ICES IBPHaddock REPORT 2016

ICES Advisory Committee

ICES CM 2016/ACOM:58

Report of the Inter-benchmark on Haddock (*Melanogrammus aeglefinus*) in Subarea 4, Division 6.a and Subdivision 3.a.20 (North Sea, West of Scotland, Skagerrak) (IBPHaddock)

29 June–29 September 2016

By correspondence

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

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

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

Recommended format for purposes of citation:

ICES. 2016. Report of the Inter-benchmark on Haddock (*Melanogrammus aeglefinus*) in Subarea 4, Division 6.a and Subdivision 3.a.20 (North Sea, West of Scotland, Skagerrak) (IBPHaddock), 29 June–29 September 2016, by correspondence. ICES CM 2016/ACOM:58. 65 pp.

Contents

Executive summary

The Inter-benchmark Protocol on Haddock in Subarea 4, Division 6.a and Subdivision 3.a.20 (IBPHaddock 2016), chaired by José De Oliveira (UK) took place by correspondence during four meetings spread over several weeks (29 June–29 September 2016). There were eight participants, including two external reviewers (both from the USA) and scientists from the UK and Germany. The main focus of the IBP was to investigate the cause of the apparent failure of the TSA model, to remedy this failure, if possible, or to consider alternative models, if not, and to re-estimate reference points based on the newly selected model. The IBP identified the problem as a retrospective pattern caused by the way in which the larger post-1999 recruitment events were treated, and was able to find a TSA model configuration that remedied this problem; this was achieved by not treating any of the post-1999 year classes as "outstanding". The post-1999 period was then used as a basis for estimating reference points, apart from B_{lim} which was taken as the lowest SSB that produced an outstanding year class (1979).

1 Introduction

1.1 Background

Northern Shelf haddock underwent a benchmark in 2014 during which two haddock stocks (the West of Scotland stock and the North Sea and Skagerrak stock) were combined into one, and a TSA assessment was developed for the newly defined stock (ICES, WKHAD 2014). This model was applied in 2014 and 2015 with no apparent issues, but during 2016 a retrospective-pattern problem arose that was unexpected and had not been seen before, which led to inconsistent management advice and rejection by the Working Group, which met in April/May 2016.

In order to support mixed fishery advice soon after the Working Group, which required haddock input, the Working Group put forward an alternative model, XSA, which was previously used for the North Sea and Skagerrak stock of haddock; it was felt that, because investigations into the retrospective problem had not been able to resolve the issue in the short time available, and because the North Sea and Skagerrak stock component forms the bulk of the Northern Shelf stock, that the previous XSA model applied to the new stock could provide a stop-gap solution until further investigations could be carried out. The work done as part of this process is reported in Annex 2.

There was an ACOM review group that reviewed the document in Annex 2 above, rejected the XSA assessment (pointing to the lack of tuning indices for ages 6 upwards [but all assessment models will suffer from this], and the hanging plus group issue that is a feature of XSA [but the plusgroup accounts for mostly less than 1% but never more than 5% of all fish 1 year and older for the entire time-series, apart from 2013 when it was 11%]), and concluded that either the TSA could be used (along with formerly derived reference points based on this model), or the advice postponed. This review document is given in Annex 3.

The Working Group continued to reject the TSA with its retrospective problem, and in particular the use of the formerly derived reference points based on this model (which, in any case, no longer complied with ICES guidelines for reference points). ACOM agreed to postponement of advice, and this Inter-benchmark process was initiated as a result.

A further issue that was realised during this process leading up to the Inter-benchmark was that the retrospective pattern of 2016 is substantially different from that of 2015, which does not make sense back in time because all that was done in 2016 was to add one more year of data. This issue formed part of the investigations in the Inter-benchmark process.

The Inter-benchmark took place by correspondence during four meetings spread over several weeks (29 June–29 September 2016) with reviewers participating and providing feedback in all of the meetings apart from the final one, which dealt with reference points.

1.2 Terms of Reference

2016/2/ACOM33 An Inter-benchmark Workshop on Haddock (*Melanogrammus aeglefinus*) in Subarea 4, Division 6.a and Subdivision 3.a.20 (North Sea, West of Scotland, Skagerrak) (IBPHaddock), chaired by José De Oliveira (UK) and reviewed by

Gavin Fay (USA) and Jim Ianelli (USA), will be established and meet by correspondence until October 14, 2016 to improve the assessment model in order to provide advice according to ICES standards. The main activities to be undertaken are:

a) Determine whether there are any model settings for TSA that adequately deal with the retrospective problems that led to its rejection during WGNSSK 2016; use available model diagnostics and retrospective performance to determine whether the revised model settings are acceptable.

b) If there are no model settings for TSA that meet the criteria for an acceptable model fit, evaluate model settings for SAM and XSA to investigate whether these produce acceptable model fits, and select from among these.

c) Evaluate the appropriateness of existing biological reference points and MSY ranges. If necessary, estimate new biological reference points and MSY ranges using the selected model, according to ICES Technical Guidelines.

d) Update the stock annex as appropriate.

The work will be conducted by correspondence. Working documents should be provided to the reviewers by 30 September 2016. The Inter-benchmark Workshop will report by 14 October 2016 for the attention of ACOM.

Stock	Name, Institute	Role
Had-3a46	Coby Needle and Harriet Cole (Marine Scotland)	Stock coordinators and stock assessors

2 ToR a: Model settings for TSA

2.1 What was the problem with TSA?

There were discrepancies between the 2015 TSA assessment and the 2016 roll-over TSA assessment presented to the Working Group in May. These occurred because there was a bug in the code used in the 2015 assessment, which was corrected for the 2016 assessment. Essentially, in the 2015 code, the counter that picks out the variances on the observations of landings-at-age, discards-at-age and survey indices-at-age was misaligned by 1. The implication of this is that, among other things, age 1 landings data will have been given too little weight, age 0 discards data too much weight, and so on. However, predicting the overall effect of this on the final assessment would be very difficult. (Note that this bug did not affect the assessments of 6.a cod and whiting, which are also assessed using TSA, as these do not include age-0 data.)

We now focus attention on the retrospective pattern in the 2016 roll-over assessment. The stock summary from the roll-over assessment is shown in Figure 2.1.1 and the retrospective summary in Figure 2.1.2.

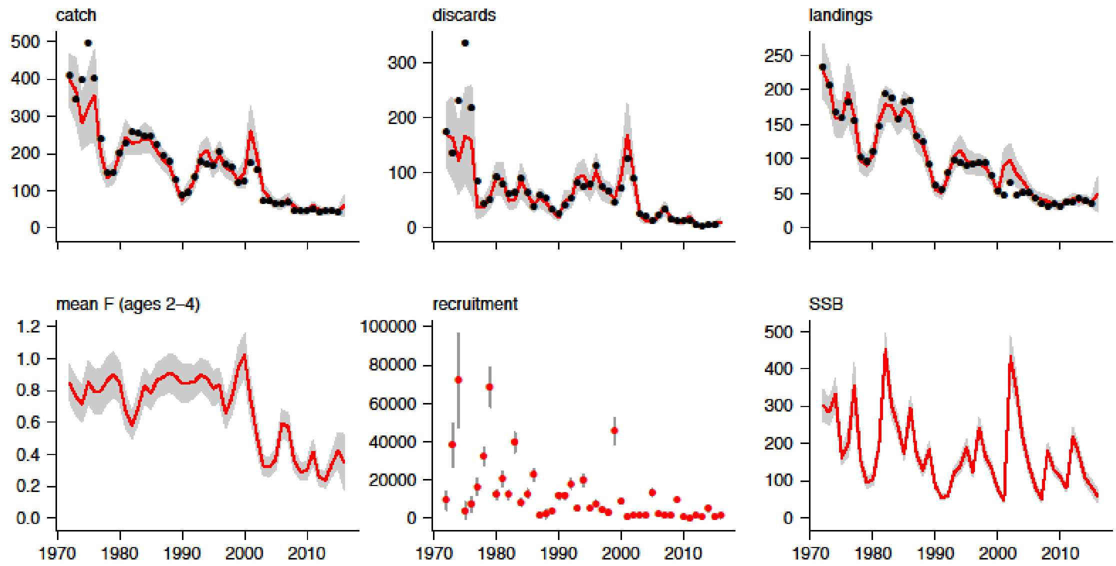

Figure 2.1.1. Roll-over assessment stock summary. Catch, discards, landings and SSB units are thousand tonnes; recruitment units are millions.

