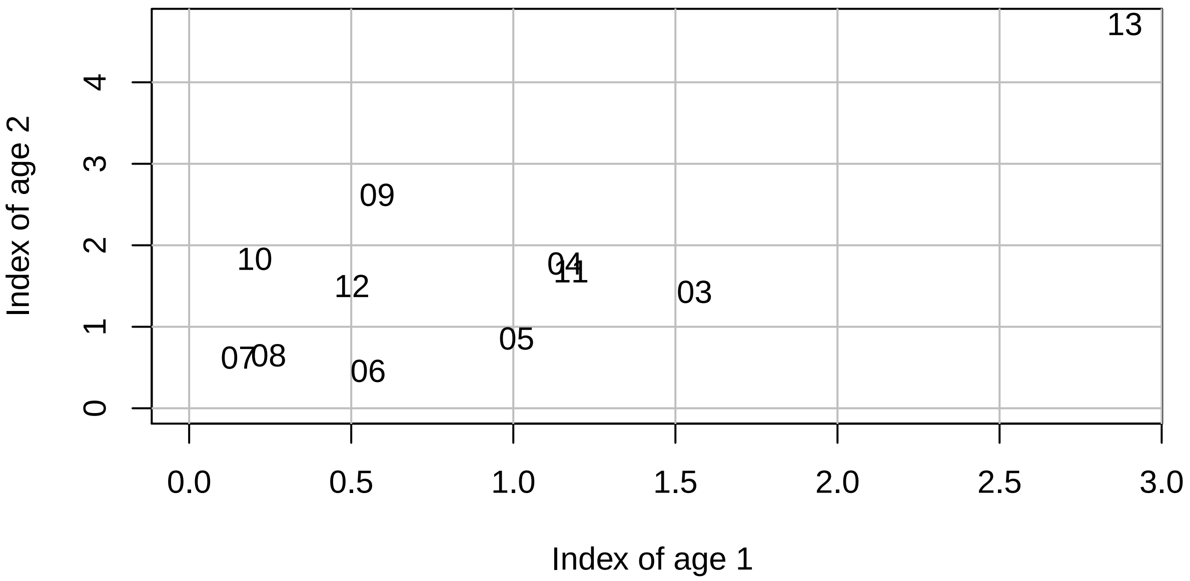


Figure 12: Number of age 2 from the IBWSS against estimated number of age 1 from IBWSS one year earlier. Labels in the figure indicate yearclass.

The result of the investigation of the acoustic surveys is that IBWSS indicates much larger 2013 and 2014 yearclasses than the acoustic survey in the Norwegian sea. The surveys could be used as two seperate measures of the total stock. The spawning stock survey would most likely get high weight as it seems to do well for yearclasses 2003-2012 so yearclasses 2013 and 2014 would be estimated very large.

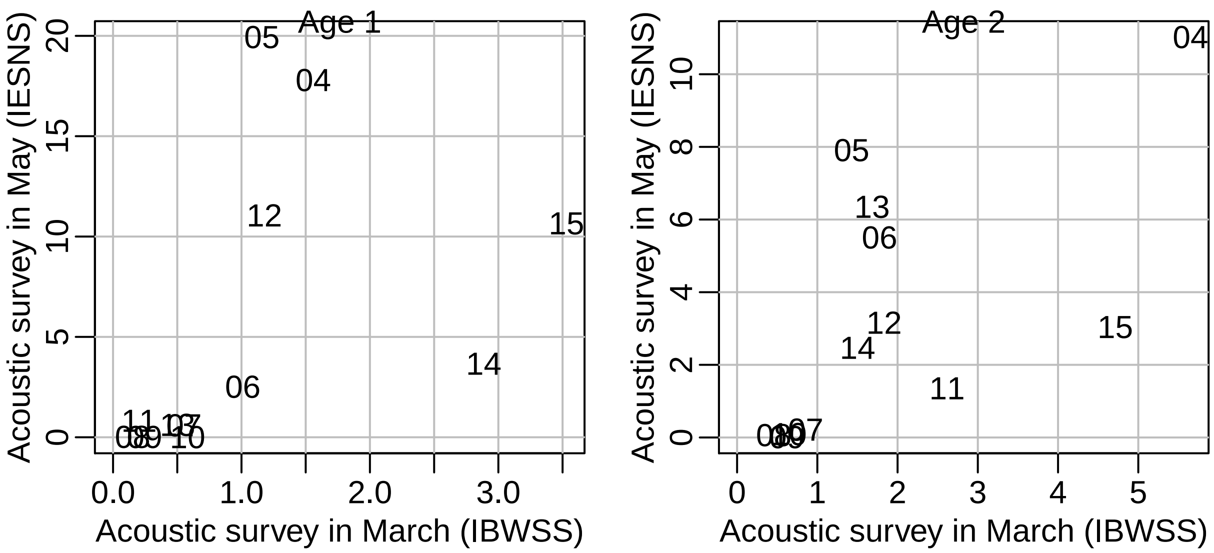


Figure 13: Number of age 1 in IESNS plotted against number of age 1 from IBWSS the same year

The acoustic indices in May (IESNS) are much higher than in March (IBWSS). Both are acoustic indices based on the same TS value, but does it mean that they are comparable and there is more fish measured in the May survey??. Assuming so and that the recruits found on the spawning grounds in March are not found in the Norwegian sea in May, adding the indices from March and May might be the right way to go. The number of age 1 in the May survey is on the average 5 times what it in the March survey, similar for small cohorts but much higher for large cohorts. Might have to do with detection level in the May survey??.

The options are to tune the assessment with survey indices from IESNS (extended area) from 2003 onwards. or to add the indices for ages 1 and 2 from IESNS and IBWSS but the time period would have to be shortened to 2004 onwards one year shorter than when only IESNS is used . Recruitment data from IBWSS in 2010 are used. What was used here is the latter option i.e. use both surveys. The period is rather short for proper retrospective analysis, though some effort was made.

2 Assessment

A separable model was run, tuned with the following settings.

1. Ages 1-10 age 10 a plus group.
2. M = 0.2 for all age groups.
3. Tuning data were indices of age 3-8 from IBWSS and age 1 and 2 from IBWSS+IESNS. 2010 data for IBWSS are not used for ages 3-8 but are included for ages 1 and 2.
4. Correlation between residuals in IBWSS was sometimes modelled but in other cases each age group rather was treated as seperate index.
5. Catchability in surveys estimated for ages 1-6 but assumed to be the same for ages 7 and 8 as for age 6.
6. CV of survey residuals estimated, one number for each survey. Pattern with age specified.
7. The residuals minimized in survey are $log(I + \delta) - log(\hat{I} + \delta)$ where delta is a relatively small number corresponding to 3-5 otholits or so.
8. 50% of the annual F on the spawning stock is assumed to occurr before the survey. This number was estimated from landings data in the first 3 months taking into account that higher proportion of the catch late in the year is not part of the SSB in March the same year. Should be replaced later by catch in numbers by age before the survey.
9. Prediction is based on 1200 thous. tonnes catch in 2015 and speicifed F (often 0.3) after that.
10. Autocorrelation of recruitment 0.3. Does not affect assessment result compared to correlation of 0.6 that was used in HCR simulations but affects short term prediction

The age range and M are just taken from the standard assessment.

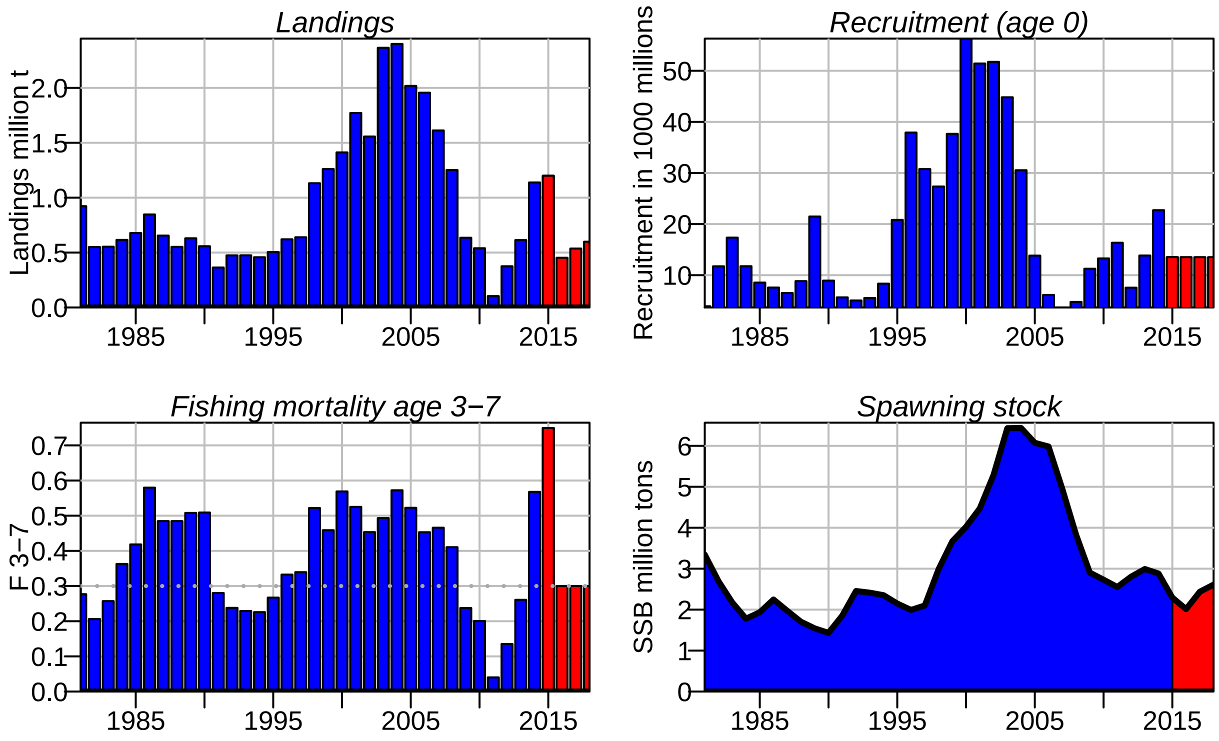


Figure 14: Summary of assessment when correlation of residuals in the survey is not modelled

Results from the assessment model are shown in figure 14. The spawning stock is predicted to stay at relatively low level in coming years even though recruitment is improving, mostly because of very high fishing mortality in 2014 and 2015. The spawning stock in 2016 is predicted to be 2012 thous. tonnes with standard error of 531 thous. tonnes or 26.5%, rather high value as a result from catch at age model that tend to underestimate uncertainty.

One of the purposes of this study was though to see if the surveys were a useful measure of recruitment. Estimated CV of the recruitment estimates is 0.62 and of recruitment 0.80 so the recruitment survey has some weight but the results are pulled towards the geometric mean. Autocorrelation of recruitment is very high, or 0.75 on logscale (lag 1 year) and 0.8 on normal scale

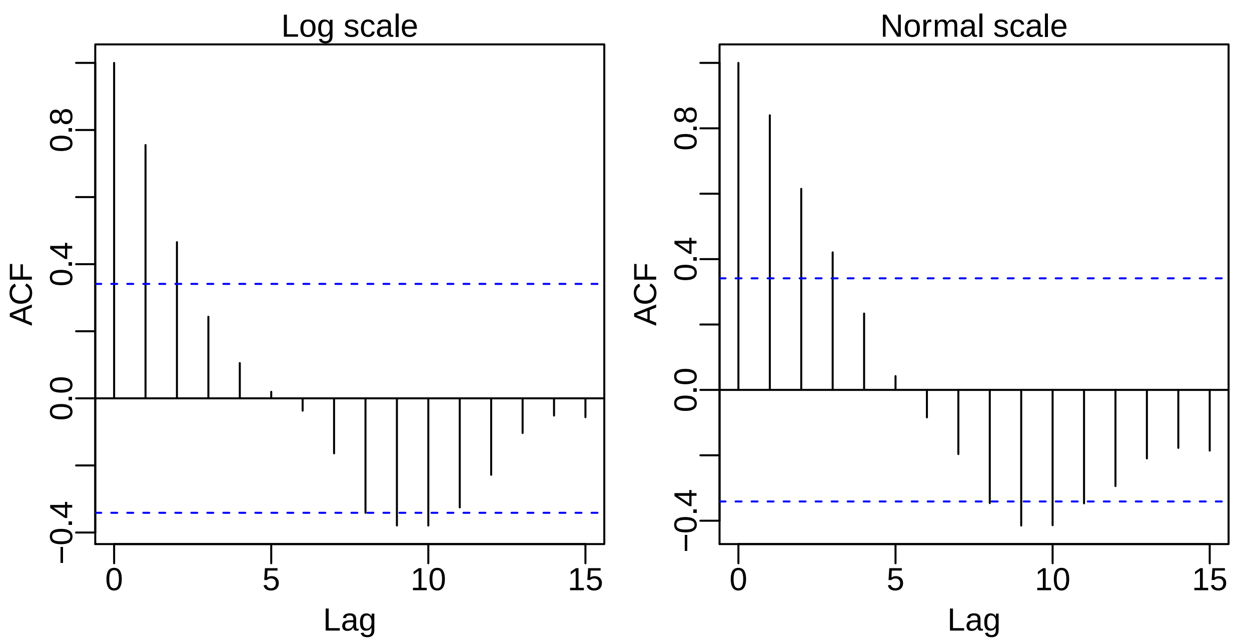


Figure 15: Autocorrelation of recruitment

A first order AR model for recruitment explains 70% of the variance in the model, more than the survey. The AR model does though not save everything as yearclasses **y-1, y-2** etc are not known when yearclass **y** is estimated. Also the autocorrelation of the error in this kind of estimate is higher than an estimate based on survey indices. Including the autocorrelation in the recruitment model would on the average improve recruitment estimates but might be problematic in transition periods, i.e. delay the response from low to high and high to low. Short data series will also limit the possibility of testing.

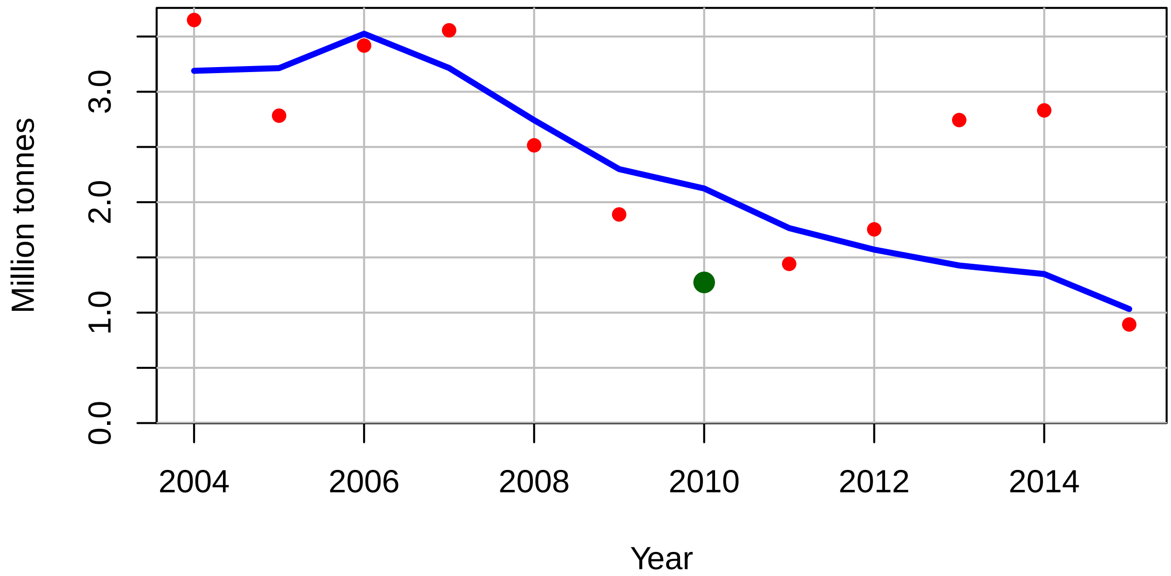


Figure 16: Observed and predicted survey biomass from IBWSS. The 2010 survey is shown as green large point.

The assessment presented here gives a more pessimistic view than the assessment used for generating advice this year, the main difference is inclusion of the 2015 IBWSS survey that is practially ignored in the "official assessment". Setting the SAM model up so it has to "listen to" the 2015 survey leads to similar result as shown here. The model presented here follows the 2015 survey but the 2013 and 2014 surveys look too high (figure 16. The 2010 survey, not used in the tuning is closer to predicted values than the 2013 and 2014 surveys. (figure 16)

The advice for 2016 of 750 thous. tonnes leads to $F_{3-7} = 0.593$ and spawning stock around 2 million tonnes until 2018 assuming 500 thous tonnes catch in 2017. Above average recruitment from yearclasses 2013 and 2014 leads to the stock not decreasing in spite of relatively high fishing mortality. The model could be underestimating these yearclasses so the advice given might be according to MSY, PA and everything else.

Calculated from the Hessian matrix the standard deviation of F_{2016} using a catch constraint of 750 thous. tonnes is 0.22 or 35%. MCMC simulations give higher value or close to 50%. Autocorrelation of assessment error is not easily obtained but periods of over and underestimation have been seen here so autocorrelation of 0.5 or higher could be expected.

Uncertainty in the predicted spawning stock is relatively high as shown in the figure below that shows the probablitity distribution of SSB_{2017} with TAC constraint of 1200 thous. tonnes in 2015 and 750 thous. tonnes in 2016. Medium value of the spawning stock is around 2700 while the 5th and 95th percentiles are 1088 and 5500 thous. tonnes. The median value is higher than the median value of SSB_{2017} with $F_{2016} = 0.3$. The reason for the difference is not clear but stochastic projections based on catch constraint can be problematic when F is as high as it is here.

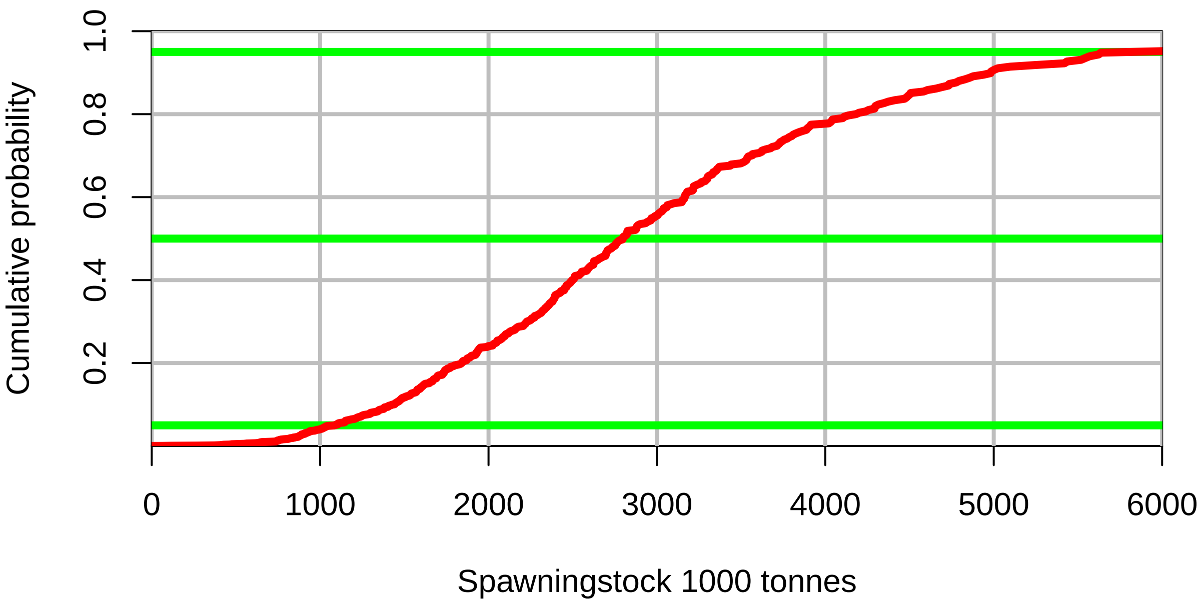


Figure 17: Cumulative probability of the spawning stock in 2017 using catch of 1200 thous. tonnes in 2015 and 750 thous. tonnes in 2016

An important thing is to look at how the model performs historically. The tuning series are very short or since 2004 so limited retrospective patterns can be investigated. The retrospective pattern from the run for assessment years 2008-2015, shows considerable variability (figure 18). This exercise might be exaggerating the variability as the survey series are very short in the beginning.It must also be stressed that the retros show prediction 2 years ahead i.e. until the 2 years after the assssment year, based on the observed catches. Retros with shorter prediction are also shown but the spawning stock two years after the assessment year a result of the catches in the advisory year and therefore a test on the quality of the assessment to generate advice.

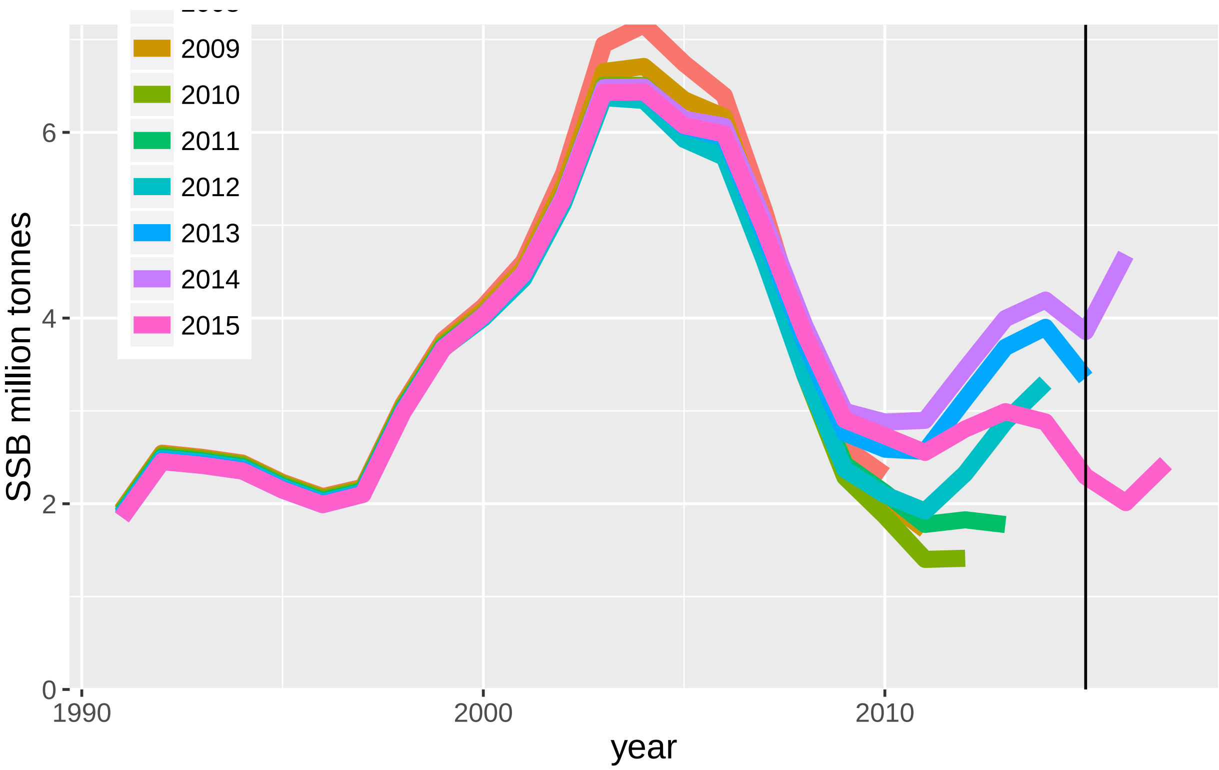


Figure 18: Retrospective pattern of spawning stock from the separable assessment model not including correlation of residuals in the survey. Each line terminates 2 years after the assessment year.

Figure 19: Retrospective pattern of recruitment from the separable assessment model not including correlation of residuals in the survey. Each line terminates 2 years after the assessment year.

A retrospective pattern terminating year after the assessment year looks a little smoother (figure 20). Of course the retrospective pattern of spawning stock 3 and 4 years after the assessment year should also be shown to prepare the ground for bi and triannual advice. Those pictures look rather noisy and are therefore not shown here.

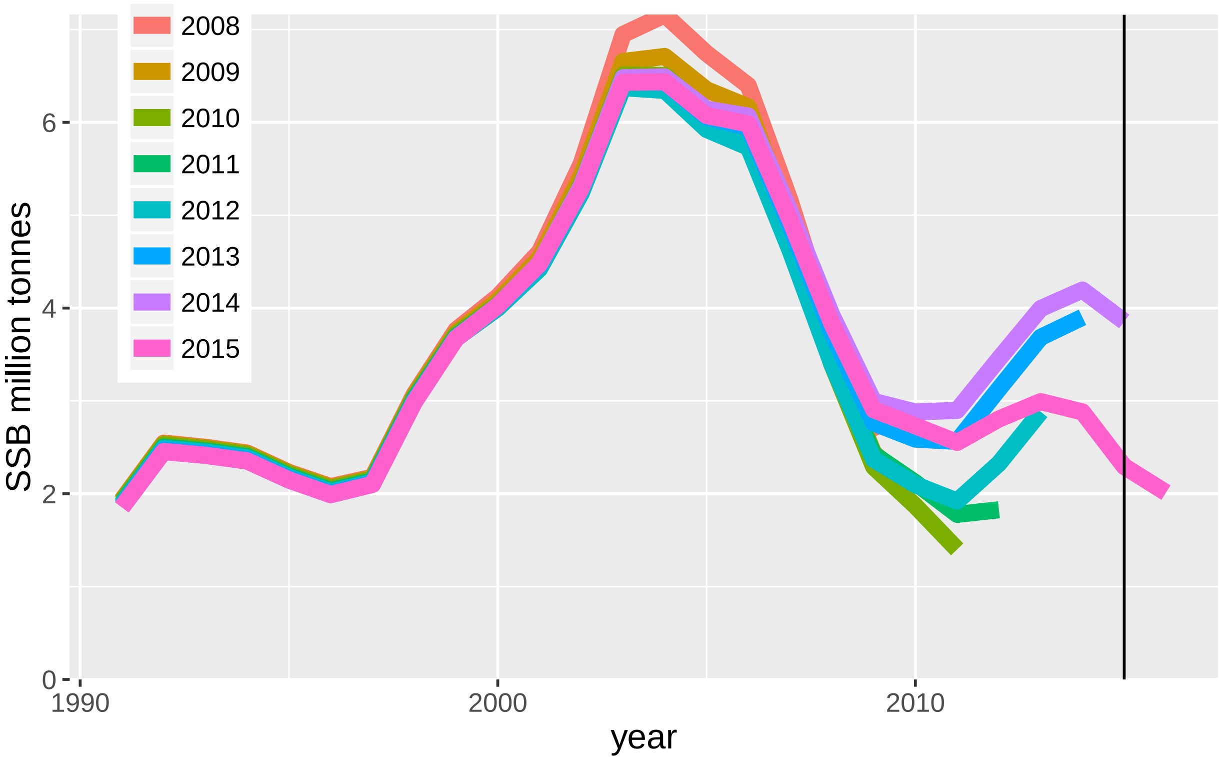


Figure 20: Retrospective pattern of spawning stock from the separable assessment model not including correlation of residuals in the survey. Each line terminates 1 year after the assessment year.

Including estimated correlation between residuals in the surveys leads to more stability in the retrospective pattern (figures 21 and figure 22. The residuals are modelled with a first order AR model withe the correlation between ages i and j given by $\rho^{|i-j}$. The value of ρ is estimated. The estimated value is rather high (0.75), reducing the weight of the survey. Ages 1 and 2 are assumed uncorrelated with the rest.

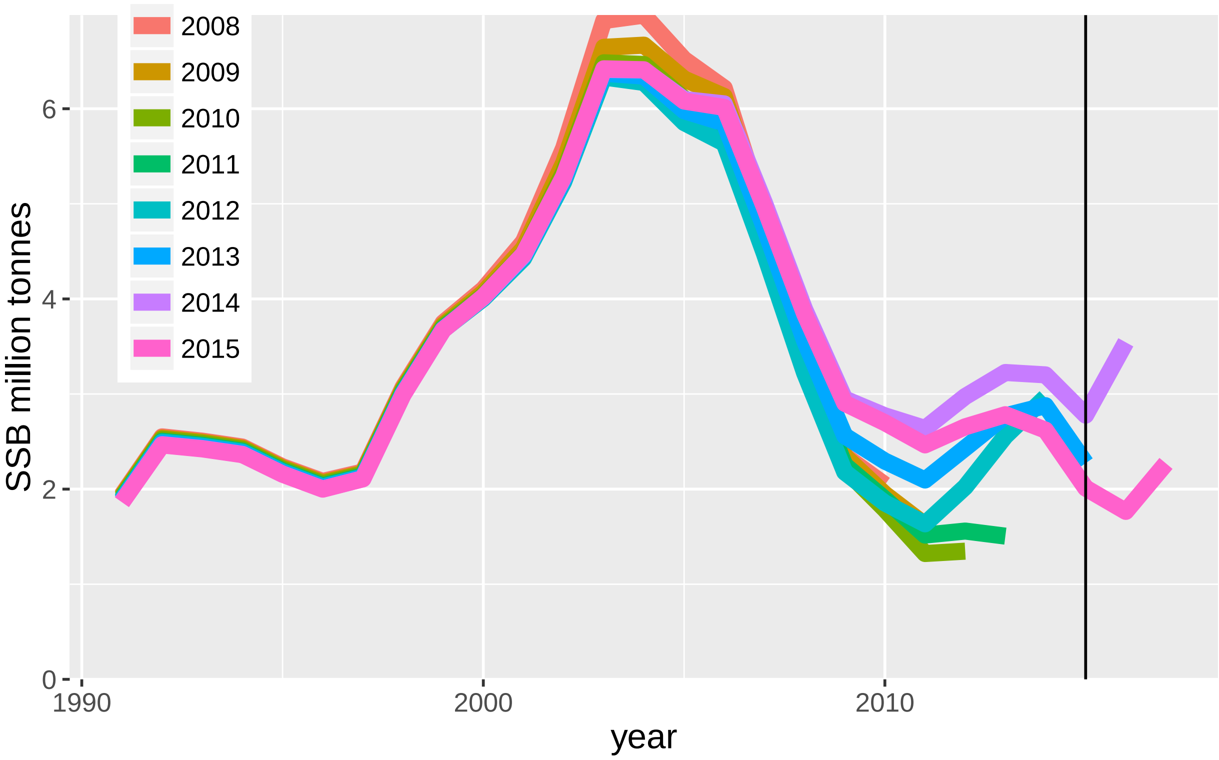


Figure 21: Retrospective pattern of spawning stock from the separable assessment model including correlation of residuals from the survey. Each line terminates 2 year after the assessmentyear.

Figure 22: Retrospective pattern of recruitment from the separable assessment model including correlation of residuals from the survey. Each line terminates 2 years after the assessmentyear.

Retro of SSB terminating a year after the assessment year is shown below (figure 23)

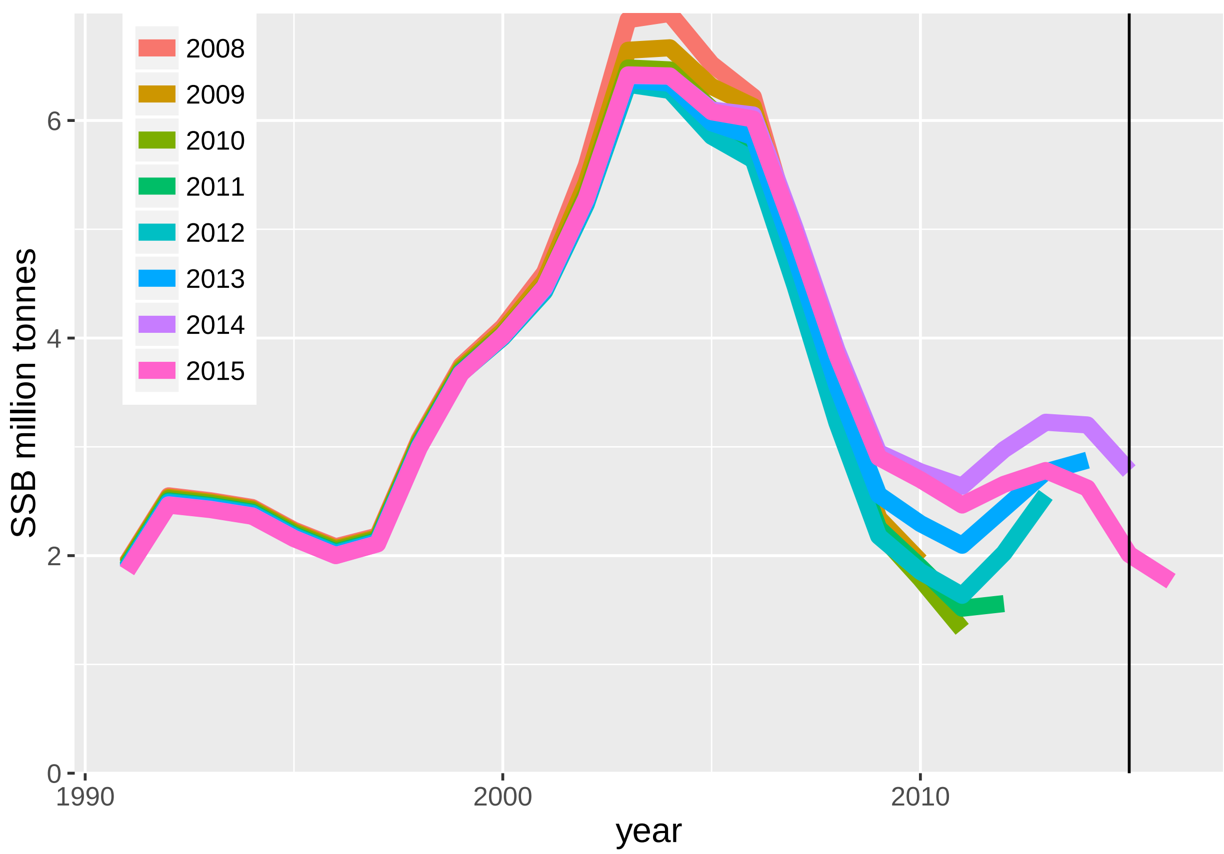


Figure 23: Retrospective pattern of recruitment from the separable assessment model including correlation of residuals from the survey. Each line terminates one year after the assessmentyear.

To show a nice retrospective pattern, one that terminates in the assessment year is shown below (figure 24). The purpose of this kind of retrospective pattern is unknown except the advice is generated directly from a biomass in the assessment year.

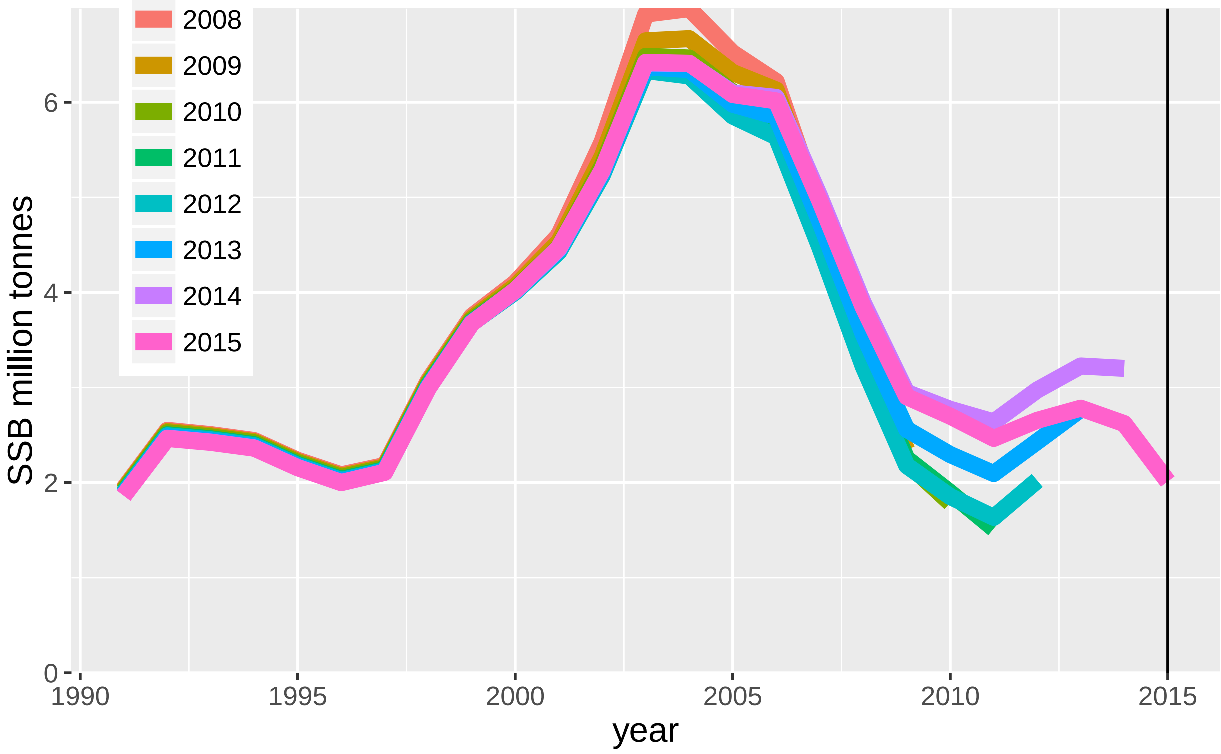


Figure 24: Retrospective pattern of recruitment from the separable assessment model including correlation of residuals from the survey. Each line terminates in the assessmentyear.

	2008	2009	2010	2011	2012	2013	2014	2015
2008	3.46	3.44	3.31	3.31	3.22	3.56	3.89	3.84
2009	2.35	2.34	2.21	2.25	2.18	2.56	2.95	2.90
2010	2.05	1.95	1.79	1.89	1.85	2.29	2.77	2.70
2011		1.62	1.32	1.52	1.63	2.10	2.64	2.46
2012			1.35	1.56	2.02	2.43	2.97	2.66
2013				1.51	2.56	2.77	3.23	2.78
2014					2.96	2.88	3.20	2.62
2015						2.27	2.78	2.01
2016							3.54	1.77

Table 1: Spawning stock 2008-2016 for assessment years 2008-2015 million tonnes. Correlation of residuals modelled.

The retrospective runs are only conducted for 9 years, i.e. assessment 2007-2015. Four or more estimates are available for the spawning stock 2008 - 2014 but what is of interest is quality of the estimate of the stock in the year after the assessment year (start of advisory year) or even one year later. The first estimates of SSB in this period have usually been underestimates as shown in table 1 (2010-2012). An advice given in those year based on fixed F would have been approximately 60% of what looks now as the "correct advice" if advice was based on constant F but lower if the estimated stock was below a trigger point.

The model not including correlations of residuals overestimates the stock more in in 2013 and 2014 than the model modelling the correlation. The quality of estimates in the underestimation period (assessment years 2009-2011) is similar (figures 18 and 22).

The model including correlation estimates SSB_{2015} as 2 million tonnes but the model not including the correlation as 2.25 million tonnes. Last 2 years the difference was much larger as the increase in the survey 2013 and 2014 was ignored when the correlations of residuals were modelled.

The figure below (figure 25) summarizes this a little bit by showing ratio of first and last estimate of SSB each year. The first estimate is in the assessment year before the year and last estimate in 2015. For the year

2016 only one point is available (of course that point is correct as indicated by the ratio 1 !!!).

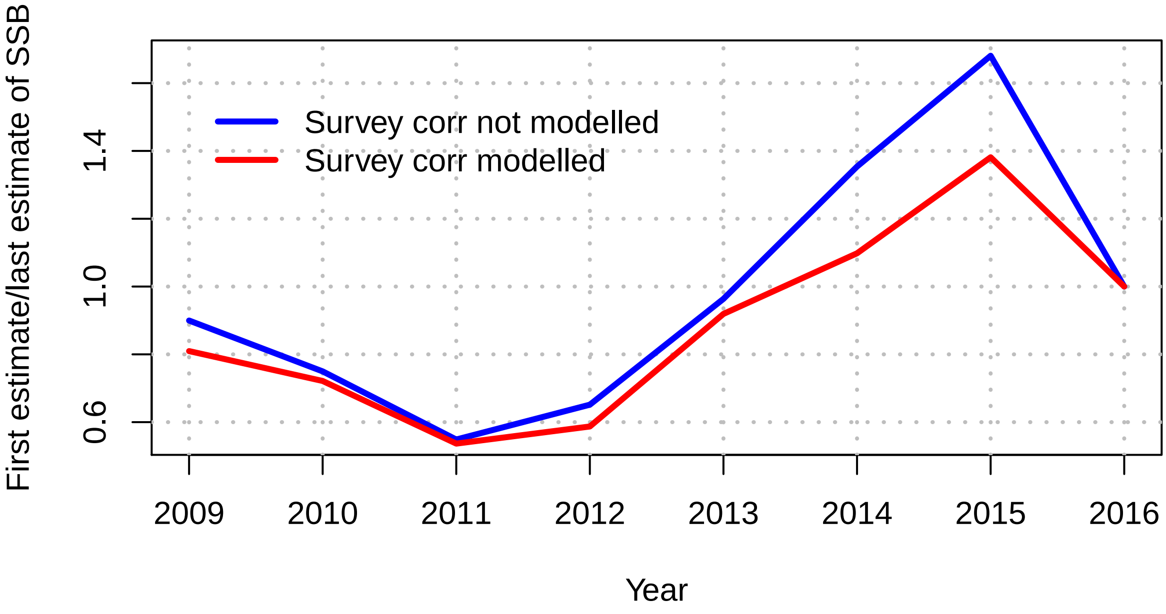


Figure 25: Last estimate of SSB each year as proportion of first estimate

3 Harvest control rule simulations

Harvest control rule simulations are presented in this section using the assessment models presented in last section plus stochastic recruitment, assessment error and mean weight at age. Selection and maturity at age are fixed.