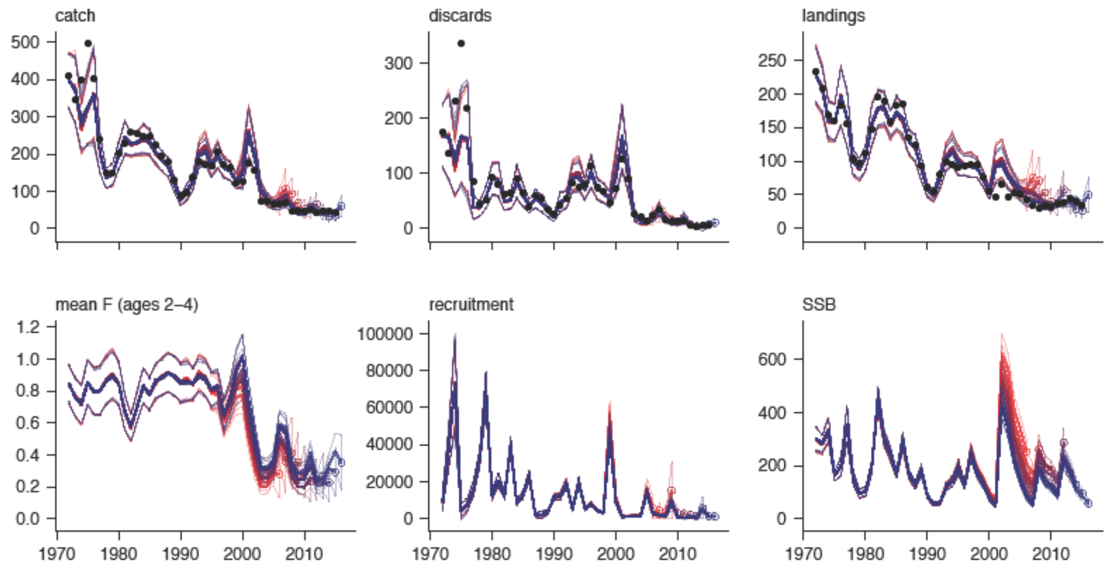

Figure 2.1.2. Roll-over retrospective showing how the estimates of SSB have reduced over time with successive assessments.

Concerns were raised about retrospective patterns, particularly in the way that estimates of SSB have reduced over time with successive assessments. However, further investigation shows that the retrospective pattern is due to changes in the estimates of recruitment (which, due to the scale used, are not evident in Figure 2.1.2). Figure 2.1.3 show the estimates of recruitment from the retrospective fits, but on the logarithmic scale. The reduction in recruitments estimates is evident. And as recruitment falls, so does SSB.

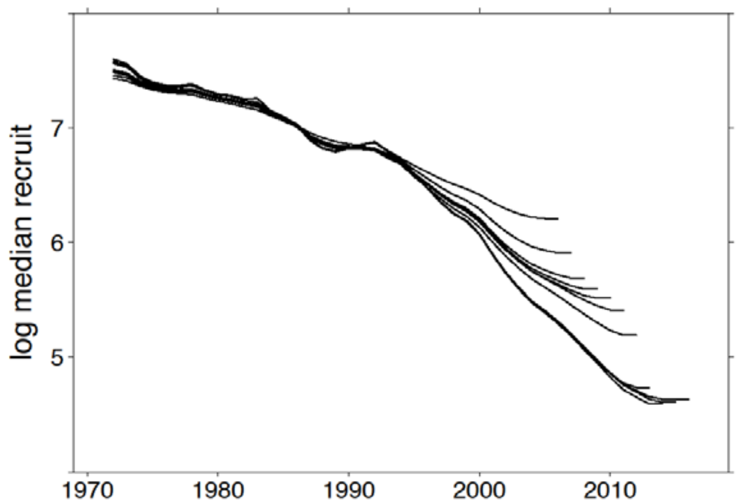

Figure 2.1.3. Retrospective pattern in recruitment from the roll-over TSA assessment. Recruitment estimates are plotted on the logarithmic scale, the scale on which recruitment is modelled.

In TSA, haddock recruitment is modelled as a random walk:

log recruit (y) = f(y) + NID(0, σ2)
f(y) = f(y-1) + NID(0, ω2)

(Note that recruitment has clearly dropped over the time-series, and attempts to model recruitment by a standard stock–recruit curve lead to markedly poorer fits and big overestimates of recruitment in recent years). However, haddock has occasional very large year classes (the 1999 year class being an obvious example), and the model above is adjusted to allow for this by writing:

log recruit (y) = log 5 + f(y) + NID(0, σ^2)

Scaling by 5 is a subjective choice, but one that usually works well in practice. Because recruitment is typically very variable, the estimates of σ are usually large, so the process can be thought of as putting a vaguely informative prior on large year-class recruitment. Clearly, using a heavier tailed distribution, or a mixture distribution would be more elegant solutions, but one that would require considerable development.

In the roll-over assessment, 1974, 1979, 1999, 2005, 2009 and 2014 were regarded as large year classes. The estimated recruitment time-series is shown in Figure 2.1.4. The reduction in median recruitment over time is clear. What is also clear is that 2005, 2009 and 2014 are large year classes relative to the years around them, but not large in the context of the whole time-series. The estimates of median recruitment in 2006 and 2016 are 119 and 103 million respectively. By contrast, Figure 2.1.5 shows the recruitment time-series from the retrospective fit with 2006 as the terminal year (landings, discards and Q3 survey data included up to 2005, Q1 survey data used up to 2006). There is a shallower decline in recruitment and the estimate of median recruitment in 2006 is 497, four times higher than in the roll-over fit. As each new year of poor recruitment is added to the time-series, the estimates of median recruitment are pulled further down, and with it goes SSB.

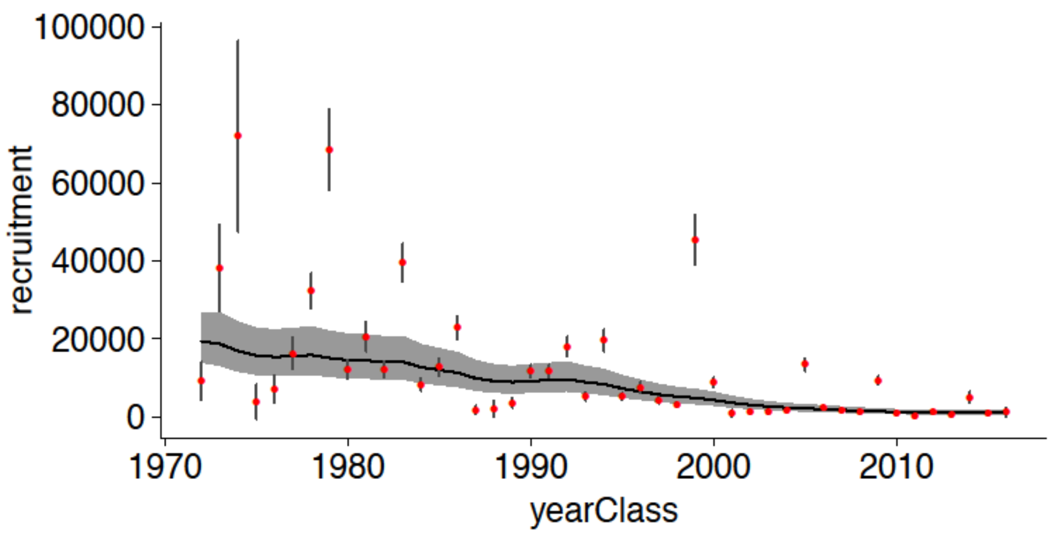

Figure 2.1.4. Recruitment time-series from the roll-over assessment.

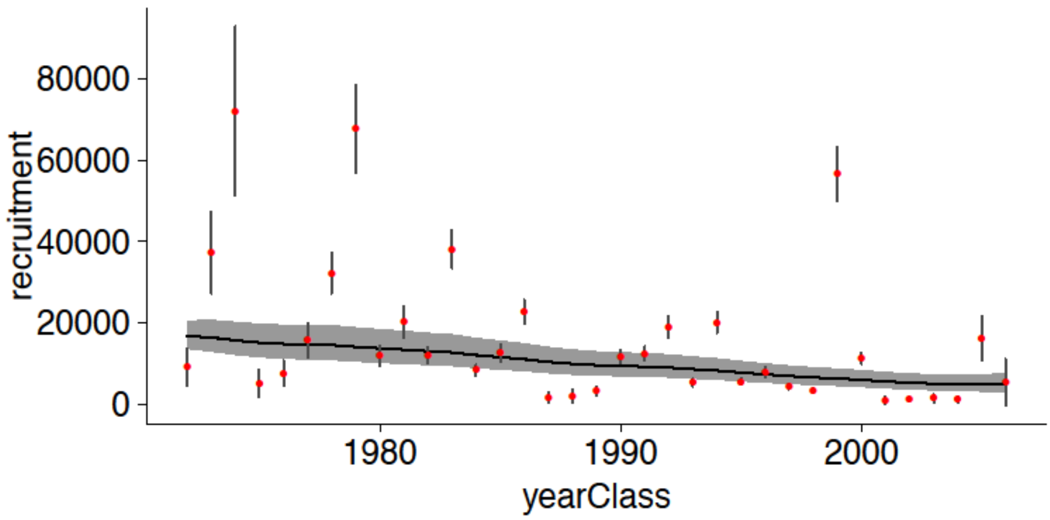

Figure 2.1.5. Recruitment time-series from the roll-over assessment based on the retrospective with 2006 as the terminal year.