Mean weight of most yearclasses follows the same pattern (figure 26). Lowest mean weight at age is observed in the years when the stock is relatively large but is only noticed for age 3-4 and older. Selection is most likely hiding the change for younger age groups.

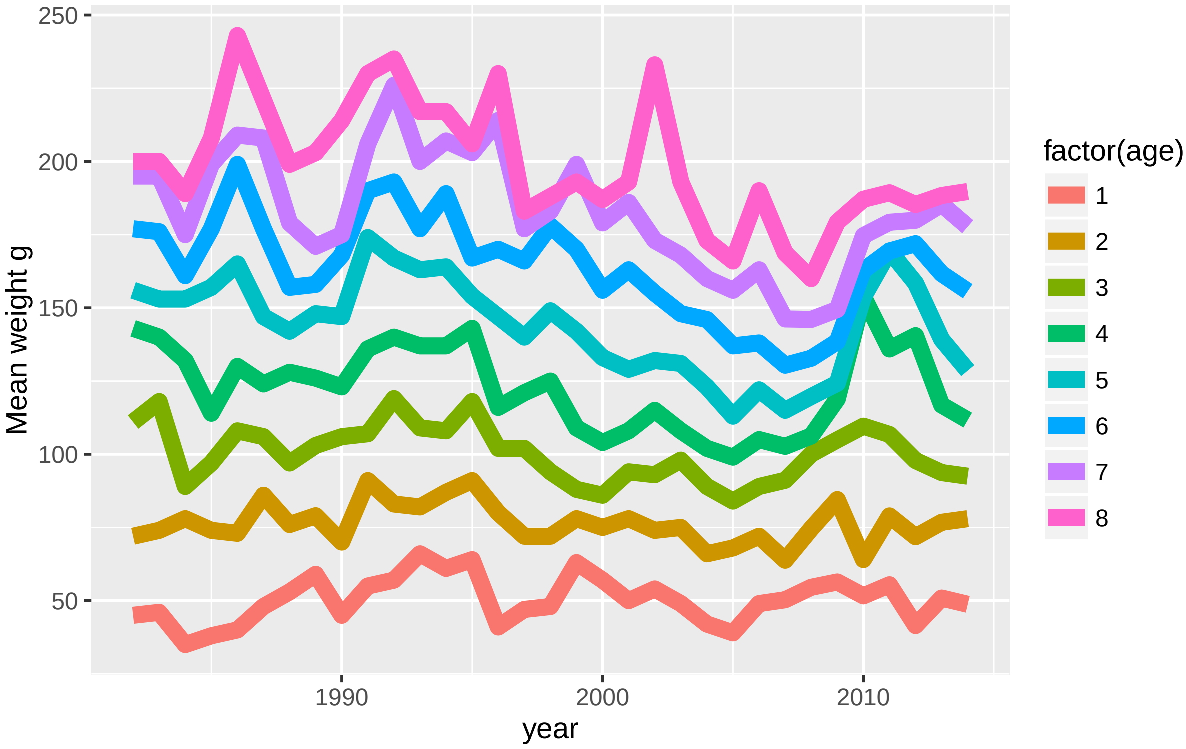


Figure 26: Mean weight at age for ages 1-9

Looking at the most important age groups (3-8) the correlation between age groups is

	3	4	5	6	7	8
3	1.00	0.79	0.67	0.58	0.54	0.51
4	0.79	1.00	0.83	0.66	0.52	0.42
5	0.67	0.83	1.00	0.90	0.77	0.65
6	0.58	0.66	0.90	1.00	0.90	0.76
7	0.54	0.52	0.77	0.90	1.00	0.80
8	0.51	0.42	0.65	0.76	0.80	1.00

Table 2: Correlation between mean weight of diffrent agegroups

Mean weight at age for ages 3-8 was modelled by a multipicative model $W_y = \delta_y \times \gamma_a$ where y is year, a age, and δ and γ estimated factors. The model is estimated as additive model by log transformation.

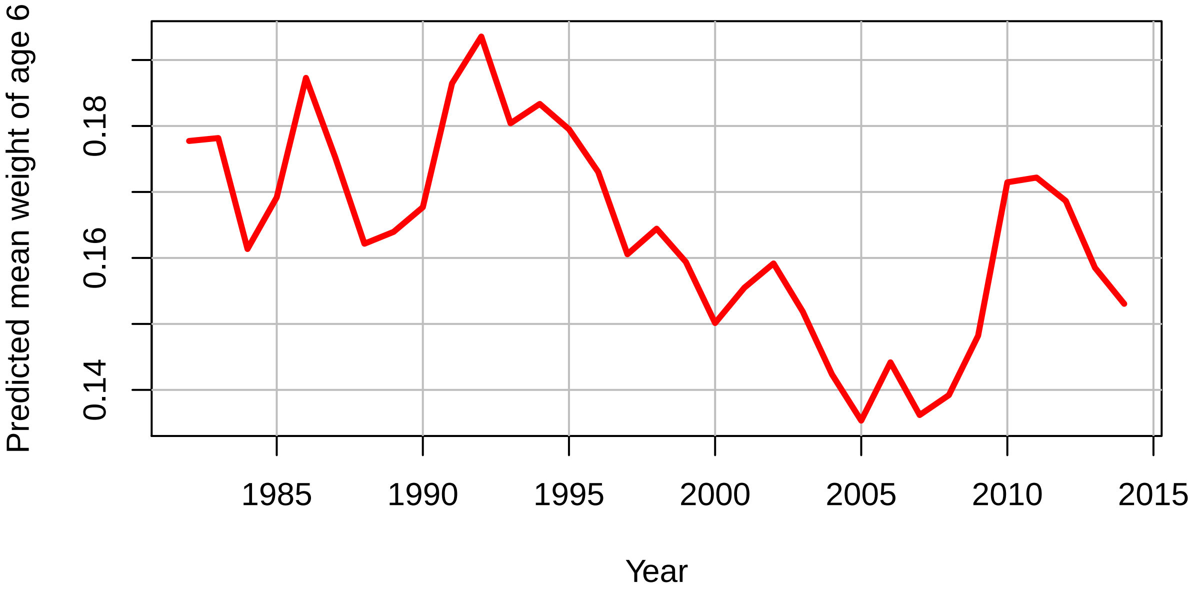


Figure 27: Plot of the yearfactor in the weight model scaled to age 6. $\sigma = 0.095$, $\rho = 0.8$, correlation with ssb = -0.7.

One good measure of mean weight at age and growth is yield and spawning stock per recruit as function of time. Reference points like F_{max} and $F_{0.1}$ can be estimated from those data. F_{max} is of no value for blue whiting as the gain in weight is less than assumed natural mortality, already at age 2 so maximum yield is obtained by fishing a yearclass up immediately. Maximum yield and spawning stock per recruit in 1995-1999 when they are lowest are only 75% of the highest values in 1991-1995. (figure 28). The yield at F=0.3 is about 90% of the maximum yield at $F = 1$ (the highest F used here).

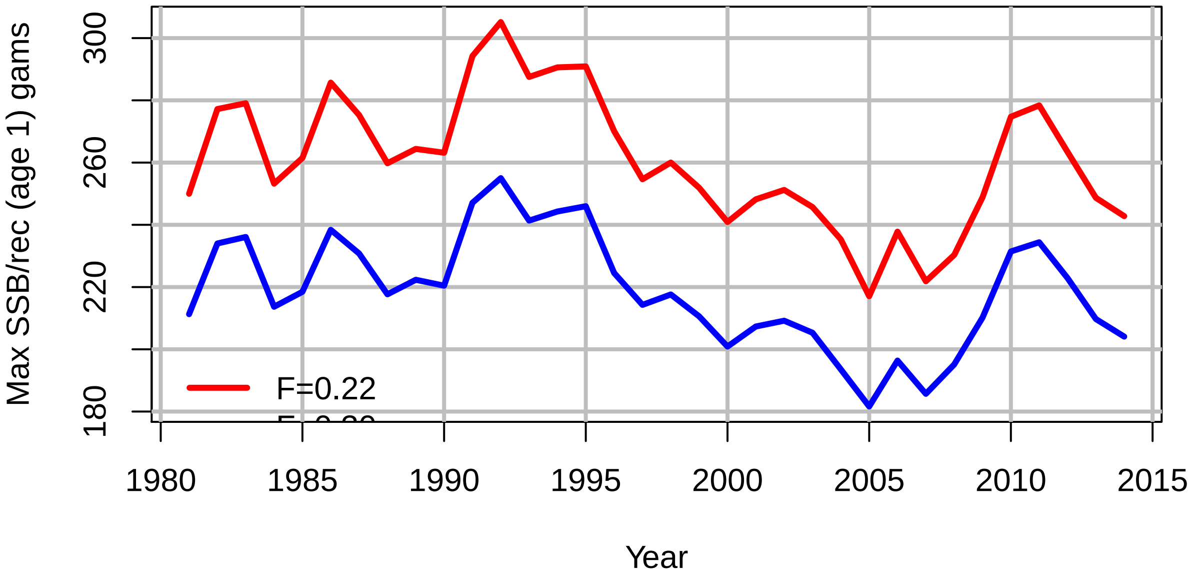


Figure 29: Spawning stock per recruit at F=0.3 and F=0.22

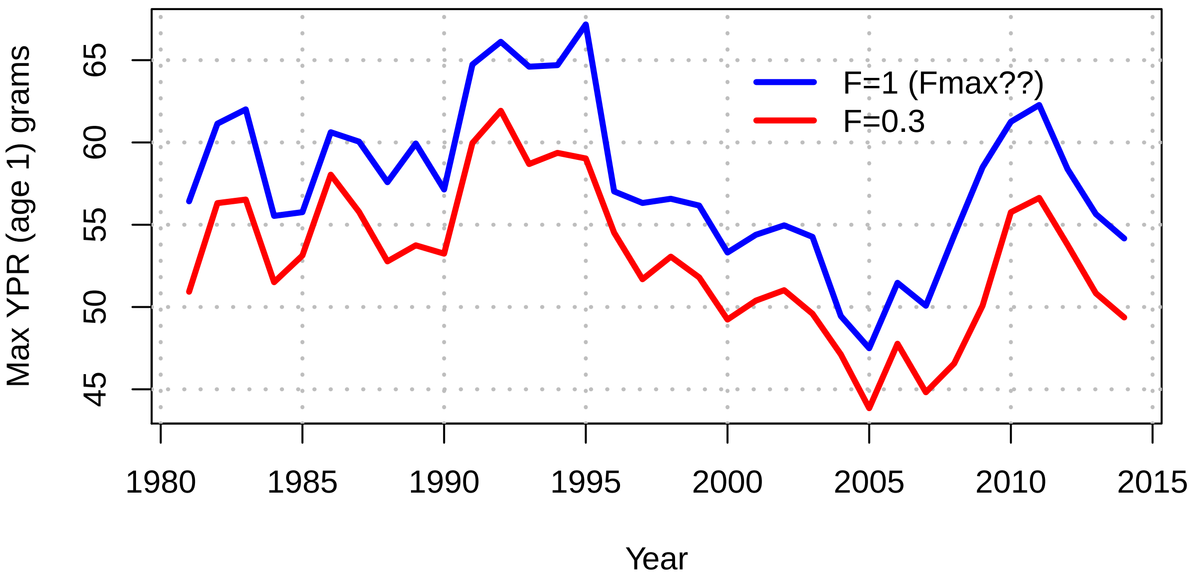


Figure 28: Maximum yield and yield at F=0.3

$F_{0.1}$ shows some varibility as function of time (figure 30) but is always close to M (0.2) as it usually is. The value 0.22 is close to $F_{0.1}$ that has always been considered a reasonable, conservative reference point which it does not have to be as M is most often assumed.

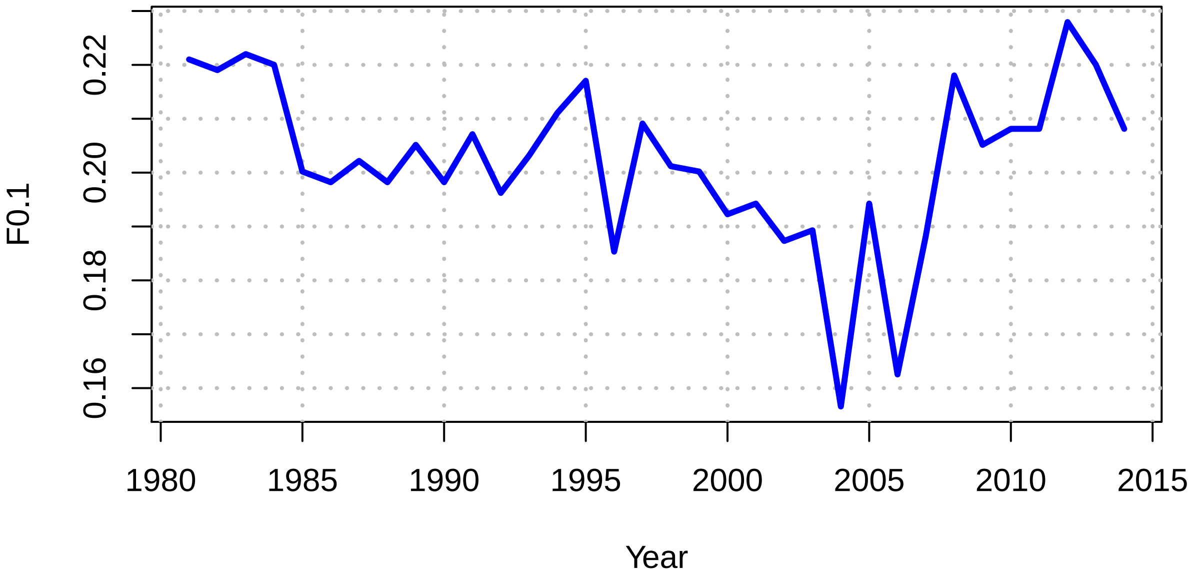


Figure 30: F_{01} as function of time

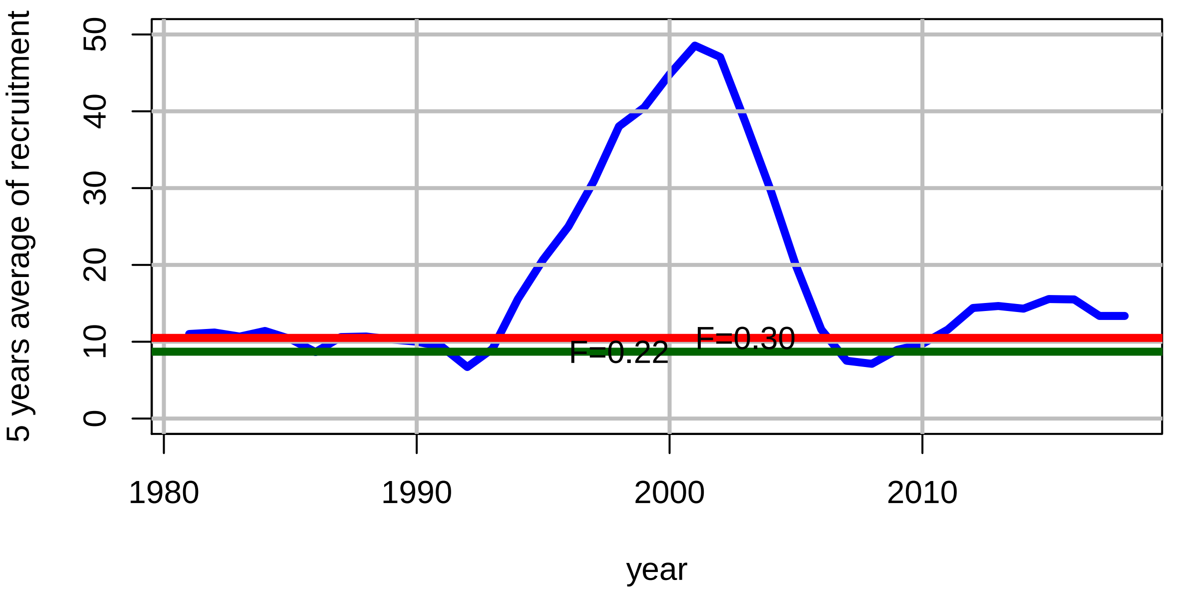


Figure 31: Five year average of recruitment and number of recruits needed to get SSB over 2250 for $F = 0.3$ and $F = 0.22$

Looking at figures 28 and 32 negative correlation between mean weight at age and stock size is suspected. Spawning stock per recruit with $F = 0.22$ is close to 260 grams in periods of poor recruitment so the average number of recruits over 5 year period (5 cohorts in the spawning stock) must not be less than 8.7 million age 1 individuals (green line in figure 31) to keep the stock above B_{pa} of 2250 thous. tonnes. If F=0.3 the average recruitment over a 5 years period would have to be 10.5 million individuals (red line in figure 31). The figure does indicate that F_{pa} is between 0.22 and 0.3.

To infer about possible effect of density dependent growth on estimated F_{pa} spawning stock per recruit scaled to an average of 1 is plotted (figure 32)

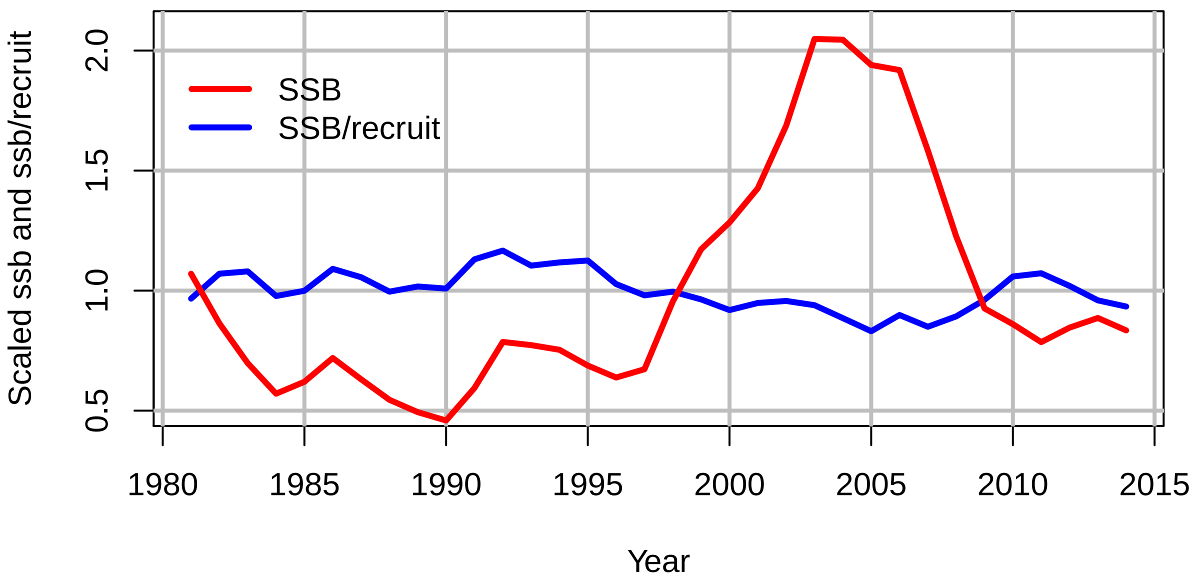


Figure 32: Spawning stock per recruit and spawning stock 1981-2015, scaled to an average of 1

Spawning stock per recruit when $SSB_{relative} < 0.7$ is 1.036. An increase of 1.036 can also be obtained by reducing fishing mortality by 0.015. The effects of including density dependent growth would therefore be to increase F_{pa} by approximately 0.015.

As shown above(figure 15) autocorrelation of recruitment is substantial for this stock or 0.75-0.8 on log scale. This is of course caused by one long period of relatively good recruitment (yearclasses 1996-2004). If the correct model of the recruitment was 1st order AR with autocorrelation of 0.8 the timeseries available would be very short for any real inference (figure 33).