Diagnostic plots are available for two types of 'errors'. Prediction errors are the scaled differences between the observations (landings, discards and survey indices) and the step-one-ahead predictions (from the forward pass of the Kalman Filter). Residuals are the more typical scaled differences between the observations and the fitted values (following both the forward and the backwards pass). Both can be useful. The prediction errors typically pick up large departures from the *status quo* (e.g. a large year class, or a sudden change in fishing effort, or just a strange observation). The residual errors are better at picking up observations that don't tie in with data later in the time-series; for example, a moderate age 0 survey index that later turns out to have been part of a large year class). Note that trends in prediction errors are expected if there are trends in the data. There shouldn't be trends in the residuals, but sometimes there are!

Figure 2.1.6 shows the prediction errors by year and age for landings, discards, the Q1 survey and the Q3 survey. Since we will be focusing on 'large' year classes, the observations associated with the 1974, 1979, 1999, 2005, 2009 and 2014 year classes (which were treated as large in the roll-over assessment) are shown in red. Figure 2.1.7 shows the corresponding residuals. Finally, Figures 2.1.8 and 2.1.9 show the prediction errors and residuals plotted by cohort, again to get more insight into the model behaviour associated with the large year classes.

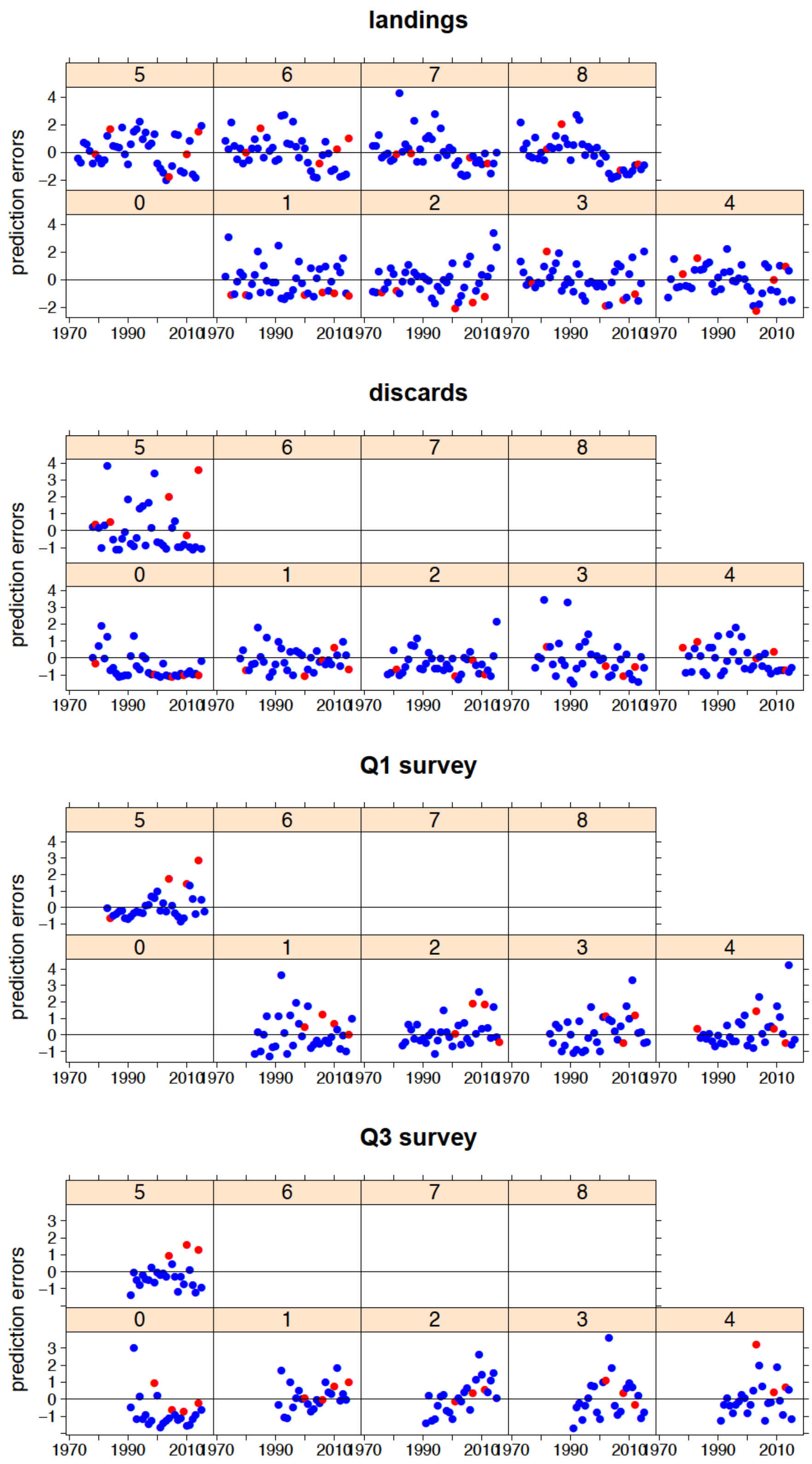

Figure 2.1.6. Prediction errors from the roll-over assessment.