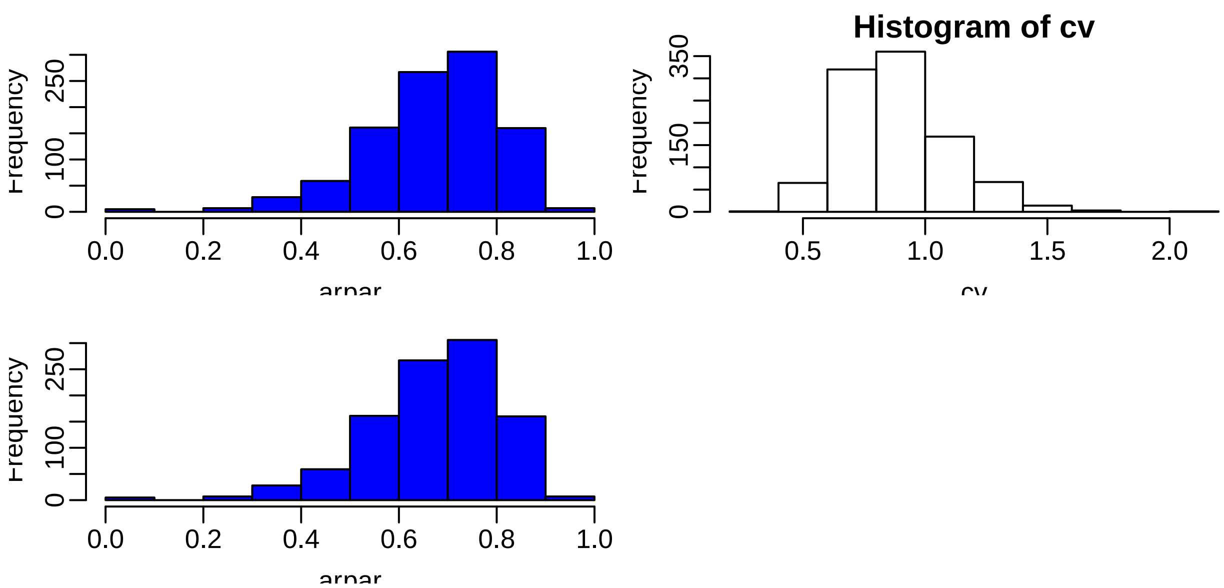


Figure 33: Inference of CV and ρ based on a number of random traws of length 35 from a timeseries with $\sigma = 1$ and $\rho = 0.8$

The first order AR model is certainly not the right model for autocorrelation, but the "right" model is not known. Modelling the distribution of recruits as lognormal might also be questionable, with small yearclasses somewhat larger than predicted by a lognormal distribution (figure 34.

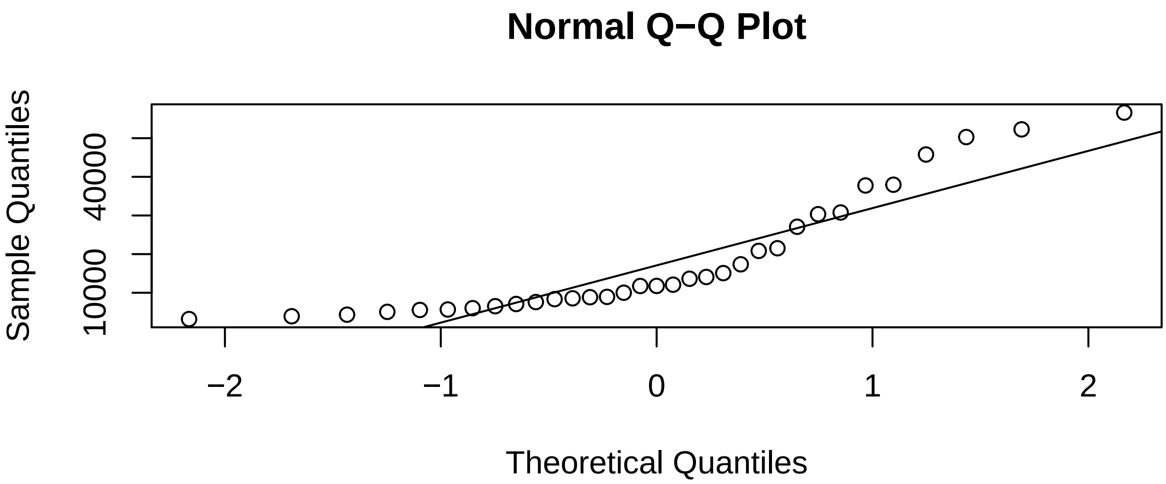


Figure 34: qqplot comparing log of “observed recruitment” and quantiles of normal distribution

The stochastic simulations presented here below will be based on the following premises.

1. Log of assessment error modelled as first order AR process. Default values will be $\sigma = 0.4$ and $\rho = 0.6$. $\sigma = 0.4$ leads to bias of $e^{\frac{0.4^2}{2}} = 1.08$. What is referred to as assessment error is uncertainty in F in the advisory year for a given TAC.

2. Mean weight at age modelled as lognormal yearfactor with CV of 0.1 and autocorrelation of 0.7.

3. Recruitment is modelled as first order AR process. Default values will be $\rho = 0.6$. σ is estimated.

4. Spawning stock - recruitment is modelled as Hockey-stick function with 3 estimated parameters, R_{max}, SSB_{break} and $sigma$. The uncertainty in the estimated parameters enters into the stochastic simulations.

5. Maturity and selection at age are fixed.

CV of recruitment is estimated between 0.65 and 0.8 on logscale (highest with no autocorrelation). As sources of uncertainty are uncorrelated variances are added and autocorrelation of recruitment (pattern of recruitment) is the most important factor in the model.

An attempt was made to estimate the first order autocorrelation of recruitment (4th parameter in SSB-R function) letting uncertainty in this parameter also into the stochastic simulations. Lead to a relatively large number of runs with the parameter close to upper bounds, that needed to be arbitrarily specified and therefore some risk of very long period of poor recruitment if the upper bounds were too high.

4 Results from simulations

As described above the base case was that autocorrelation of assessment error and recruitment variability was 0.6 for both. Other values were tested with the results summarized in figures 35 and 36.

The figures show average catch and fifth percentile of spawning stock when the effects of initial condition have disappeared. Maximum of average catch might be an indication of what could be called F_{msy} and the value of F_{target} where $SSB_{05} = B_{lim}$ might be called F_{pa}.

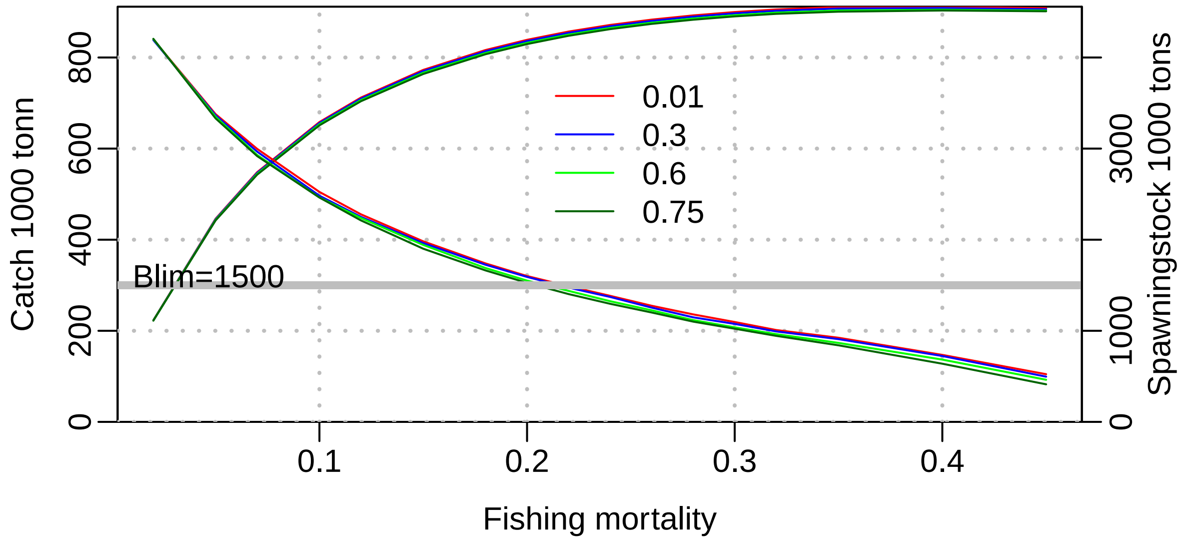


Figure 35: Average catch and fifth percentile of spawning stock as function of fishing mortality for various values of autocorrelation of assessment. Autocorrelation of recruitment 0.6

The effects of autocorrelation of the assessment error are not large (figure 35) but effects of correlation of recruitment variability are much larger (figure 36).

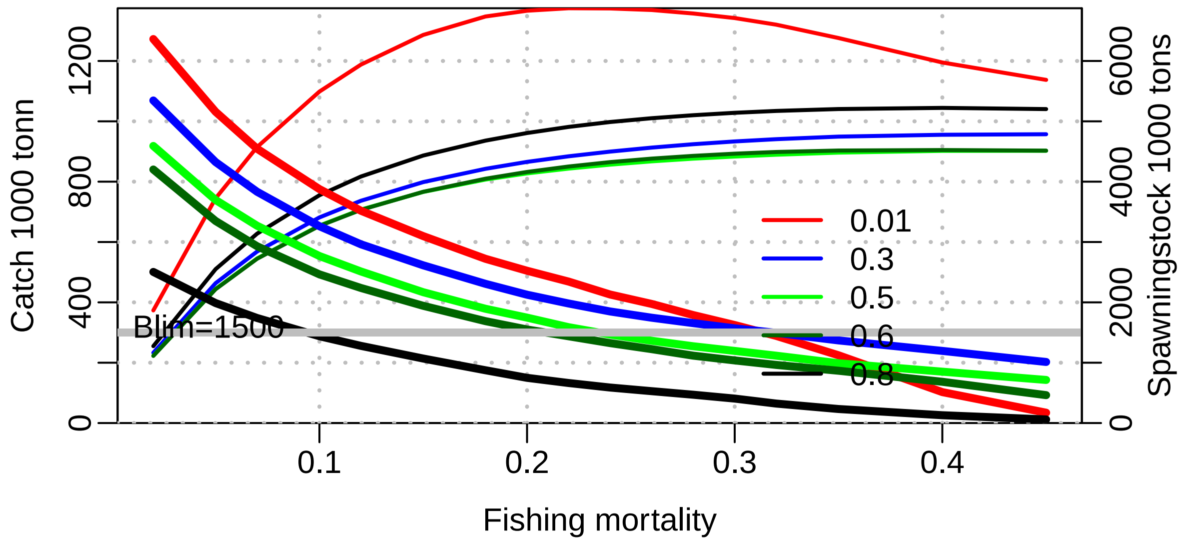


Figure 36: Average catch (thin lines) and fifth percentile of spawning stock (thick lines) as function of fishing mortality for various values of autocorrelation of recruitment variability . Autocorrelation of assessmenterror 0.6

When the effects of autocorrelation of recruitment are tested the assessment is run with the assumed value.

The most interesting feature is the difference between the run assuming no autocorrelation of recruitment and the other runs. In that case a relatively well defined peak in catch is seen at $F = 0.21$, that would then be defined as F_{msy} as SSB_{05} is well above B_{lim}.

What happens here is that SSB_{break} is estimated relatively high ($\approx$ 2.2 million tonnes). The reason is that the period of "good" recruitment is long enough to give a number of "high ssb-good recruitment" pairs. The change in mean catch with ρ is caused by increased uncertainty about R_{max} when the autocorrelation increases that effectectively shortens the series. But the main feature is that for a given F SSB_{05} increases with increased autocorrelation.

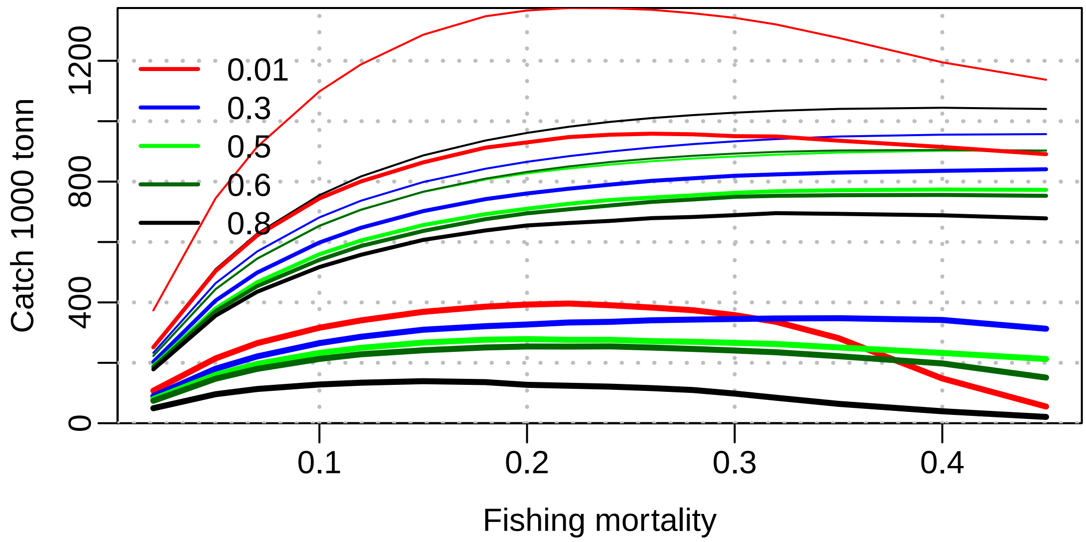


Figure 37: Average catch (thin lines), median catch and fifth percentile (thickest lines) of catch as function of fishing mortality for various values of autocorrelation of recruitment . Autocorrelation of assessmenterror 0.6

The "base case" with recruitment correlation of 0.6 leads $F_{pa} = 0.21$. (figure 38). This fishing mortality lead to average spawning stock of 4400 tonnes and median around 3800 tonnes. Adding the estimated effect of density dependent growth (figure 32) would lead to $F_{pa} \approx 0.225$.

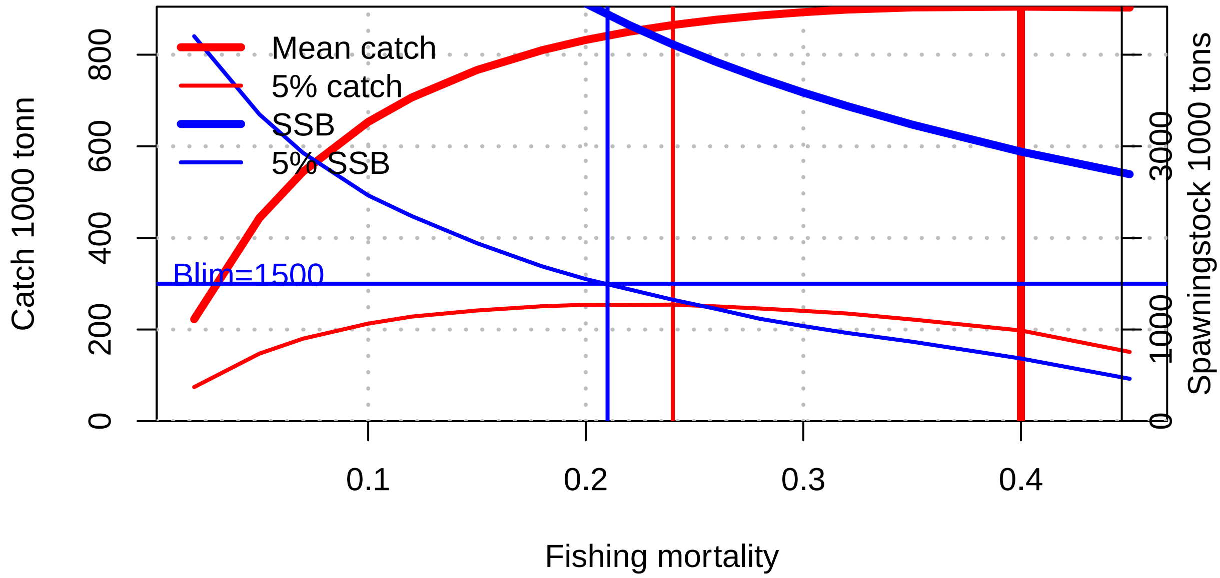


Figure 38: Average catch, fifth percentile of catch, average SSB and fifth percentile of SSB as function of fishing mortality for the base run with autocorrelation of assessment error and recruitment set to 0.6

As expected for stocks with recruitment variability as high as for blue whiting, individual runs into the future are widely different as are the confidence intervals (figure 39). Higher fishing mortality increases varibility in catches (figures 41 and 40). The effect is though most pronounced when target mortality exceeds 0.3.

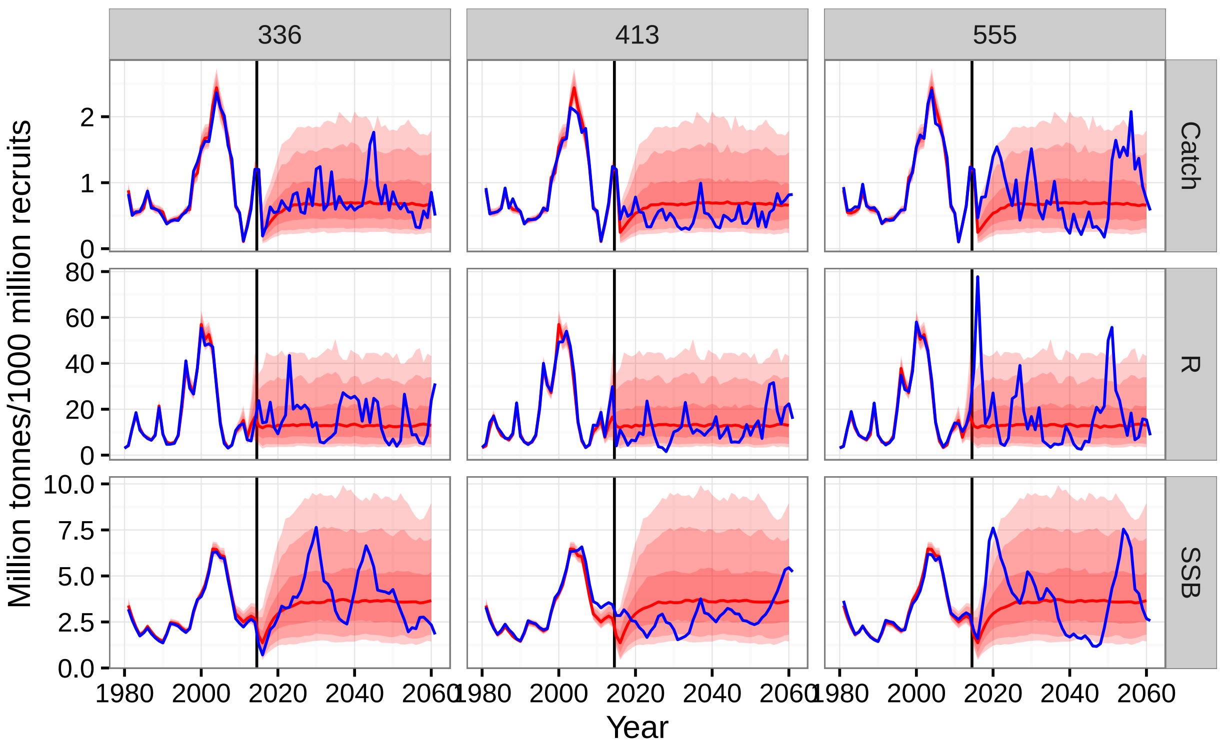


Figure 39: Summaries of spawning stock, recruitment and catch when target fishing mortality in 0.22. The shaded areas show 5, 10, 25, 75 90 and 95th percentiles and the red line the median. 3 individual runs are shown

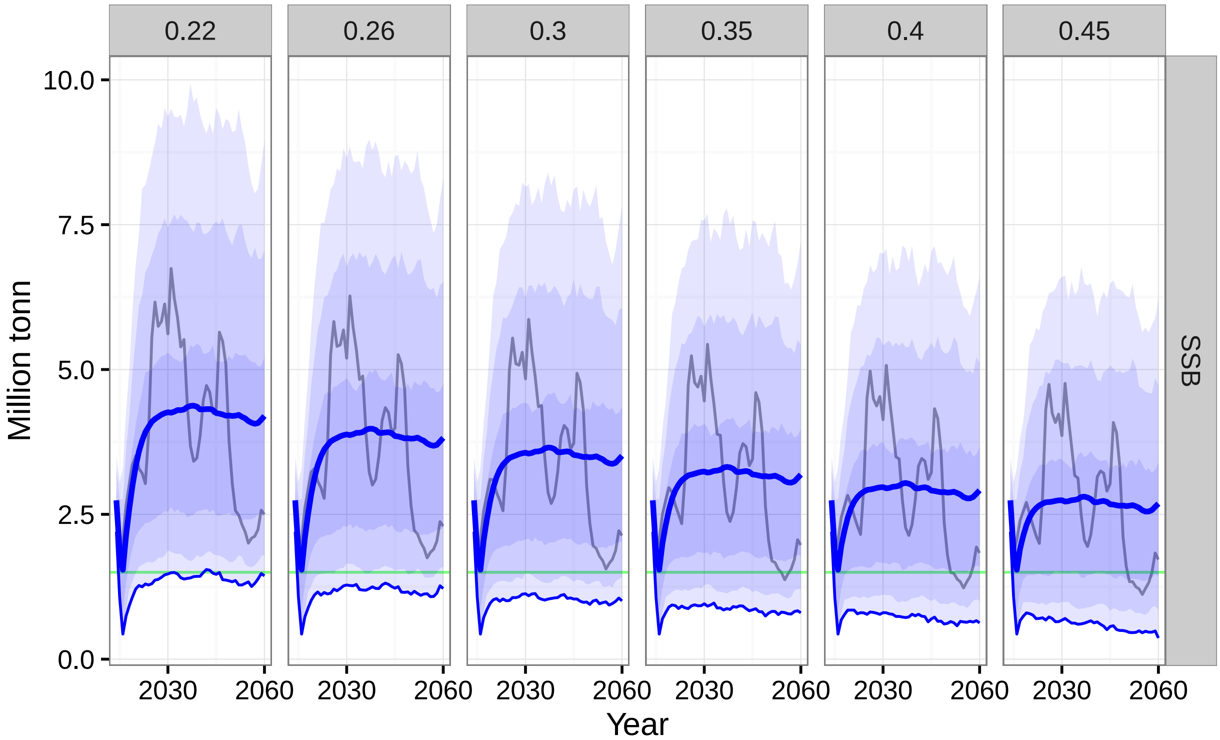


Figure 40: Development of spawning stock for 6 different target fishing mortalities. The shaded areas show 5, 10, 25, 75 90 and 95th percentiles and the blue lines the median. One individual run is shown. The horizonal lines shows B_{lim}=1.5 million tonnes.

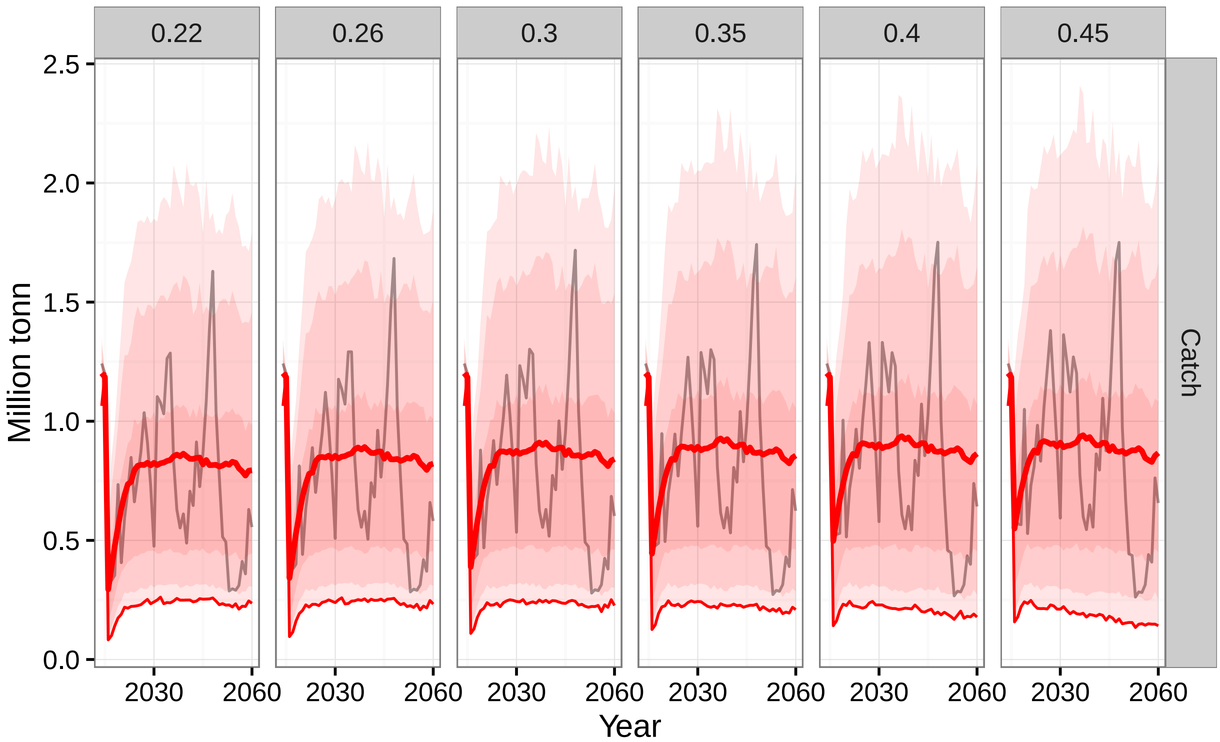


Figure 41: Development of catch for 6 different target fishing mortalities. The shaded areas show 5, 10, 25, 75 90 and 95th percentiles and the red lines the median. One individual run is shown

5 Effect of trigger

The simulations set up so far have not included any trigger point, below which fishing mortality is reduced. (figure 42). If the goal of a management plan is to have less than 5% probability of being below B_{lim} a fishing mortality of 0.22 or lower will reach the goal without any trigger. But various combinations of fishing mortality and trigger point will achieve the same goal. Trigger point and target fishing mortality are positively correlated so the same goal can be reached by higher fishing mortality and higher trigger, giving higher catch on the average but more variable.

In the simulations the trigger action was based on the "estimated" spawning stock in the beginning of the assessment year. The assessment error is based on uncertainty in F a year later that is much more uncertain so the standard deviation of the assessment error was multiplied by 0.6 when applied to SSB in the assessment year. The value of 0.6 was based on results from the assessment model and short term projection with catch constraint.

Figure 42: Fishing mortality and catch as function of spawning stock with a HCR calling for fixed F above $B_{trigger}$ and linear reduction below. The wide lines show HCR with $F_{target} = 0.22$ and $B_{trigger}$=1.5 but the thin line $F_{target} = 0.35$ and $B_{trigger}$=3.0. The labels on the y asis refers to fishing mortality.

Taking an example from figure 42, if the estimated SSB changes from 1.5 to 1.2 (red lines) or 20% the catch changes by approximately 36%. Estimated uncertainty in the spawning stock in the beginning of the assessment year is 25% (or more) potentially leading to advice changing by up to 50% if $B_{trigger}$ is high (grey line). Autocorrelation of assessment error makes those cases fewer but even worse when they occurr. Therefore management scheme for blue whiting should be based on low target fishing mortality and low $B_{trigger}$ avoiding estimating SSB below $B_{trigger}$. In figure 42 someone would claim that the rule with higher target and trigger was more precautionary as fishing mortality is lower at when the stock is small. Simulations do not support this as more fish is left from earlier years when the target is lower and variability will be a problem with the rule with higher target.

F_{target}	0.22	0.233	0.251	0.277	0.31
$B_{trigger}$	1500	2000	2500	3000	3500

Table 3: Combinations of F_{target} and SSB_{break} satisfying $P(SSB < B_{lim}) = 0.05$

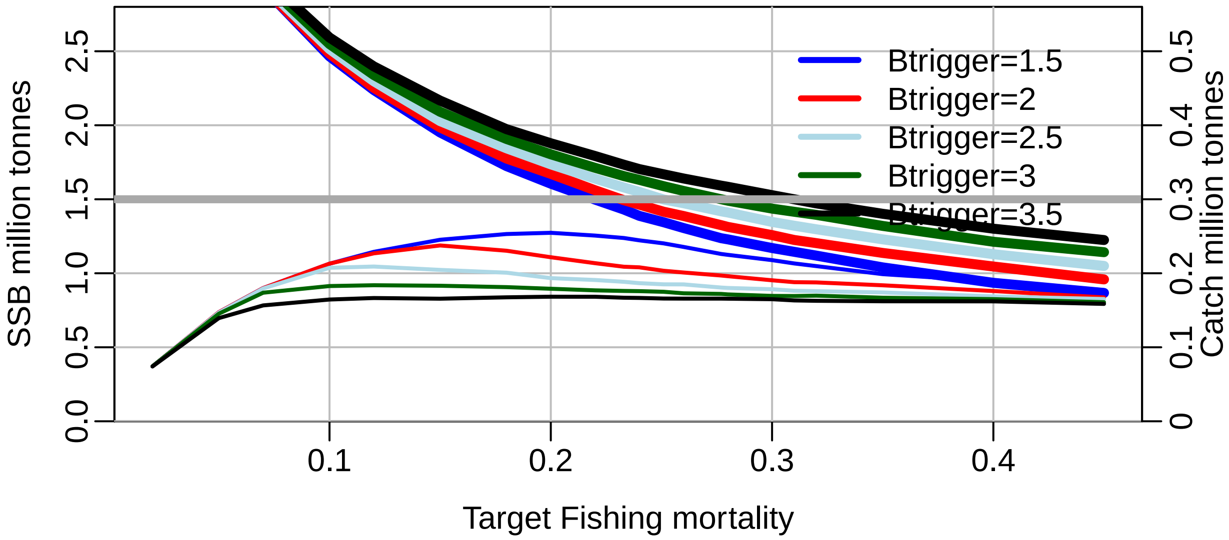


Figure 43: Fifth percentile of catch and spawning stock as function of fishing mortality and $Btrigger$. The thick lines indicate SSB and the thin lines catch

The requirement that a HCR is percautionary $p(SSB < B_{lim} < 0.05)$ leads to the following pairs of $B_{trigger}$, F_{target} (figure 43 and table 3)

With no trigger the value of F_{pa} was 0.21 but is 0.22 here with the lowest trigger. The trigger can in principle affect average fishing mortality, even though it is never reached as long as "estimated spawning stock" is lower than $B_{trigger}$ and vice versa.

Variability in catch as function of average catches, increases with higher $B_{trigger}$ (figure 44), somewhat more if positive correlation between F_{target} and $B_{trigger}$ is taken into account.

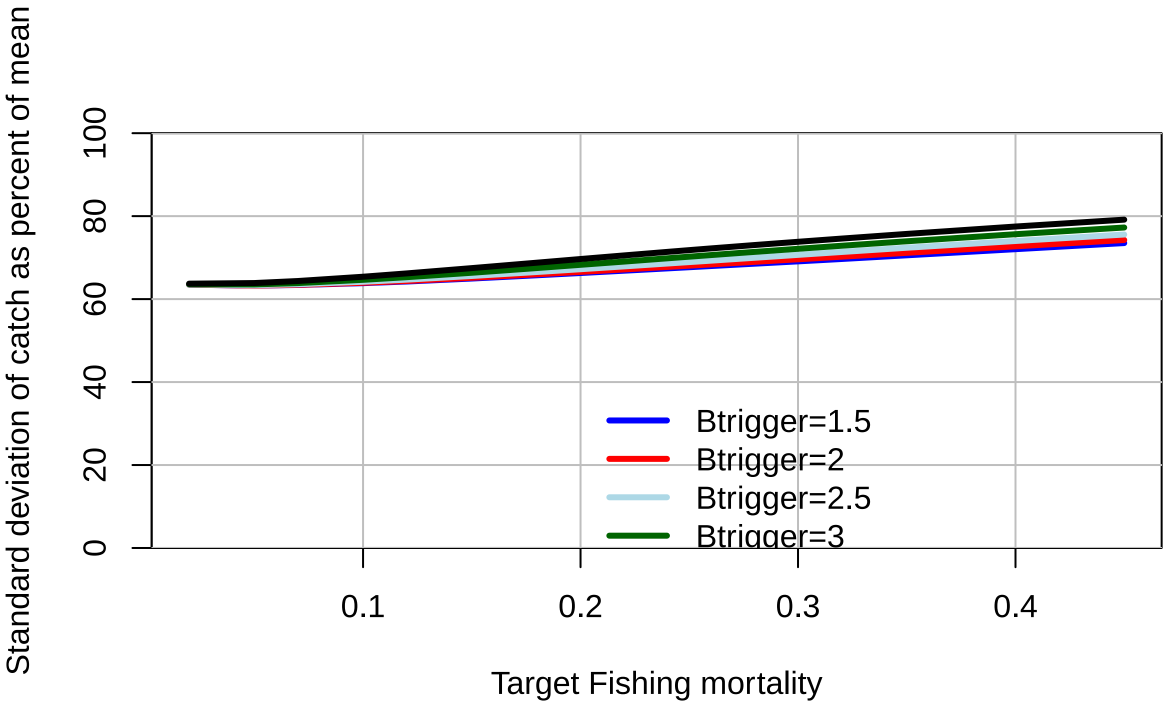


Figure 44: Standard deviation of catch divided by mean

Looking at average catch figure (45) the trigger point does not affect it much (individual lines can not be identified) so the "best policy" if the goal is to maximize average catch, is to have as high trigger high fishing mortality as possible to satisfy $P(SSB < B_{lim}) < 0.05$.

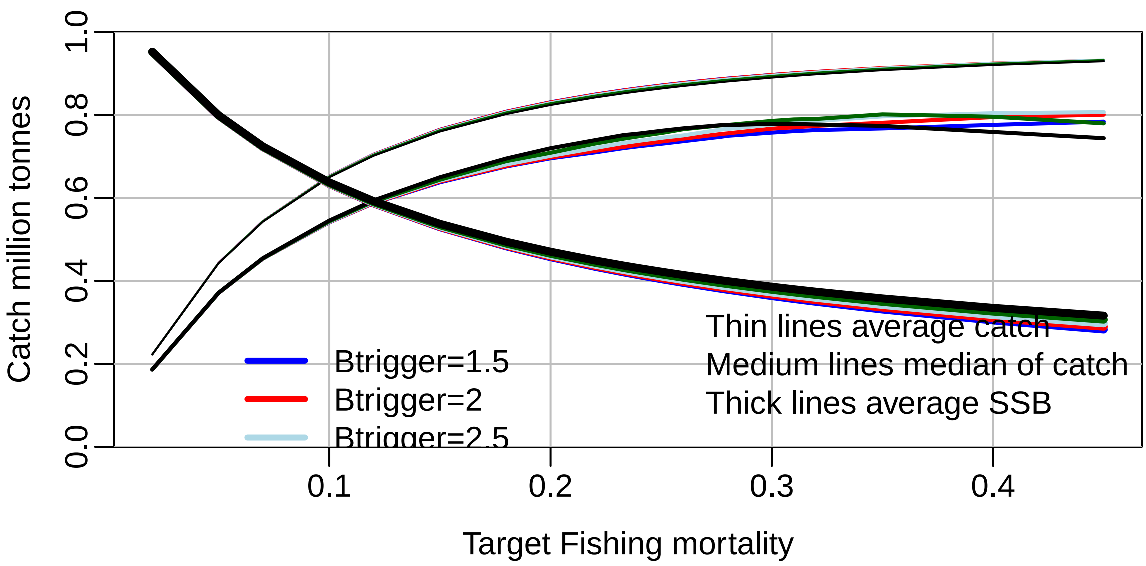


Figure 45: Average catch, median catch and average spawning stock as function of F_{target} and $B_{trigger}$

Looking at the PA combinations of F_{target} and $B_{trigger}$ (table 3) ,nothing surprising happens, variability and average value of catch increase with increase $B_{trigger}$ (table 4).

The looking at development and individual runs some diffence may be noted (figurez 46 and 47). The difference is though not very large but all of the alternatives shown could be classified as precautionary according to some critera.

	1	2	3	4	5
Btrigger	1500.00	2000.00	2500.00	3000.00	3500.00
Average catch	843.40	854.20	866.02	878.93	890.95
CV of catch	0.69	0.69	0.71	0.73	0.75
Ftarget	0.22	0.23	0.25	0.28	0.31

Table 4: Average catch and CV of catch for PA combinations of Ftarget and SSBtrigger

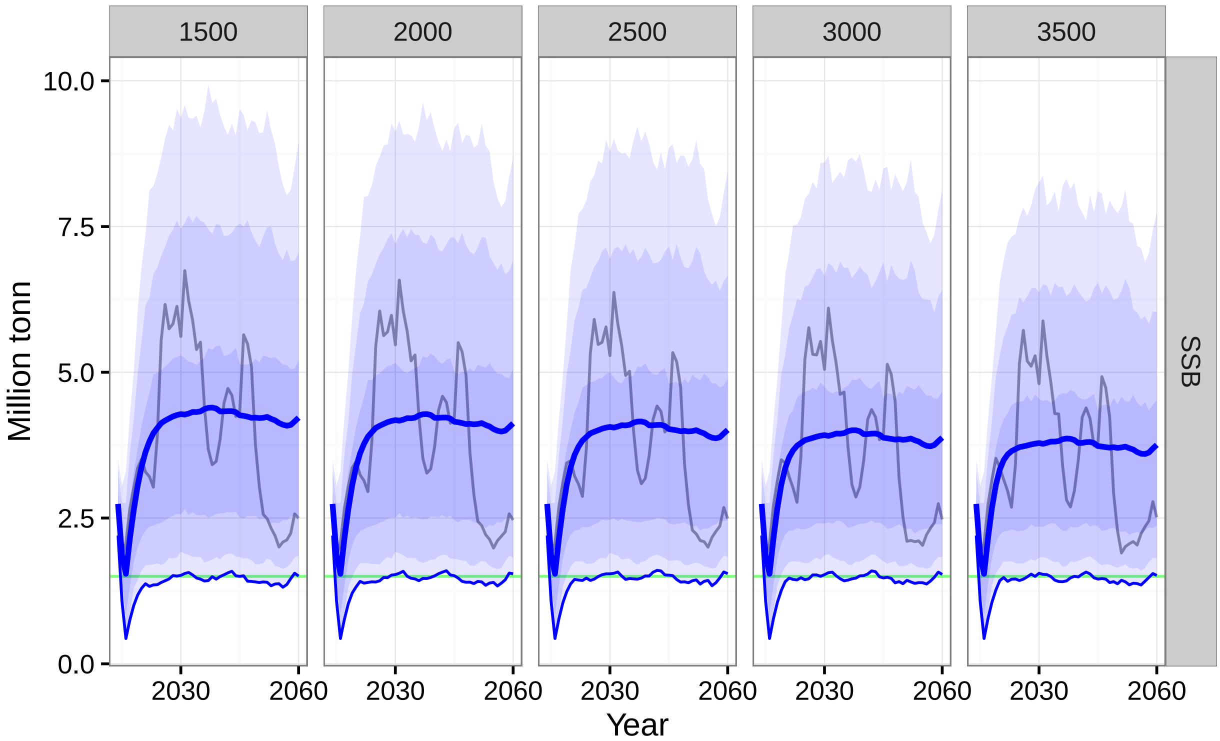


Figure 46: Development of spawning stock for 5 different combinations of F_{target} and $B_{trigger}$ satisfying $P(SSB < B_{lim} = 0.05)$. The shaded areas show 5, 10, 25, 75 90 and 95th percentiles and the red line the median. One individual run is shown. Horizontal line shows B_{lim}=1.5 million tonnes

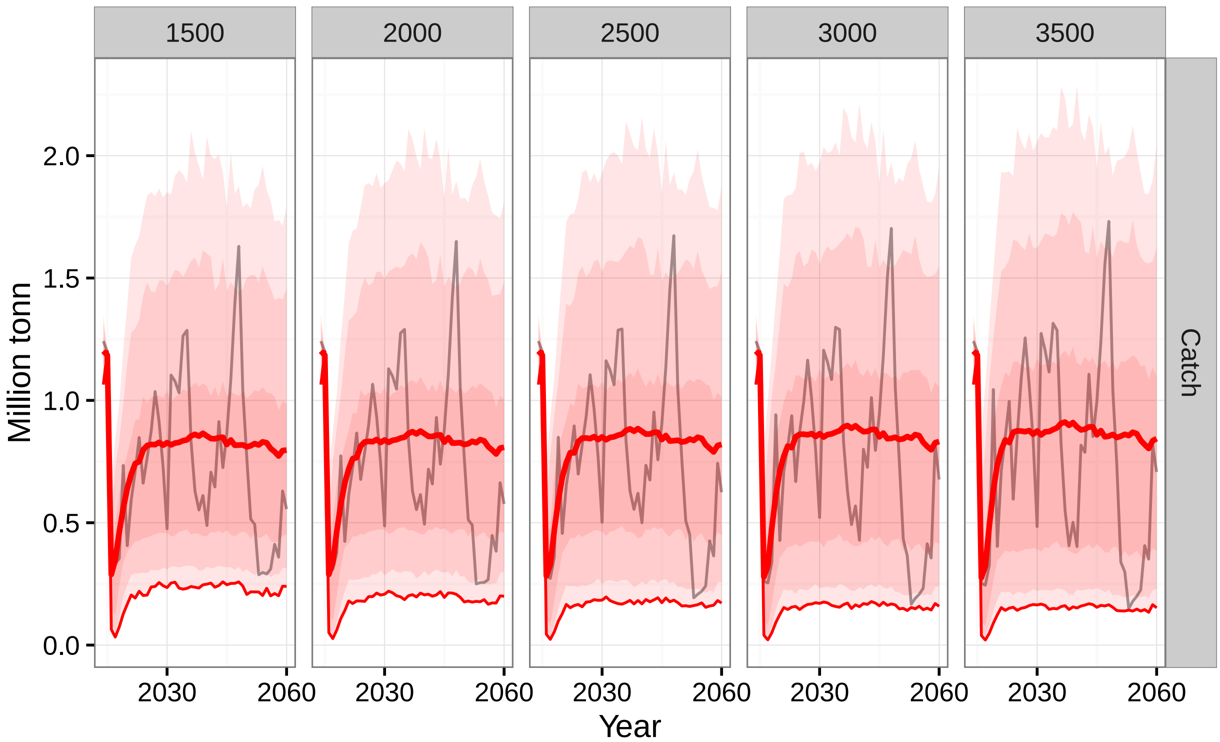


Figure 47: Development of catch and 5 different combinations of F_{target} and $B_{trigger}$ satisfying $P(SSB < B_{lim} = 0.05)$. The shaded areas show 5, 10, 25, 75 90 and 95th percentiles and the red line the median. One individual run is shown

Variability has until now been presented as standard deviation of all simulation records within certain time interval. What would be of even more interest is the interannual variability or variability over 4 years following each individual run. In "equilibrium" both those values are zero taking an average over all iterations.