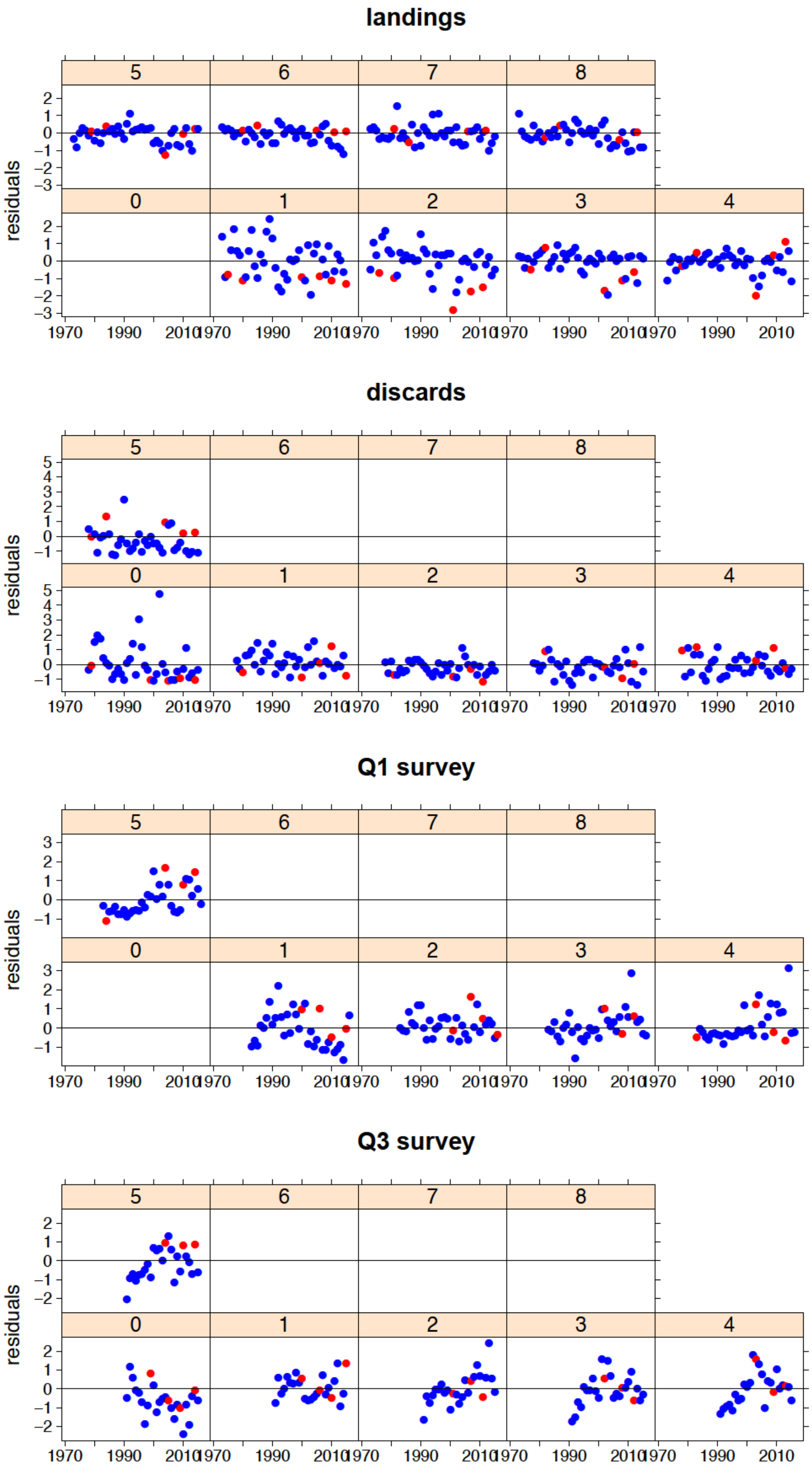

Figure 2.1.7. Residuals from the roll-over assessment.

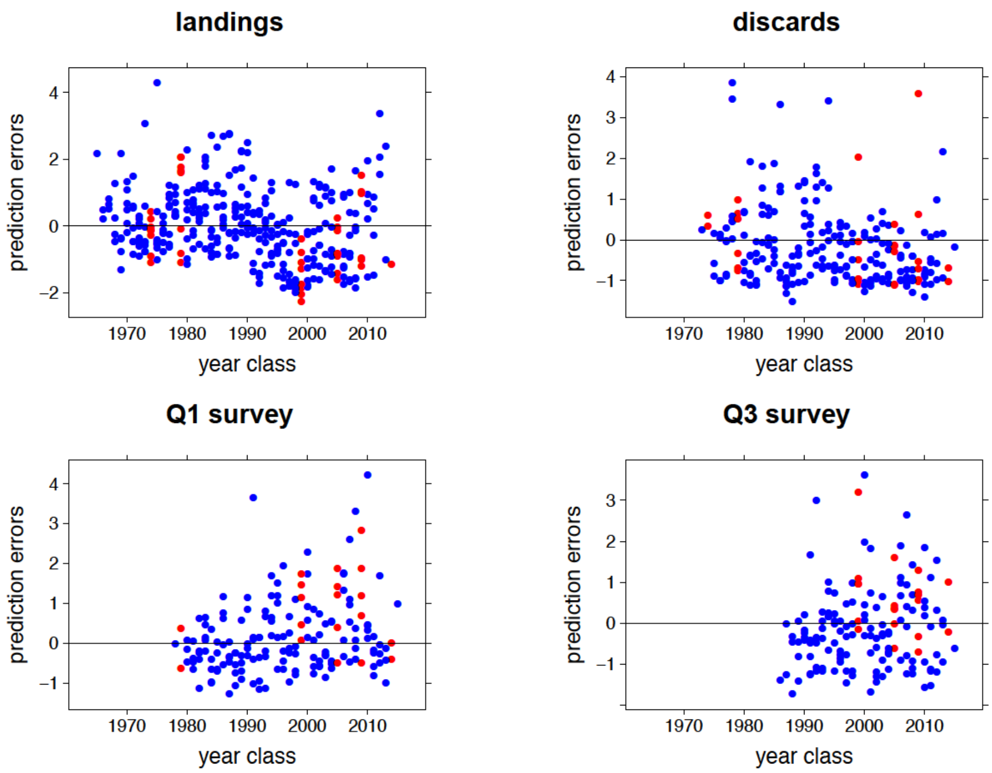

Figure 2.1.8. Prediction errors by year class from the roll-over assessment.

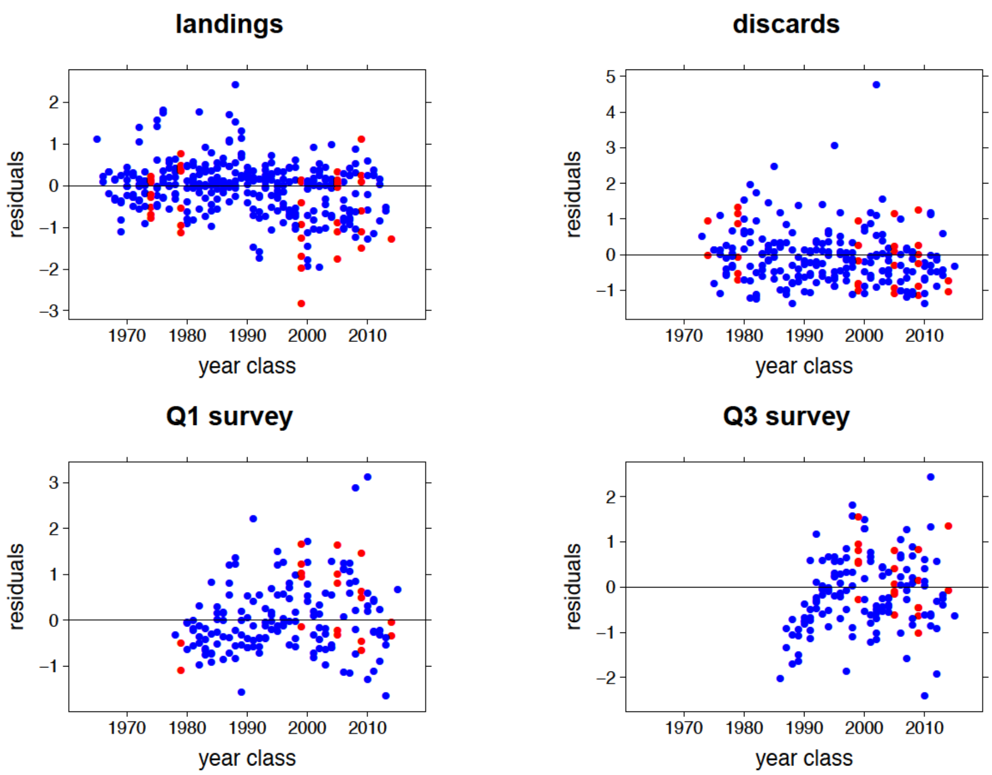

Figure 2.1.9. Residuals by year class from the roll-over assessment.

2.2 Alternative settings that provide solutions

To address the retrospective pattern in the roll-over assessment, we considered three alternative model specifications, labelled M1, M2 and M3.

1) M1 made minor changes to the fishing selection pattern and used external data to provide better weights for the landings and discards data. These changes were motivated by discussions during and subsequent to the Working Group in May, but made little difference to the model fit or the retrospective pattern.

2) M2 carried forward the changes made in M1 and also treated 2009 and 2014 as 'normal' year classes in the recruitment model. Thus, the large year classes were 1974, 1979, 1999 and 2005. This model reduced the retrospective pattern.

3) M3 carried forward the changes made in M1 and also treated 2005, 2009 and 2014 as 'normal' year classes in the recruitment model. Thus, the large year classes were 1974, 1979 and 1999. This model also reduced the retrospective pattern.

These models are described in more detail below.

Model M1

M1 makes two changes to the specification of the roll-over assessment:

1) In TSA, the fishing selection pattern is allowed to evolve stochastically over time, but is assumed to be flat above a specified age. In the roll-over fit, it was assumed to be flat for ages 5 and above. However, other assessment models suggest a domed selection pattern. To allow more scope for this, M1 assumes the selection pattern to be flat for ages 7 and above (above is the 8+ category), thus allowing more flexibility for ages 5, 6 and 7.

2) The landings data are assumed to be distributed with a common CV, φ_l say. This CV is estimated by maximum likelihood. However, prediction error and residual plots often suggest that some the landings data are more variable for some ages than others. The landings data are therefore weighted with age-specific weights which are based on the prediction error and residual plots. In the roll-over assessment, the landings for ages 1 through 8+ were assumed to have CVs of (2, 1, 1, 1, 1, 1, 2, 2) φ_l respectively. Thus, the CVs for ages 1, 7 and 8+ were assumed to be twice that of ages 2, 3, 4, 5 and 6. Similarly the discards for ages 0 through 5 were assumed to have CVs of (2, 1, 1, 1, 2, 2) φ_d respectively. (Note that all age 0 fish are assumed to be discarded, and all age 6+ fish are assumed to be landed).