The variability is here presented as the average interannual change as proportion of catch, and the proportion of cases where catch reduces by certain amount in 1 and 3 years. (figure 48). As expected lower trigger and fishing mortality lead to more stability.

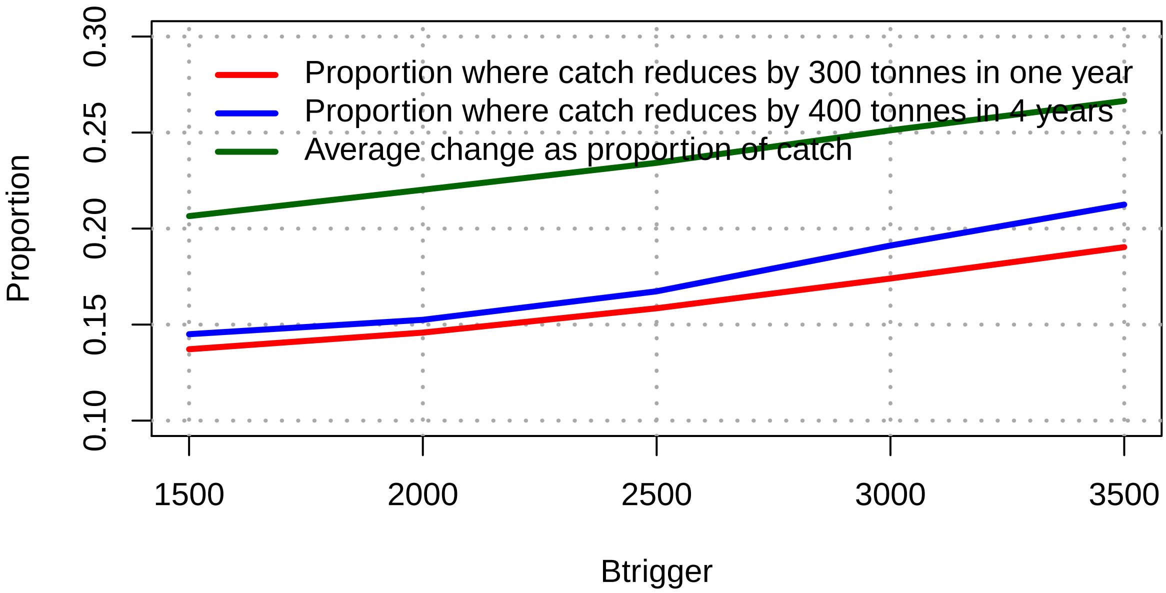


Figure 48: Different measures of variability for combinations of F_{target} and $B_{trigger}$ satisfying $P(SSB < B_{lim} = 0.05)$

6 Value and cost of fisheries

One of the things that might be considered when evaluating suitable fishing mortality for blue whiting stock is cost of the fisheries. Data from Iceland from 1999-2002 showed that use of oil was approximately 90 l/ton (EyþórBjörnsson04) and as was as proportion of value of catch up to 8 times of what is was in demeral fisheries in Iceland. Price of blue whiting is usually relatively low and the analysis here are based on 2 different prices, 0.2 EUR/kg and 0.1 EUR/kg. Usually price of larger blue whiting is higher so lower fishing mortality should lead to higher price (older and larger fish). Price of oil is assumed 0.67 EUROS/l.

CPUE is assumed to be linked to spawning stock size so the cost is calculated as $C = 90 \times 0.67 \times \frac{SSB_{ref}}{SSB}$ EUROS/ton. SSB_{ref} is the average spawning stock 1999-2002 that is 4300 tonnes.

Results from those analysis (figure 49) show that optimal harvest rate occurs at F=0.12 with the lower price (0.1 Euros/kg) and F=0.2 with the higher price (0.2 Euros/kg) (figure 49).

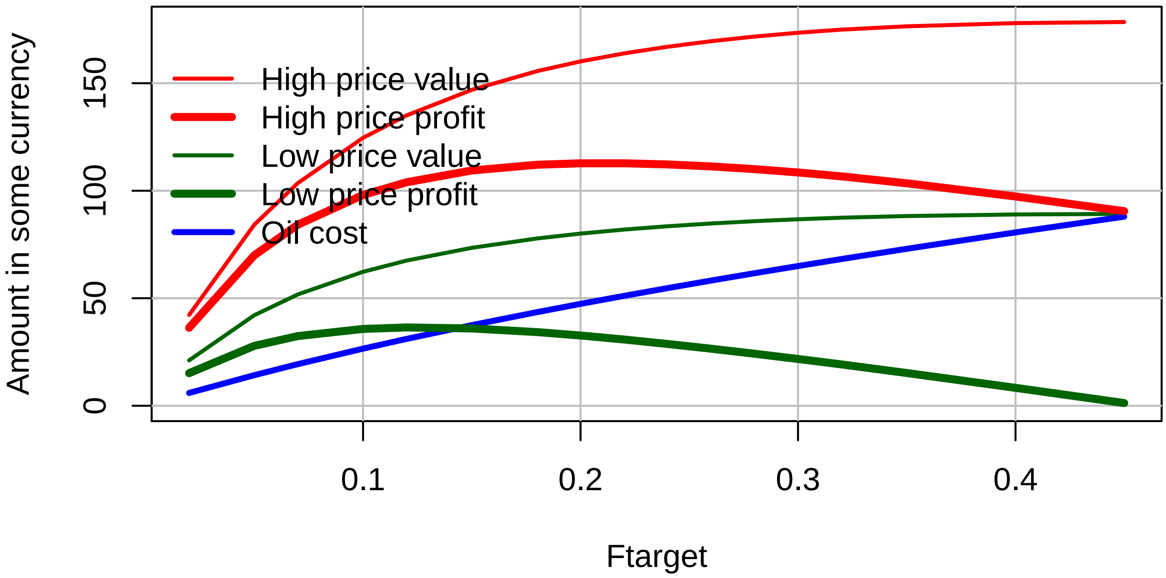


Figure 49: Income, oilcost and revenue (Income - oilcost) base on two different assumptions about the price.

The analysis done here are somewhat limited, things that could be looked at are.

1. Link between size of fish and price.
2. Estimate reduction in growth when the stock is large.
3. Only cost of oil is included here. Cost of vessels, gear , wages of crew etc are not included. When evaluating optmimal HCR for Icelandic cod in 1994 and 2003 wages of fishermen were considered a social revenue rather than cost. Taking the same approach oil cost can be considered revenue for Norway and Scotland.
4. Check if the cost and price of fish used here are realistic.
5. Penalty for oil use (CO_2) should be the norm rather than exception when evaluating HCR today. Here the penalty is probably too low.

7 Summary

This report has touched points regarding assessment, recruitment estimated and HCR for blue whiting. The main conclusions are.

- A HCR should be based on F_{3-7} in the range 0.22-0.26 and $B_{trigger}$ from 1500-2500. F and $B_{trigger}$ so the highest values of F are associated with the highest values of $B_{trigger}$and vice versa.
- Recruitment estimates are poor but a combination of the May and March survey might be feasible. Important to continue surveying over the whole are surveyed since 2003 in May.
- Assessment is very uncertain, mostly due to yearblocks in the IBWSS survey and lack of recruitment data. The 2013 and 2014 survey look somewhat like outliers.
- Recruitment is highly variable, with one long period of good recruitment (1996 - 2003). The length of the data series not sufficient to infer about the frequency of such phenomena (1 has occurred). Recruitment in the selected HCR simulations was generated by an AR model with autocorrelation lower than estimated from data (0.6 vs 0.8).
- F_{pa} i.e. fishing mortality leading to $P(SSB < B_{lim} = 0.05)$ decreases with increased autocorrelation of recruitment as expected.

- Effect of assessment is error less than of recruitment variability.
- Taking cost or environmental effects (CO_2) into account leads to considerably F_{3-7} than shown here.
- Value of fish with size and mesh penetration in pelagic fisheries are not taking into account, both factors calling for lower fishing mortality.
- No implementation error is included. Is that realistic?. Including implementation error is prectially impossible. It will just lead to F_{target} divided by the implementation error and then exceeded by implementation error squared.

In addition to most stability of catches the rule with lowest B and lowest $B_{trigger}$ i.e. $F_{target} = 0.22$ and $B_{trigger} = 1500$ is best suited to the world of annual advice (traffic lights) where a stock could potentially get a red light when the stock size is **estimated** below $B_{trigger}$, something that could happen often with this kind of stock with high assessment uncertainty and large variation and autocorrelation of recruitment.

An alternative rule instead of $F_{3-7} = 0.22$ would be 20% of the SSB in the assessment year (19.7%).

Catch stabilisers have not been considered in this report but it was emphasized that the best stabilizer is low F_{target}, low $B_{trigger}$. A PA HCR with low F and low $B_{trigger}$ could most likely be augmented by 20% stabilizer, without much increase in risk. The most difficult time is the transition from high to low recruitment regime but low F_{target} should smooth the transition, the large cohorts last longer.

What is currently happening with the assessment and advice of this stock brings out the question how the assessment and advisory process should be modelled. According to the adopted assessment SSB_{2015} is 3.3 million tonnes. The assessment model presented here estimate the spawning stock between 2 and 2.3 million tonnes. This difference is too large taking into account serious overexploitation of the stock at the moment. The value of 3.3 million can not even be justified "as the result of the SAM" model, as they are based on abuse of that model, normal settings lead to $SSB_{2015} \approx 2.5Mtons$. (See appendix A)

8 Appendix A. Problems withe the current blue whiting assessment.

This section was written earlier and presented to the working group. Therefore some repetitions from what has written before are found. There was also some rewriting of text but the earlier version was apparently poorly written and difficult to understand.

One of the parameters set in the SAM model is proportion of F before survey. It was set to 0.15 in the 2014 assessment. Analysis of the catches show that since 2004 (when the tuning series start) 45% of the catches are taken in January - March and 72% in January - April (The survey starts sometime in the beginning of April). As all the catch in January - April is mature fish but part of the catch in the latter half of the year mature fish average proportion of F before survey is higher than 0.45 (propably over 50%). Here 50% is suggested but variable proportion taking into account the proportion of catches taken each year would be more appropriate. Similar considerations need to be taken into account when compiling the spawning stock that is somewhat depleted in years of heavy fishing compared to "spawning stock compiled in the beginning of the year"

Sam run with the default settings ($pF = 0.15$) shows the spawning stock in 2015 around 2 million tonnes. Last years assessment is shown for comparison indicating SSB_{2015} between 5 and 6 million tonnes . Alternative 2015 assessment with proportion of F before survey higher ($pF \approx 0.2$??) is closer to the old result. As shown below the assessment is sensitive to this parameter (pF), with indications that the likelihood function might have multiple optima (figure 50).

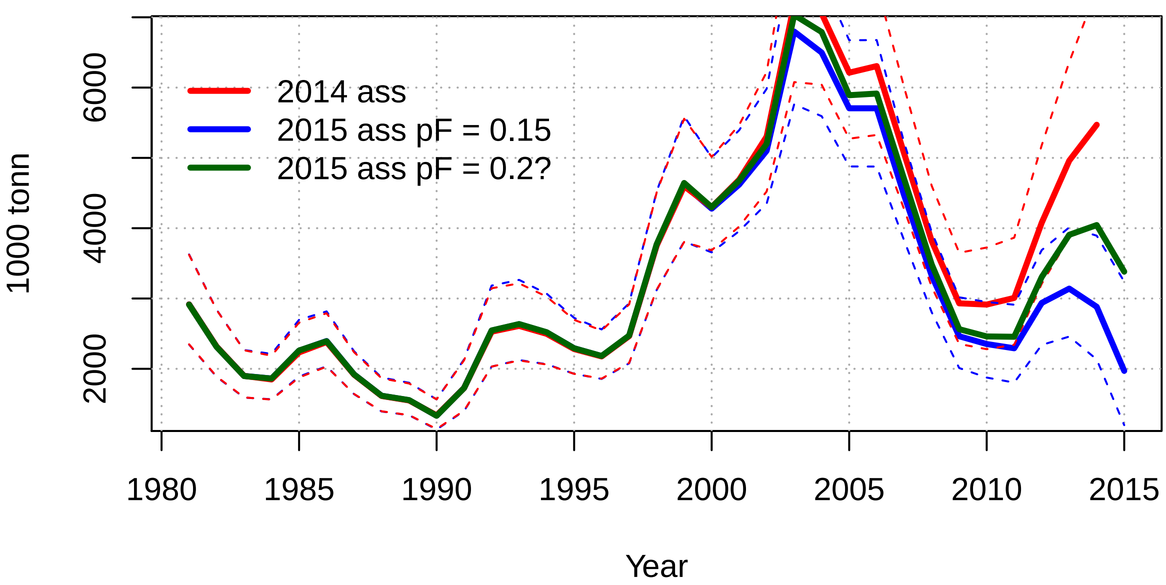
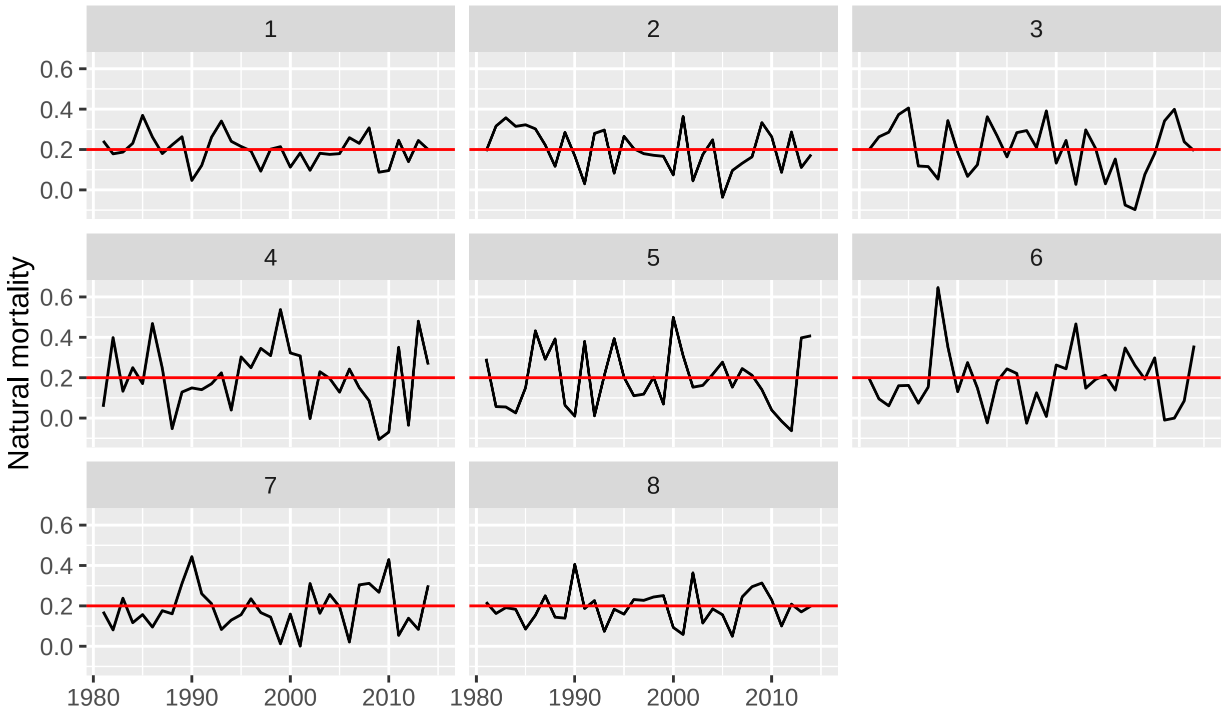


Figure 50: Estimated spawning stock from the SAM model, 2014 and 2015. Two 2015 runs are shown with relatively low and somewhat higher proportion of F before survey (pF). The lower value (0.15) was used in the 2014 assessment

SAM allows for random variability in M (process error) on top of the **0.2** used as fixed value (figure 51. The average value of this deviation is 0 but the variance is estimated, seperately for age 0 and for age 1-12. As may be seen, the process error can be quite high but it does not show as large blocks as for example seen for mackerel.



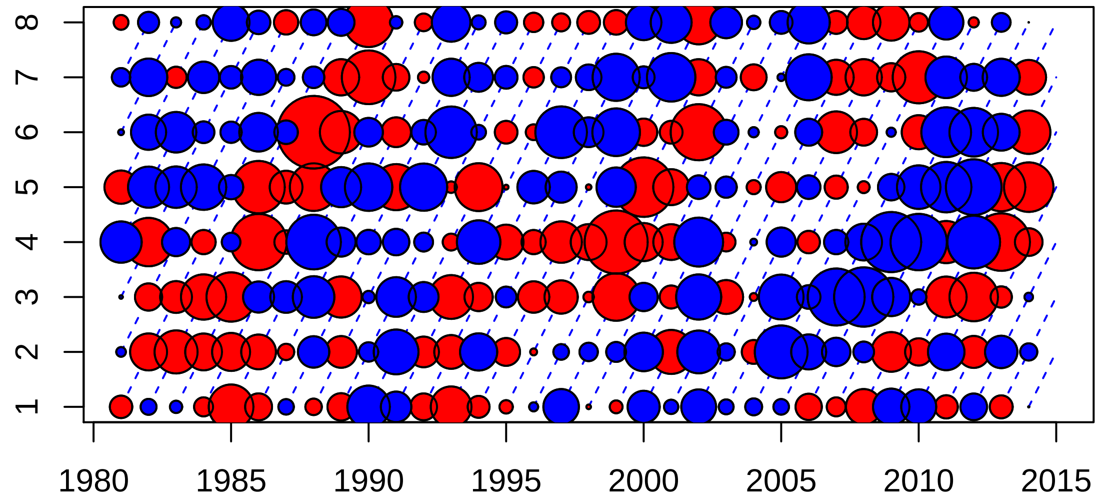


Figure 51: Deviations in natural mortality predicted by the SAM model (The model selected for prediction 2015)

Showing the process error (residuals in natural mortality) as biomass shows somewhat different pattern in recent years for the solution giving higher biomass (3.2 million tonnes in 2015) and the solution giving lower biomass (1.9 million tonnes in 2015) (figure 52). The difference is though far from enough to explain the difference in SSB_{2015} so something else must be explaining that difference.

Average value of the process error is below zero (-38 or -45 thous tonnes/year). This can happen even though the average of the process error is supposed to be zero, the reason being different biomass behind each point.

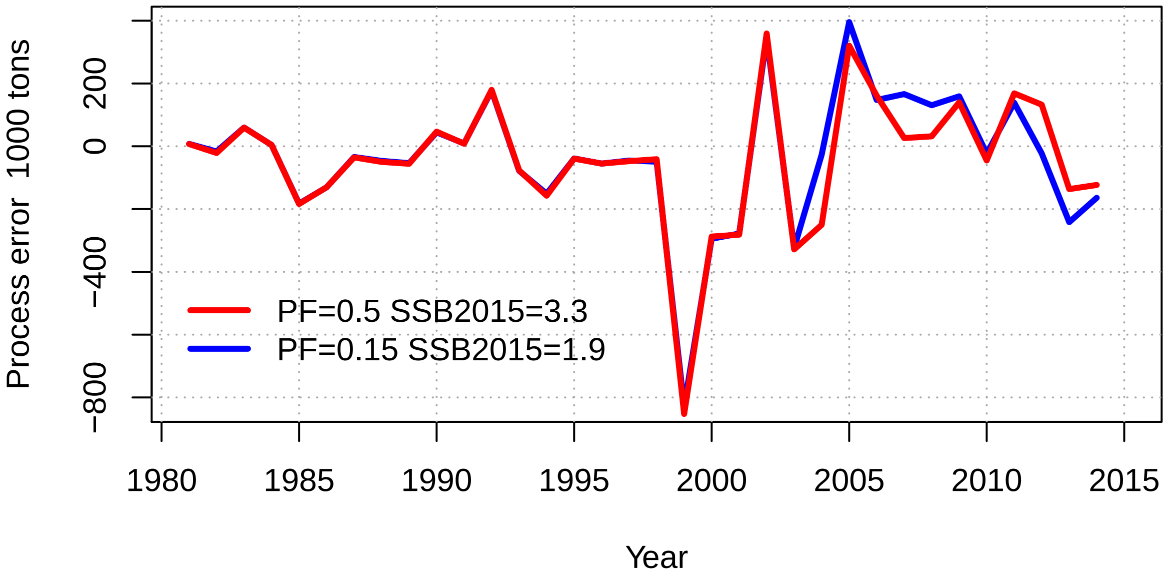


Figure 52: Process error shown as biomass from 2 Sam models giving different SSB in 2015. Difference between the model is in pF. The values shown are only for age 2 and older. Positive values in the picture mean addition to the stock.

The reason for loss of fish from the stock in 1999 is not clear. Fishing mortality was increasing rapidly after 1997. This increase was not picked up due to the random walk constraint on fishing mortality (figure 53) so the model needs to compensate for low fishing mortality by loosing fish in other ways. The separable model does not have this type of constraints . Problems with the random walk constraint might be occurring today as the interannual variability in catch has been very high and they are real. Catches decreased from 1251 to 103 thous. tonnes from 2008-2011, increasing back to 1138 thous. tonnes in 2014. Fishing mortality from the separable model varies from 0.04 to 0.3 in that period. (In TSA runs for Icelandic cod estimated parameter has been added when large change in F is known to occur).

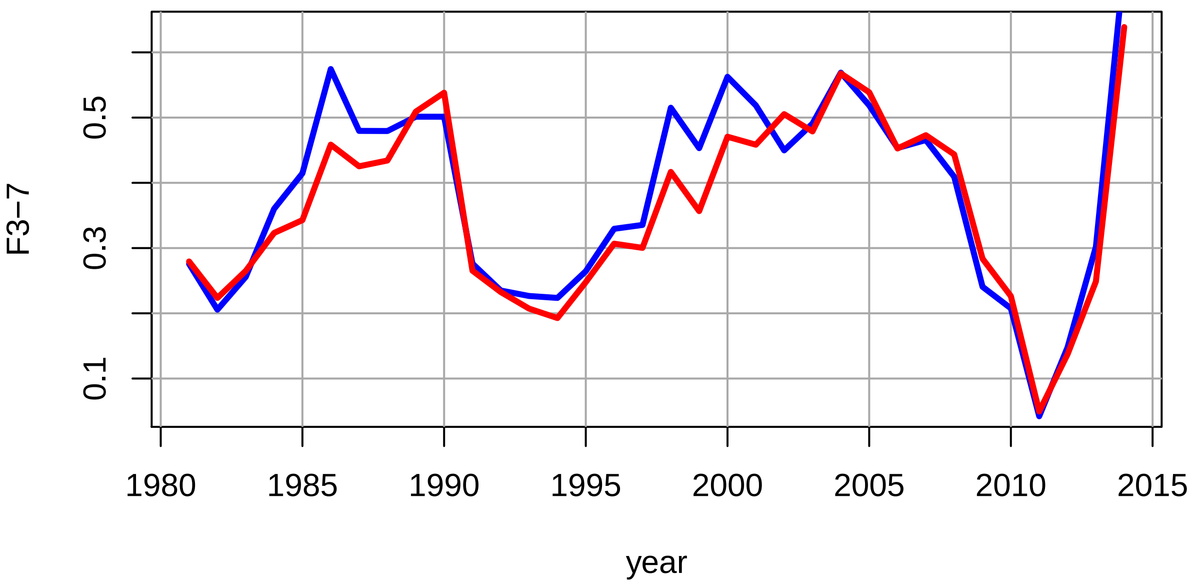


Figure 53: Comparison of fishing mortalty from SAM and the separabel model shown earlier. The separable model is based on modelling correlation of residuals in the survey and the sam model is the one with low pF (not used for advice this year)

The separable model shown in earlier sections were also tested. The model does allow the selection to change at prespecified times but analysis indicated that same selection pattern all the time was the best choice. There are some short term changes like targeting of small fish after 2000. Results of SAM model indicate the same thing, the parameter of the correlated random walk is close to **0.99**, indicating close to separable model.

In the separable model correlation between the survey residuals in same year are modelled with first order AR model ($\sigma_{i,j} = \sigma \times \rho^{|i-j|}$ where i and j denote age). The estimated correlation parameter ρ is estimated rather high or 0.78 reducing weight of the survey measurment compared six independent series, one for each age group. High value of ρ approach modelling a yearfactor in the survey. The effect of including the correlation is currently to reduce estimated stock size (figure below). Results from the separable model not including correlation of the survey residuals are similar to the "lower" SAM model results. The separable model is the same one as used earlier for retrospective runs and HCR simulations.

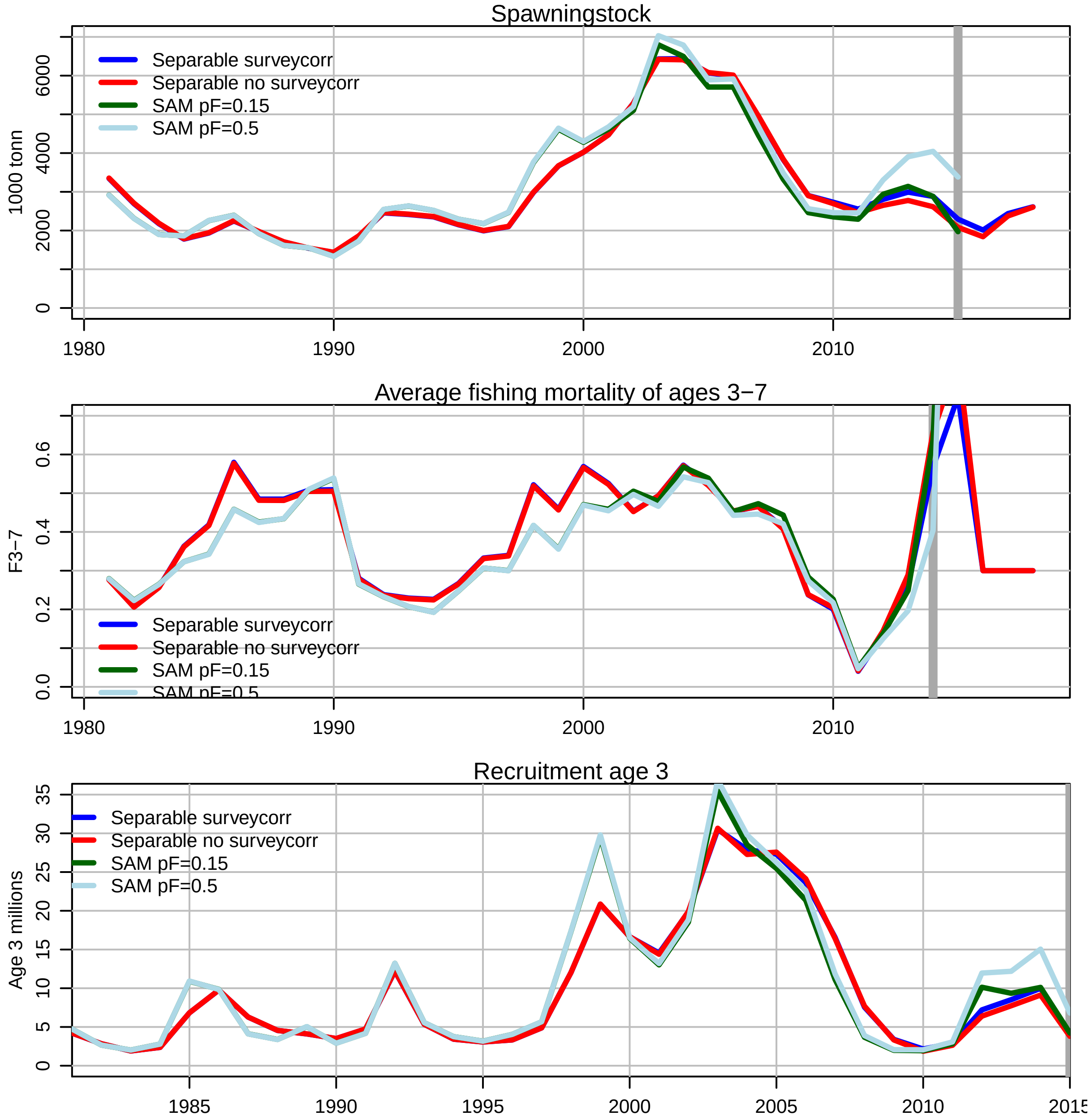


Figure 54: Comparison of results from two versions of the Sam model and two versions of the separable model

The models seem to follow the survey biomass reasonably well except in the years 2013 and 2014 (figure 55). Predicted survey biomass from the 2014 runs (green and lightblue lines) is alwo well below observed survey biomass for the years 2013 and 2014. The model taking into account correlation of survey residuals does not follow the survey biomass as well as the model where this correlation is ignored. The 2010 point shown darkgreen in figure 55 is closer to prediction than the 2013 and 2014 points.

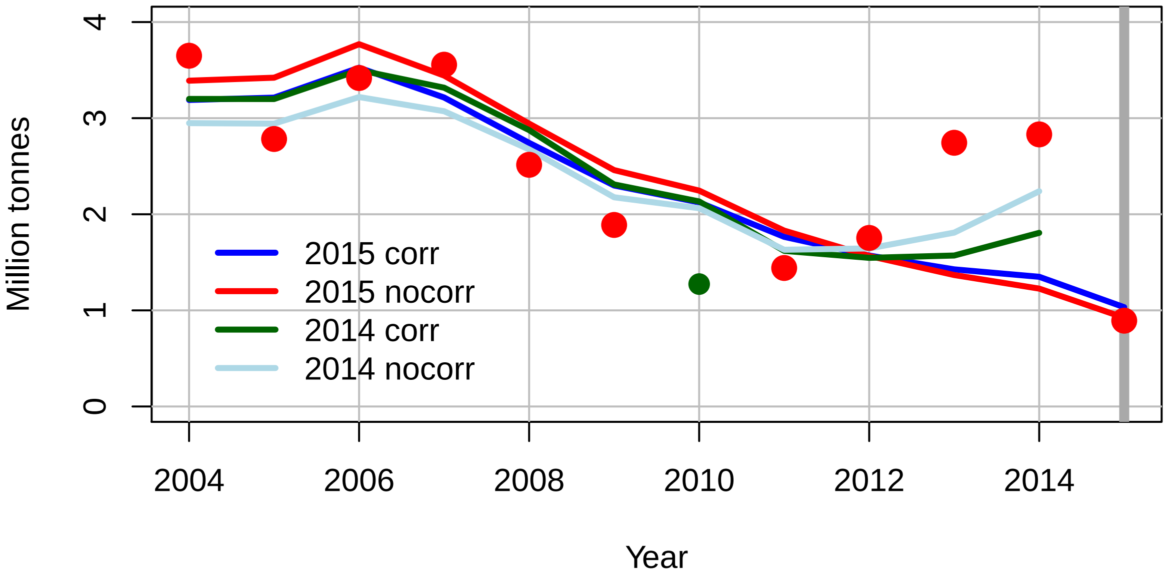


Figure 55: Observed and predicted biomass from various runs of the separable model. The 2010 survey point not used in tuning is shown as darkgreen

As described above two SAM solutions are available based on current data, which one the models finds depends on the initial conditions and proportion of F before survey. Which of the solutions the model selects depends on the parameter proportion of F before the survey (pF). If pF is larger than 0.17 the model goes for the higher solution. Then F_{2015} is estimated to be very high or in the range 3-6 and F before the survey $4 \times 0.17 = 0.68$ or higher or just enough to follow the indices in 2013 and 2014 and at the same time also the 2015 index. The spawning stock in the beginning of the year is 3.5 million tonnes but the spawning stock at survey time just 2 million tonnes. There is no catch constraint in 2015 so the only thing limiting F_{2014} is the random walk constraint that is relatively loose as the fishing mortality has been highly variable. A nice corrolary of this behaviour is that around $pF = 0.17$ the solution selected depends on the initial conditions.

Short survey series increases the problem described but Q in the survey is estimated 20% higher in the selected SAM solution compared to the one showing lower biomass.

Most model run indicate that the spawning stock in 2015 is around 2 million tonnes. The landings in 2015 are predicted to be around 1.2 million tonnes, a large part of the spawning stock. Little information is avaliable on recruitment ??, at least none that is included in the assessment.

As the assessment is presented now the stock might be seriously depleted by the end 2015. This assessment is driven by the 2015 survey that might of course be misleading. There are though indications that the 2013 and 2014 surveys might be problems in the series. This assessment is selected in period of very high exploitation rate and can lead to severe depletion of the stock.

Blue Whiting, TOR c) examine the impact of including Q1 catches in the assessment year?

WD to the Inter-benchmark for Blue whiting, spring 2016.

Thomas Brunel, IMARES

Morten Vinther, DTU Aqua

1.1 TOR

a) evaluate the robustness of the SAM model in situations with clear "year effects" in survey indices as observed in the IBWSS 2015, test appropriate model modifications and /or make criteria for posterior discarding of survey indices;
b) estimate recruitment for short-term forecast from the present available survey indices;
c) retrospectively examine the impact of including Q1 catches in the assessment year;
d) estimate Biological Reference Points.

This WD documents progress in relation to TOR C.

Most of the blue whiting catches occur early in the year. It was therefore suggested that catches during the beginning of the assessment year (either Q1 or Q1+Q2) could be used to predict the annual catches in this year. By incorporating these raised catches in the catch at age matrix used for the assessment, the model would have an additional source of information to confront with the most recent survey index, which might result in terminal year estimates being less sensitive to year effects in the survey.

In this working document, 1) the potential for raising preliminary annual catches for Q1 or Q1+Q2 catches is investigated and 2) the consequences for the assessment from using these raised catches is examined.

1.2. Estimating preliminary annual catches from Q1 or Q1+Q2 data.

Quarterly catch numbers-at-age data were available for the periods 2000 to 2006 and 2010 to 2015. The proportion of catches taken in Q1 varied in average between a very low percentage for the incoming year-class, to between 40 and 50% for ages 3 and older (figure 1). This proportion was quite variable in time with, for instance, an increase for most ages between the end of the earlier period and the start of the second period. Interannual variation can be also large, with changes of +-30% frequently observed. Ages 4 to 9 were slightly more stable in time. Due to these large variations, the average over the last three years does not appear to be good predictor of the proportion of catches in Q1 in a given year.

The proportion of catches in Q1+Q2 varied between 10% for the 0 group and 80% for age 3 to 10 (figure 2). Although the proportion in Q1+Q2 showed less year to year variations than proportion in Q1, there were still substantial changes in the longer term, namely an increase from the earlier period to the more recent period. The average over the last 3 years seem to be an acceptable predictor of the proportion caught in Q1+Q2 for most ages, except ages 1 and 2 which showed a higher temporal variability than the other age groups.

Using the average of the proportion caught during Q1 in 2012-2014 as an estimate for 2015, the catches from Q1 in 2015 were raised to the whole year, and added to the catch-at-age data used in the WGWIDE 2015 assessment (which ended in 2014). The catch-at-age matrices corresponding to the 2014 and 2013 assessment were updated by the addition of the in-year catches raised in the same way. Finally, the same catch-at-age matrix were also produced using the proportion caught in Q1+Q2 to raise the in-year catches instead of Q1 alone.

The SAM assessment model for blue whiting was then run for each of these 6 new catch matrices, in addition to the original 2013, 2014 and 2015 WGWIDE assessments.

1.3 Improving the assessment by using preliminary annual catch data

To investigate the impact of the preliminary catch data on the assessment results a series of assessment (using Model 3 configuration, see WD 1) were made. Each assessment was extended by one year using the preliminary catch at age number. Other annual data (e.g. mean weight at age or M at age) were copied from the previous year. Table 1 shows an overview of assessment runs.

Table 1. Overview of assessment runs

Assessment run	Catch and other assessment data	IBWSS data
2014_p2015Q1	1981-2014 & 2015 (Q1 preliminary)	2004-2015
2014	1981-2014	2004-2015
2013_p2014Q1	1981-2013 & 2014 (Q1 preliminary)	2004-2014
2013_p2014Q12	1981-2013 & 2014 (Q1+Q2 preliminary)	2004-2014
2013	1981-2012	2004-2014
2012_p2013Q1	1981-2012 & 2013 (Q1 preliminary)	2004-2013
2012_p2013Q12	1981-2012 & 2013 (Q1+Q2 preliminary)	2004-2013
2012	1981-2011	2004-2013

1.3.1 Results

The estimated Fbar for the various runs are shown in Figure 3. The values show a highly unstable F which makes it difficult to see improvements by using preliminary annual catches. However, if we assume that assessment results are more correctly estimated in a succeeding assessment, it is possible to evaluate the effect of using preliminary catches. Table 2 shows estimated F and such comparisons of F. If we assume that the results from the “2014” assessment are the most correct, F in 2014 (the last assessment year) was estimated to 0.606. The “2013” assessment estimated F(2014) to 0.178 which is 29% of the F estimated by the “2014” run. The run with 2013 catch data and preliminary data for 2014 based on Q1 data (run “2013_p2014Q1”) estimates F(2014) to 0.278, which is 46% of F(2014) from the “2014” assessment. This is not a precise estimate, however much better than the assessment without the preliminary catches. Using both Q1 and Q2 data for preliminary catches of 2014 gives almost the same result as just using Q1 data. The same conclusions can be made if we assume that the “2013” assessment results are the most correct. F(2013) is estimated to 0.178, which is close to (98%) of the F estimated by the 2012 assessment with preliminary catches, however far from the F(2013) estimated by the “2012” assessment.

For estimation of TAC, F in the “intermediate” year as estimated by the SAM model is really not used, as it is substituted by F calculated on the basis of expected catch in the intermediate year. With this in mind there seems to be no gains by using preliminary catches. For example, the “2014” assessment estimates F(2013) to 0.270, whereas the “2013” assessment estimates F(2013) to 0.178. In the 2013 assessment with preliminary 2014 catches F(2013) are even further away from the “best” estimate.

Doing the same analysis for SSB (Table 3) shows that the use of preliminary catch data does not improves the precision of the assessment, using the same criteria as for the analysis of F.

The parameter estimate for the various runs (Table 4) show that the longest time series of IBWSS has the highest observation variance. Observation variance of catches seems more independent of time series length. The main problem with the assessment, that an addition of a data from a new survey year creates rather different results is clearly shown by the change in catchability over time (Figure 6). After a relatively constant catchability up to the 2013 survey, catchability for the run including the 2014 survey data went up considerably and down again when the 2015 survey data were used.

1.3.2 Conclusion

The proportion of catches occurring in Q1 appears to be too variable to be used as a basis for raising the Q1 catches to the whole year. The proportion in Q1+Q2, however, is much less variable, especially for ages 3 and older, which represent the bulk of the catches.

The assessment results are highly variable from one year to the next (strong retrospective noise) such that a stable value of F or SSB is only obtained for the period 3-4 years before the last assessment year. The addition of a new survey year seems to influence the results more than addition of additional catch at age data. Preliminary catch data seem not to improve the retrospective pattern except for F in the “intermediate year”, which however is not used in the presently used procedure to calculate TAC.

On a side note, this exercise showed that the proportion of fish caught in Q1 showed some substantial changes over time. This indicates that the input data “proportion of fishing mortality occurring before spawning”, which is currently a constant value across years and across ages should be made variable and updated regularly.