External estimates of the CVs of the landings and discards data are available from 2009–2015. In M1, the median of these CVs by age was used to weight the landings and discards data for the whole time-series. Thus, the landings for ages 1 through 8+ were assumed to have CVs of (3.7, 1.3, 1, 1.1, 1.4, 1.6, 2.7, 2.8) φ_l and the discards for ages 0 through 5 were assumed to have CVs of (2.0, 1.7, 1, 1.5, 1.8, 2.4) φ_d. The parameters φ_l and φ_d were still estimated by maximum likelihood.

The fit of model M1 is similar to that of the roll-over assessment. The stock summary is shown in Figure 2.2.1, but the other diagnostic plots are not shown. Although not strictly comparable (because the weightings are different), the deviance (- 2 log-likelihood) of the two models are similar (roll-over: 651.7; M1: 650.2).

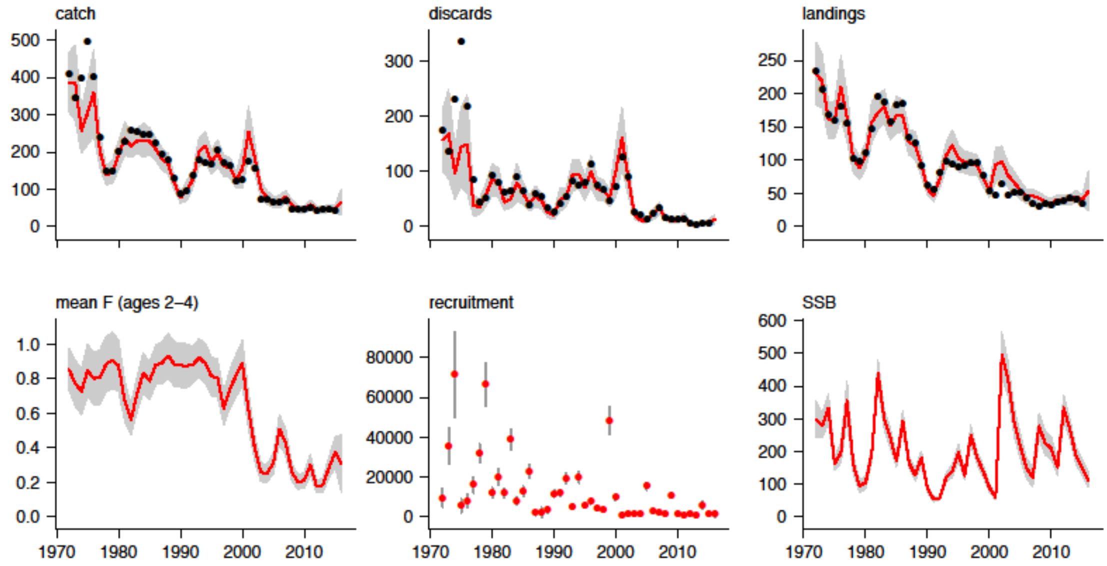

Figure 2.2.1. M1 stock summary.

Model M2

One of the reasons that recruitment in the roll-over assessment is dragged lower is because 2009 and 2014 are regarded as large year classes, so have little influence on the estimates of median recruitment. Instead, median recruitment tracks the surrounding year classes which all have low recruitment. To compensate for this, model M2 treats 2009 and 2014 as normal year classes, in the expectation that median recruitment will be pulled up to accommodate them. M2 thus regards only 1974, 1979, 1999 and 2005 as large year classes. (Note that M2 also adjusts the fishing selection pattern and the weights on the landings and discards data in the same way as M1).

Figure 2.2.2 shows the stock summary, 2.2.3 shows the retrospective summary, and 2.2.4 shows the recruitment retrospective. The retrospective pattern is markedly reduced (c.f. Figure 2.1.3).

Figure 2.2.5 shows the recruitment time-series. As expected, the estimates of median recruitment are higher than in the roll-over assessment (Figure 2.1.4). The M2 estimates of median recruitment in 2006 and 2016 are 386 and 318 respectively. The M2 estimate in 2006 is again lower than the estimate of 536 for 2006 from the fit when 2006 is the terminal year, but the difference is much less than for the roll-over assessment.

Figures 2.2.6 and 2.2.7 show the prediction errors and residuals by year and age, with the year classes treated as large in the roll-over assessment (1974, 1979, 1999, 2005, 2009, 2014) again highlighted in red. Figures 2.2.8 and 2.2.9 show the corresponding plots by year class. There is no evidence from the diagnostics that 2009 and 2014 are particularly unusual year classes.

There is no formal criterion for comparing models M2 and M1 and 'testing' whether 2009 and 2014 should be treated as large year classes. The deviances of the two models (M1: 650.2; M2: 687.4) suggest a big improvement in fit when 2009 and 2014 are treated as large. But the comparison is inappropriate to model selection because 2009 and 2014 have effectively been cherry-picked for special treatment. A certain element of expert judgement is called for!

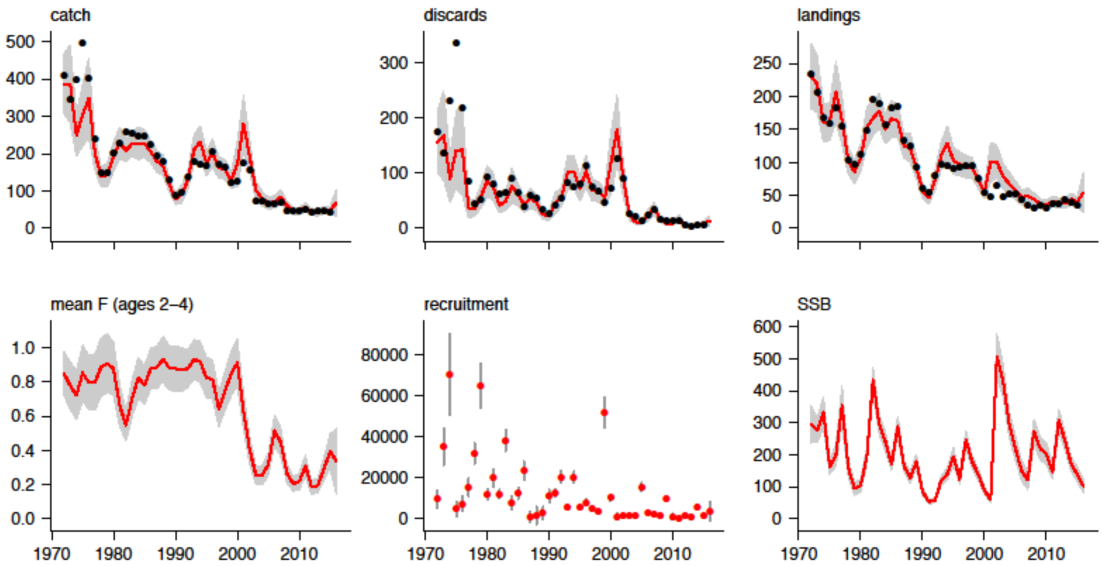

Figure 2.2.2. M2 stock summary.

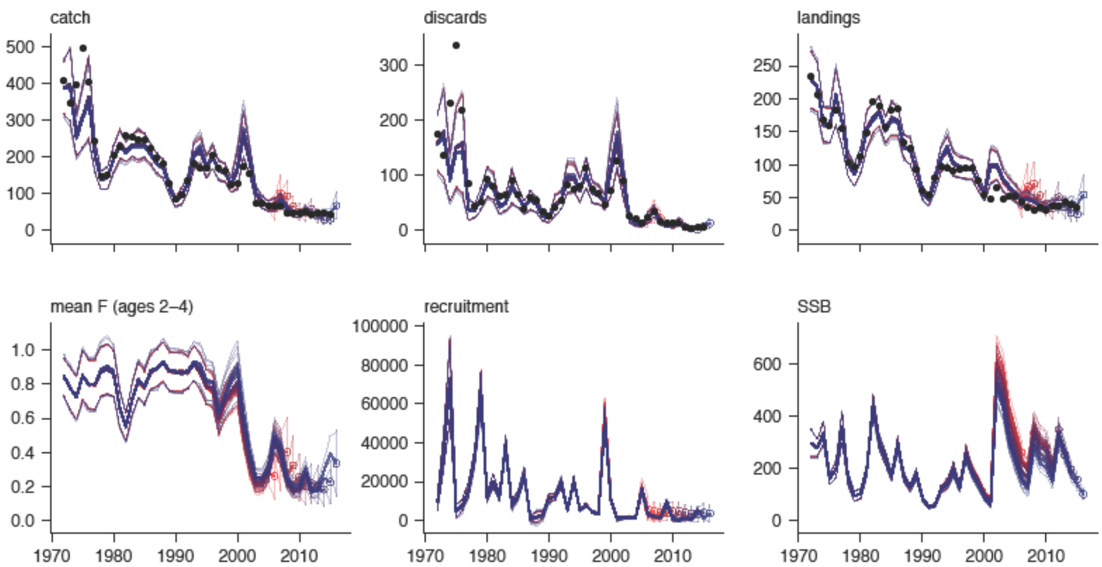

Figure 2.2.3. M2 retrospective.