Table 2. Fbar by assessment run and year (upper part) and F in assessment year relative to assessment year+1 (lower part)

Run	2009	2010	2011	2012	2013	2014	2015
2014_p2015Q1	0.291	0.206	0.06	0.13	0.23	0.463	0.675
2014	0.302	0.219	0.065	0.146	0.270	0.606	0.746
2013_p2014Q1	0.263	0.180	0.052	0.104	0.166	0.278	
2013_p2014Q12	0.264	0.180	0.052	0.104	0.166	0.274	
2013	0.266	0.185	0.055	0.112	0.178	0.178	
2012_p2013Q1	0.283	0.196	0.062	0.119	0.175		
2012_p2013Q12	0.283	0.196	0.062	0.119	0.175		
2012	0.275	0.192	0.059	0.111	0.108		
	2009	2010	2011	2012	2013	2014	
2013_p2014Q1	87%	82%	80%	71%	61%	46%	
2013_p2014Q12	87%	82%	80%	71%	61%	45%	
2013	88%	84%	85%	77%	66%	29%	
2012_p2013Q1	106%	106%	113%	106%	98%		
2012_p2013Q12	106%	106%	113%	106%	98%		
2012	103%	104%	107%	99%	61%		

Table 3. SSB (1000 tonnes) by assessment run and year (upper part) and F in assessment year relative to assessment year+1 (lower part)

Run	2009	2010	2011	2012	2013	2014	2015
2014_p2015Q1	2548	2365	2333	2999	3315	3468	3199
2014	2469	2253	2173	2741	2955	2958	2395
2013_p2014Q1	2806	2713	2863	3917	4780	5494	
2013_p2014Q12	2797	2703	2855	3908	4776	5490	
2013	2775	2662	2778	3748	4502	5007	
2012_p2013Q1	2660	2510	2566	3382	3748		
2012_p2013Q12	2660	2510	2566	3382	3748		
2012	2668	2536	2634	3586	4342		
	2009	2010	2011	2012	2013	2014	
2013_p2014Q1	114%	120%	132%	143%	162%	186%	
2013_p2014Q12	113%	120%	131%	143%	162%	186%	
2013	112%	118%	128%	137%	152%	169%	
2012_p2013Q1	96%	94%	92%	90%	83%		
2012_p2013Q12	96%	94%	92%	90%	83%		
2012	96%	95%	95%	96%	96%		

Table 4. Parameter estimates by run

	2014_ p2015Q1	2014	2013_ p2014Q1	2013_ p2014Q12	2013	2012_ p2013Q1	2012_ p2013Q12	2012
Random walk variance								
"--- F	0.396	0.403	0.381	0.379	0.366	0.352	0.352	0.351
--- log(N@age1)	0.596	0.593	0.599	0.601	0.605	0.611	0.611	0.616
Process error								
--- log(N@age2 to 10+	0.175	0.173	0.169	0.169	0.164	0.167	0.167	0.164
Observation variances								
--- Catch age 1	0.423	0.421	0.415	0.413	0.418	0.434	0.434	0.420
--- Catch age 2	0.335	0.327	0.323	0.324	0.330	0.340	0.340	0.342
--- Catch age 3-8	0.198	0.202	0.200	0.201	0.199	0.208	0.208	0.198
--- Catch age 9-10	0.401	0.409	0.410	0.410	0.423	0.431	0.431	0.432
--- IBWSS age 3	0.530	0.527	0.396	0.395	0.417	0.478	0.478	0.412
--- IBWSS age 4-6	0.409	0.414	0.251	0.251	0.254	0.263	0.263	0.275
--- IBWSS age 7-8	0.441	0.376	0.339	0.339	0.329	0.348	0.348	0.349
Survey catchability								
--- IBWSS age 3	0.405	0.439	0.351	0.350	0.368	0.383	0.383	0.362
--- IBWSS age 4	0.716	0.767	0.640	0.640	0.657	0.667	0.667	0.671
--- IBWSS age 5-8	0.943	1.026	0.913	0.915	0.936	0.967	0.967	0.981
Rho	0.925	0.931	0.919	0.919	0.917	0.914	0.914	0.916

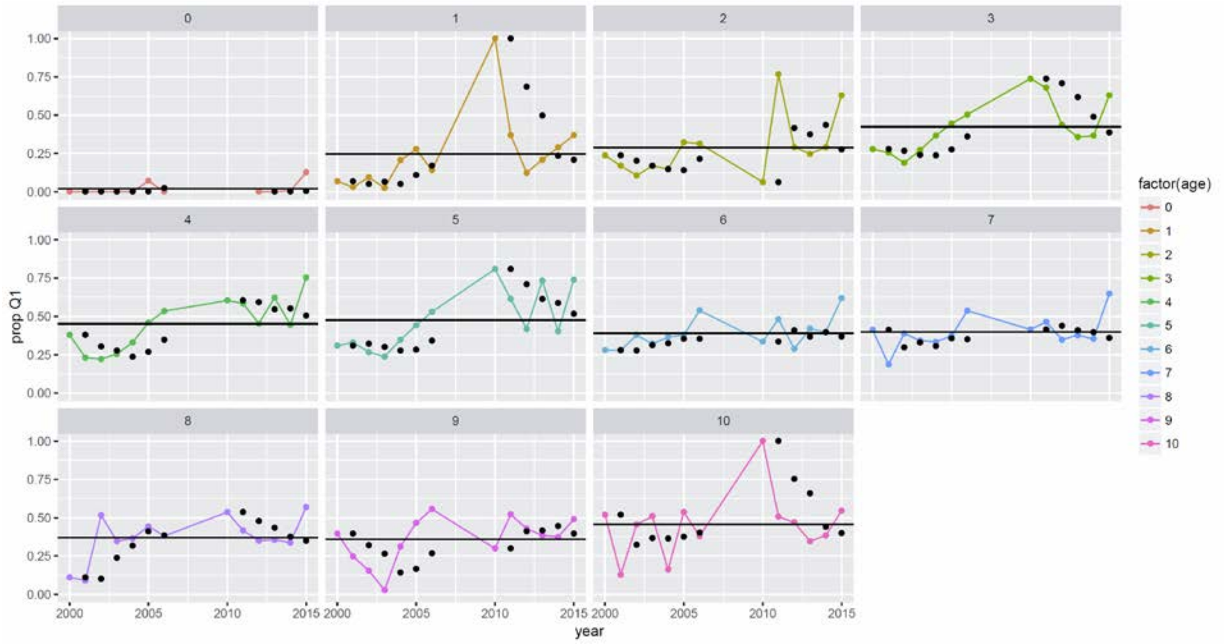


Figure 1 . Proportion of the catches-at-age occurring in Q1 (colour lines with dots) and predicted proportion for the coming year based on the mean of the last 3 years (black dots) and overall mean (horizontal line)

Figure 2 . Proportion of the catches-at-age occurring in Q1+Q2 (colour lines with dots) and predicted proportion for the coming year based on the mean of the last 3 years (black dots) and overall mean (horizontal line)

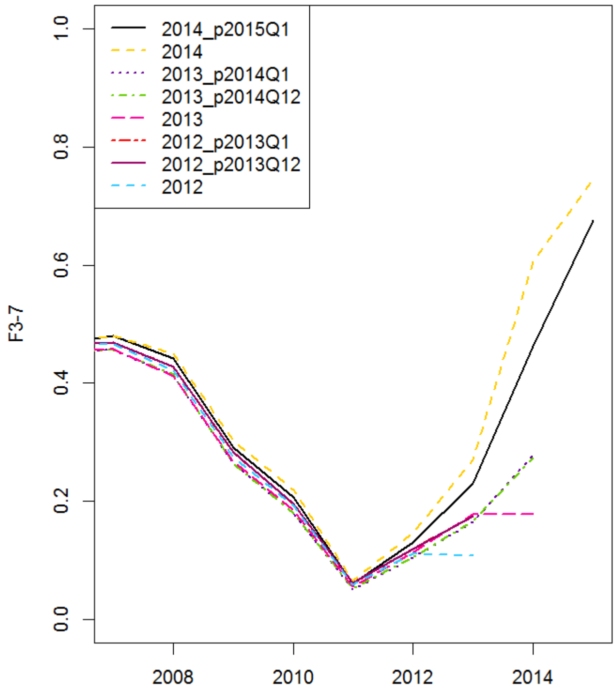


Figure 3 Fbar by assessment (see Table 1 for legends).

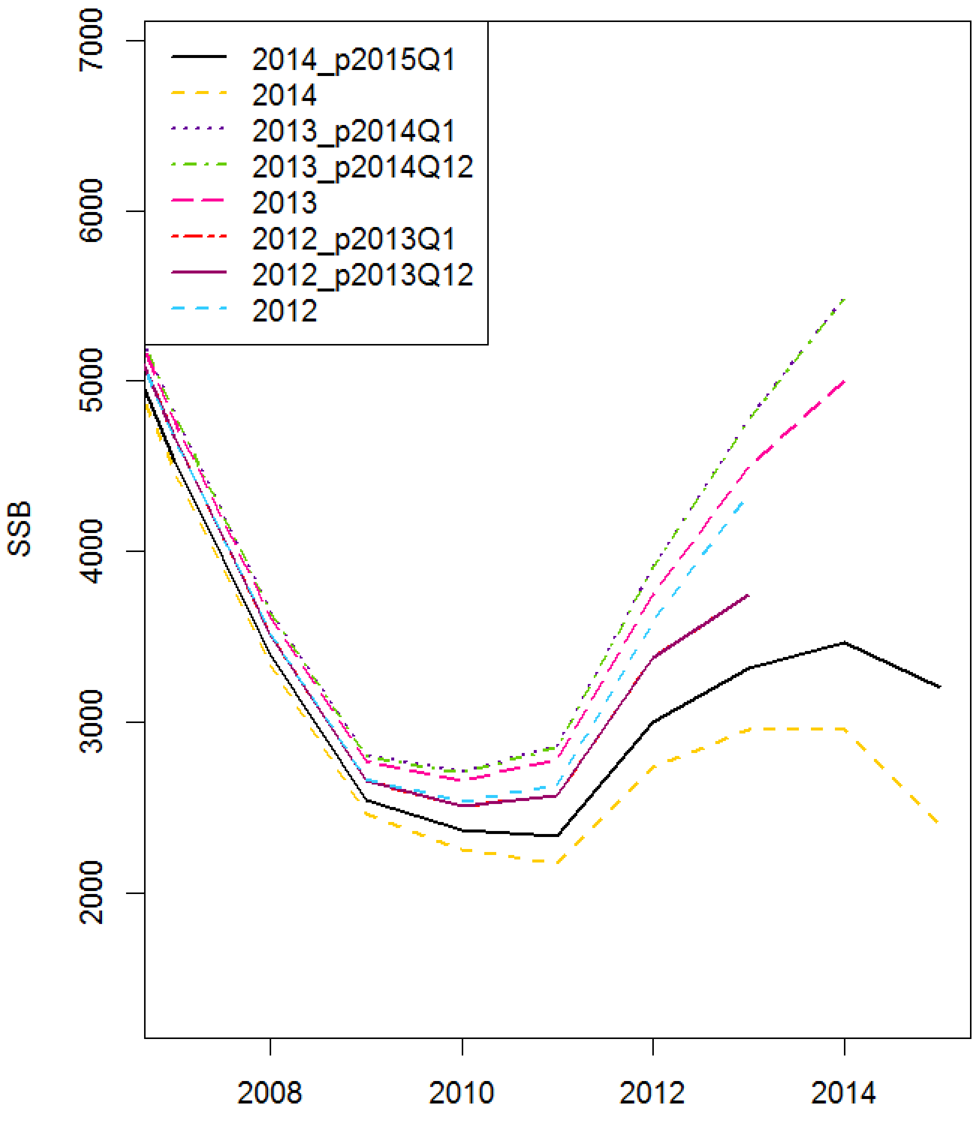


Figure 4. SSB by assessment (see Table 1 for legends).

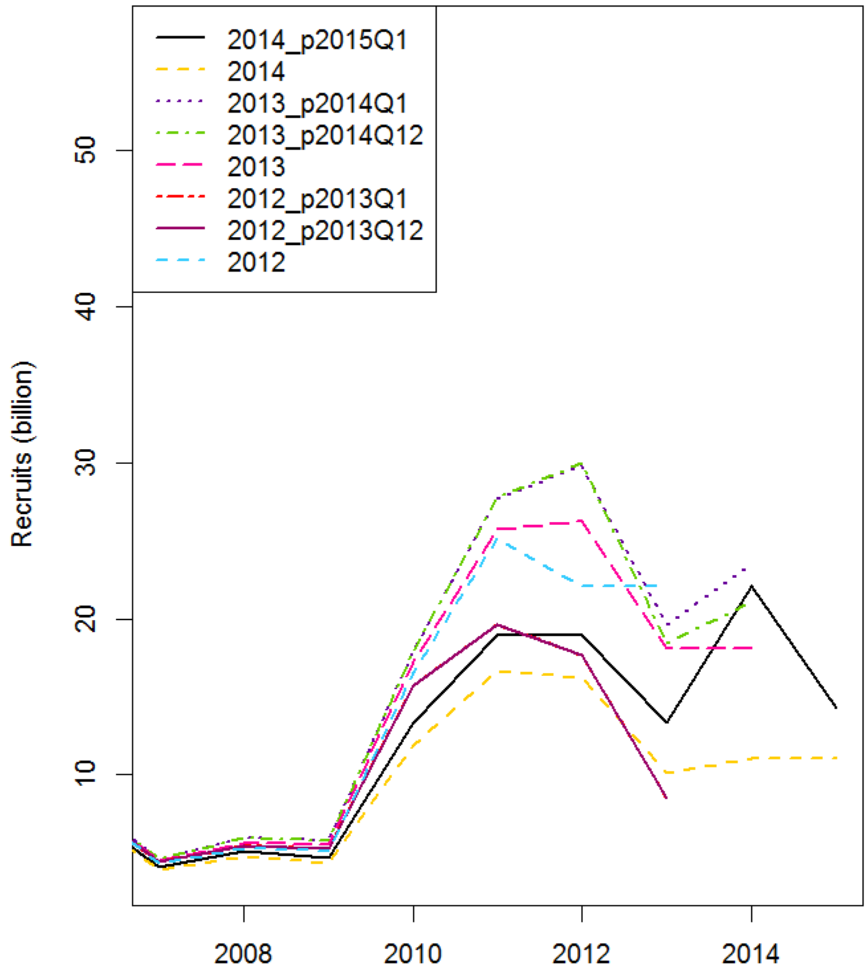


Figure 5. Recruitment by assessment (see Table 1 for legends).

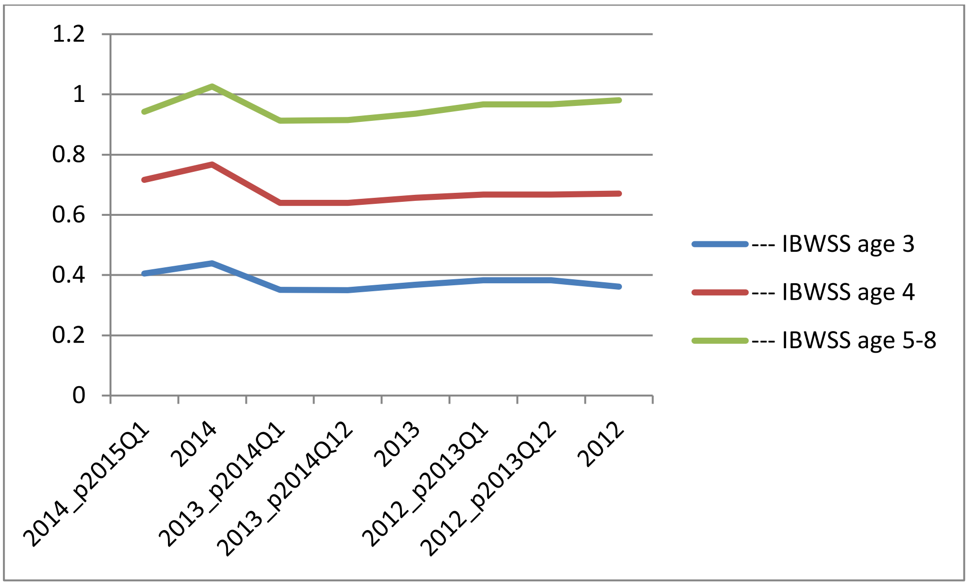
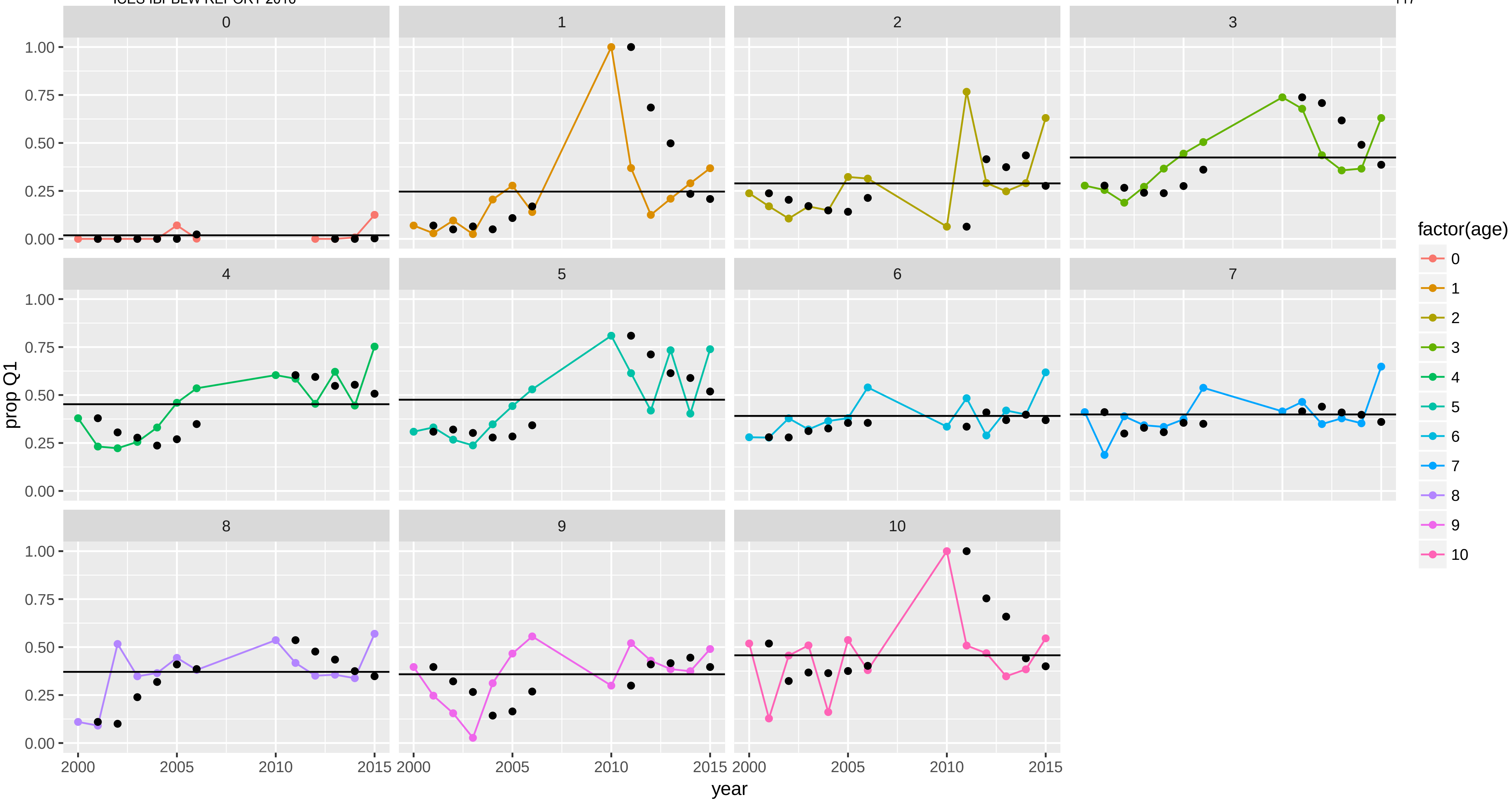


Figure 6. Catchability by age groups in the IBWSS survey estimated by run

observed (line with dots) and predicted proportions based on the mean of the last 3 year (black dots) and overall mean (horizontal line) of catches in Q1



Annex 3: Stock Annex Blue Whiting (Subareas I–IX, XII and XIV)

The table below provides an overview of the WGWIDE Stock Annexes. Stock Annexes for other stocks are available on the ICES website Library under the Publication Type "Stock Annexes". Use the search facility to find a particular Stock Annex, refining your search in the left-hand column to include the *year*, *ecoregion*, *species,* and *acronym* of the relevant ICES expert group.

STOCK ID	STOCK NAME	LAST UPDATED	LINK
whb-comb	Blue whiting (*Micromesistius poutassou*) in Subareas I–IX, XII and XIV	May 2016	Blue whiting