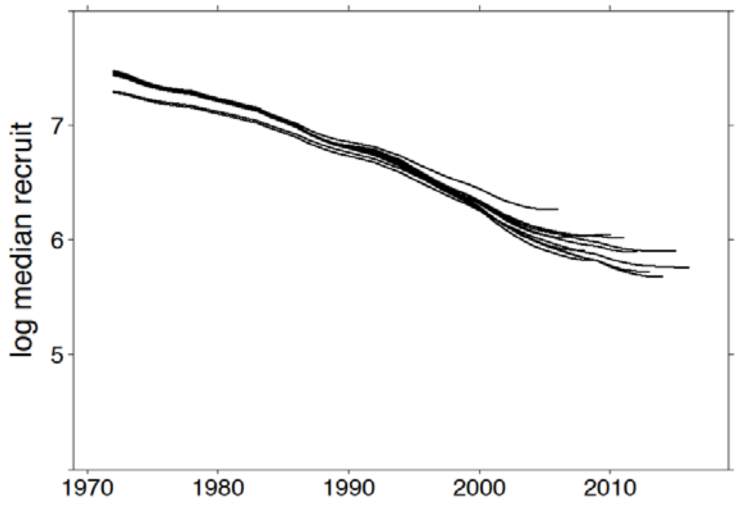

Figure 2.2.4. M2 recruitment retrospective. Recruitment estimates are plotted on the logarithmic scale, the scale on which recruitment is modelled. The limits on the vertical axis are the same as in Figure 2.1.3, the equivalent plot for the roll-over assessment.

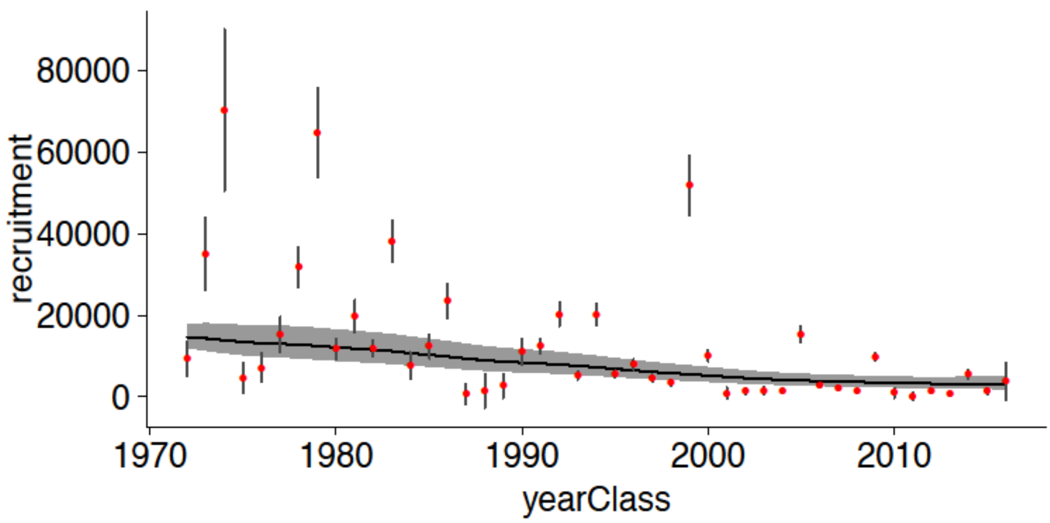

Figure 2.2.5. M2 recruitment time-series.

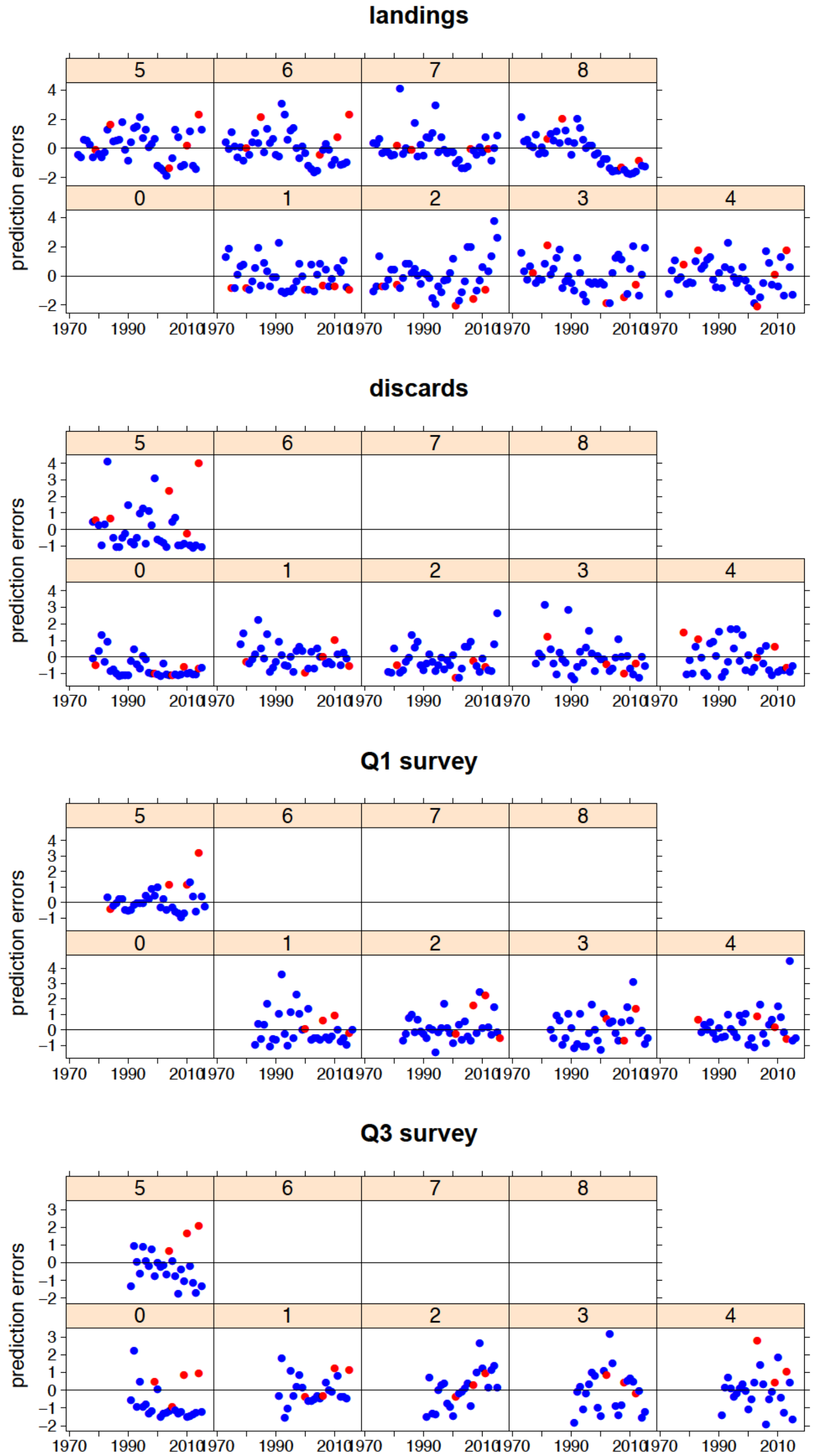

Figure 2.2.6. M2 prediction errors.

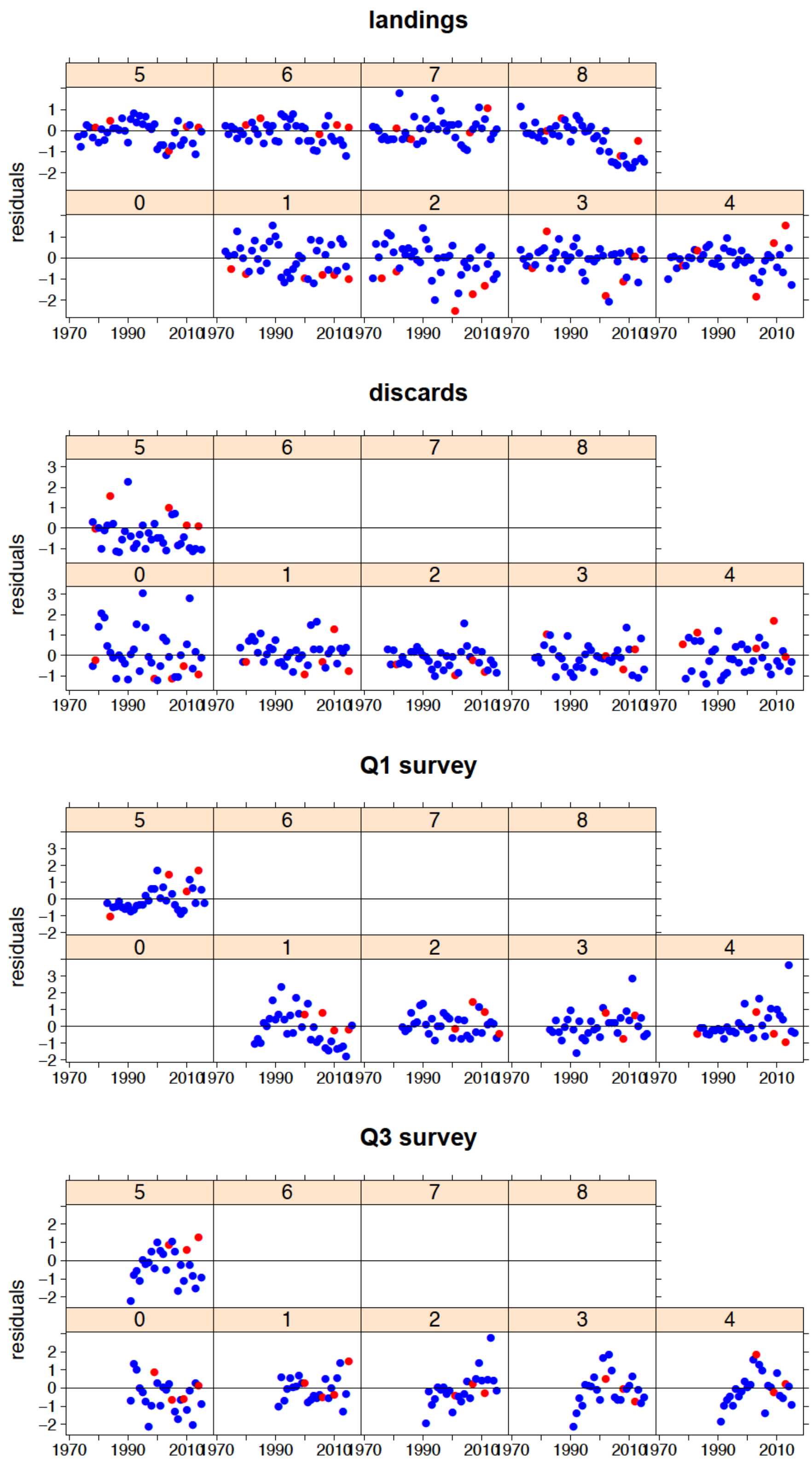

Figure 2.2.7. M2 residuals.

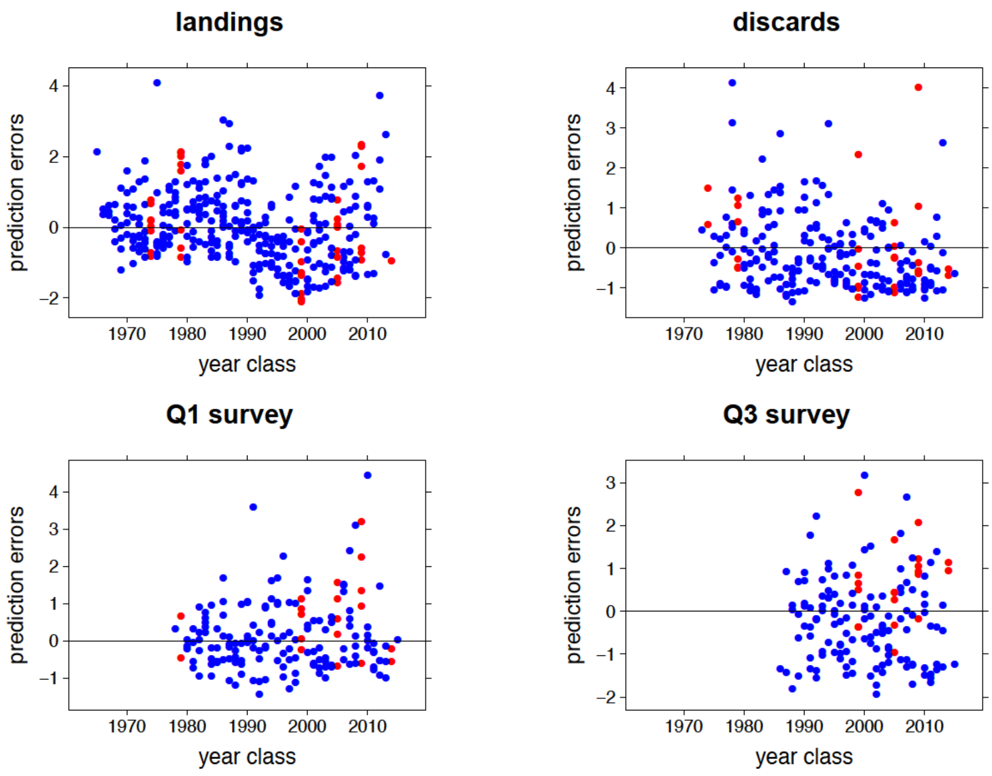

Figure 2.2.8. M2 prediction errors by year class.

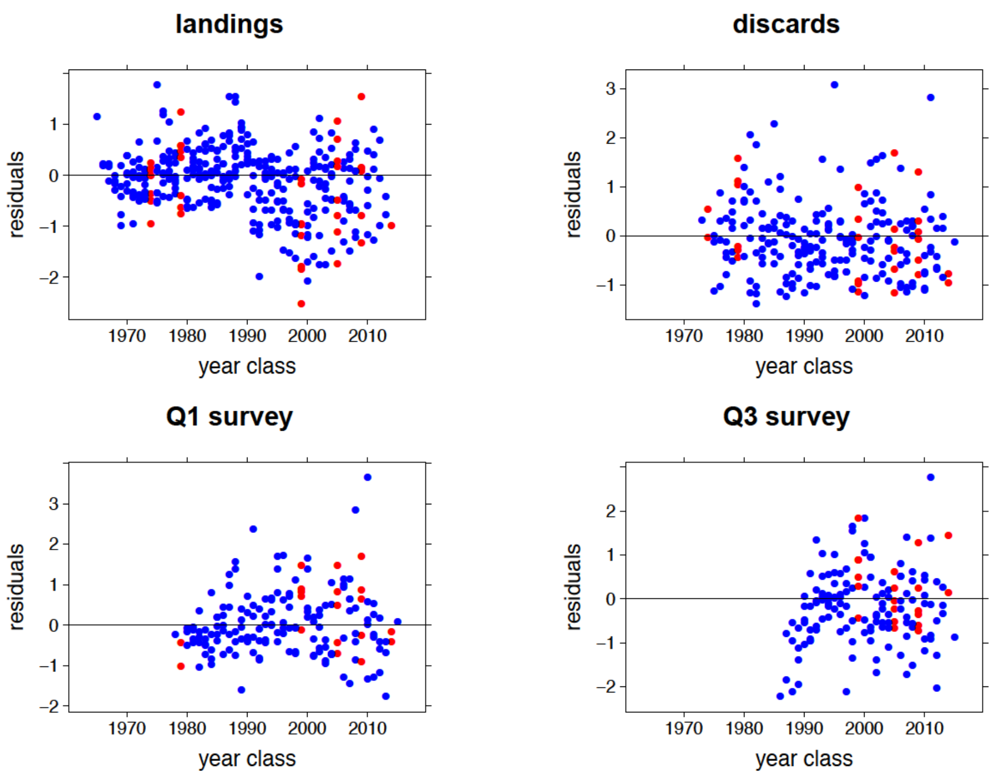

Figure 2.2.9. M2 residuals by year class.

Model M3

M3 extends M2 by treating 2005 as a normal year class. M3 thus regards only 1974, 1979 and 1999 as large year classes.

Figure 2.2.10 shows the stock summary, 2.2.11 shows the retrospective summary, and 2.2.12 shows the recruitment retrospective. Again, the retrospective pattern is much less than for the roll-over assessment.

Figure 2.2.13 shows the recruitment time-series. The M3 estimates of median recruitment in 2006 and 2016 are 462 and 378 respectively. The M3 estimate for 2006 is again lower than the estimate of 722 for 2006 from the fit when 2006 is the terminal year. But of course, 2005 is now treated as a normal year class, so when 2006 is the terminal year, the estimate of recruitment in 2006 will be influenced far more by the higher recruitment seen in 2005.

Figures 2.2.14 and 2.2.15 show the prediction errors and residuals by year and age, with the year classes treated as large in the roll-over assessment again highlighted in red. Figures 2.2.16 and 2.2.17 show the corresponding plots by year class. There is no evidence from the diagnostics that 2005, 2009 and 2014 are particularly unusual year classes.

The deviances of models M1, M2 and M3 are 650.2, 687.4 and 681.6 respectively.

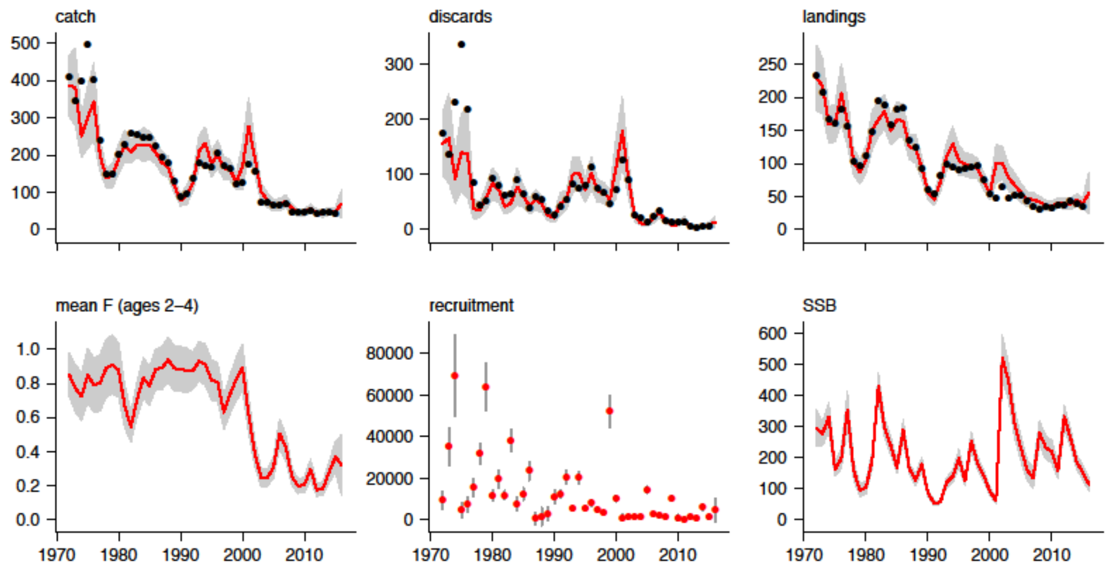

Figure 2.2.10. M3 stock summary

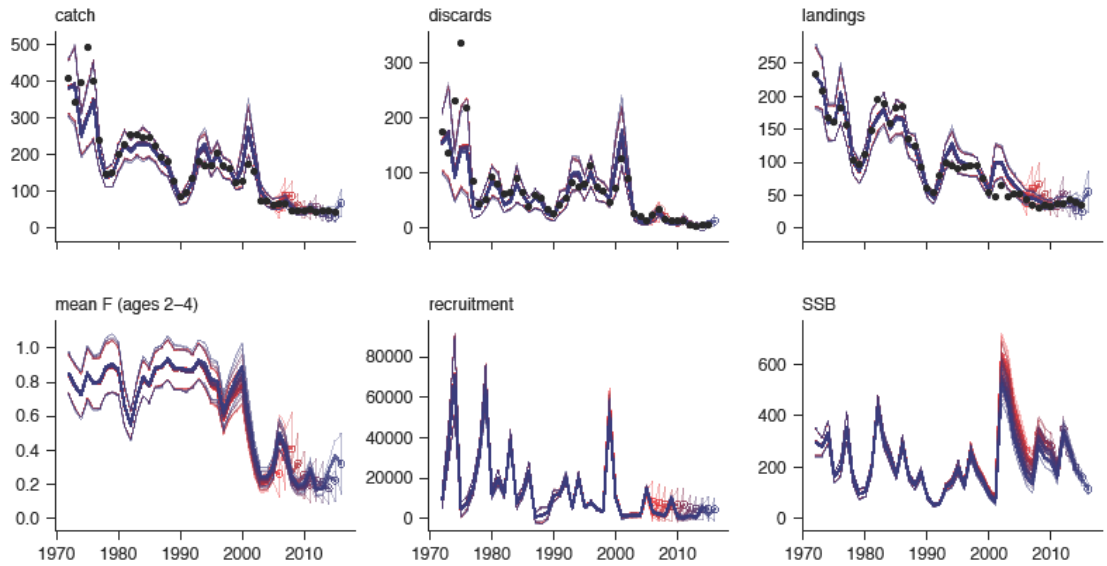

Figure 2.2.11. M3 retrospective.

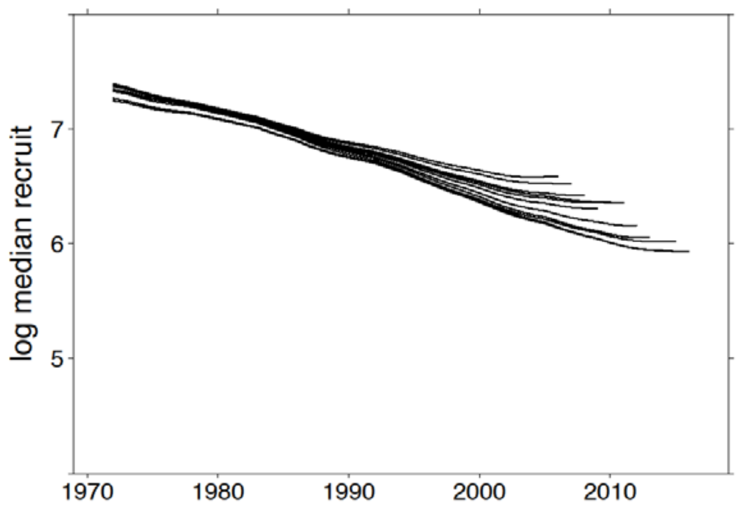

Figure 2.2.12. M3 recruitment retrospective. Recruitment estimates are plotted on the logarithmic scale, the scale on which recruitment is modelled. The limits on the vertical axis are the same as in Figure 2.1.3, the equivalent plot for the roll-over assessment.

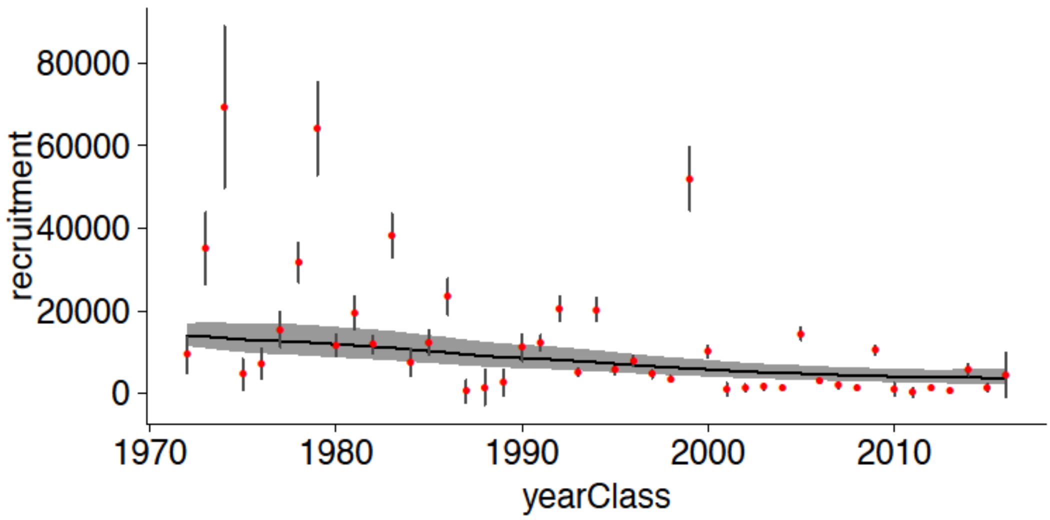

Figure 2.2.13. M3 recruitment time-series.

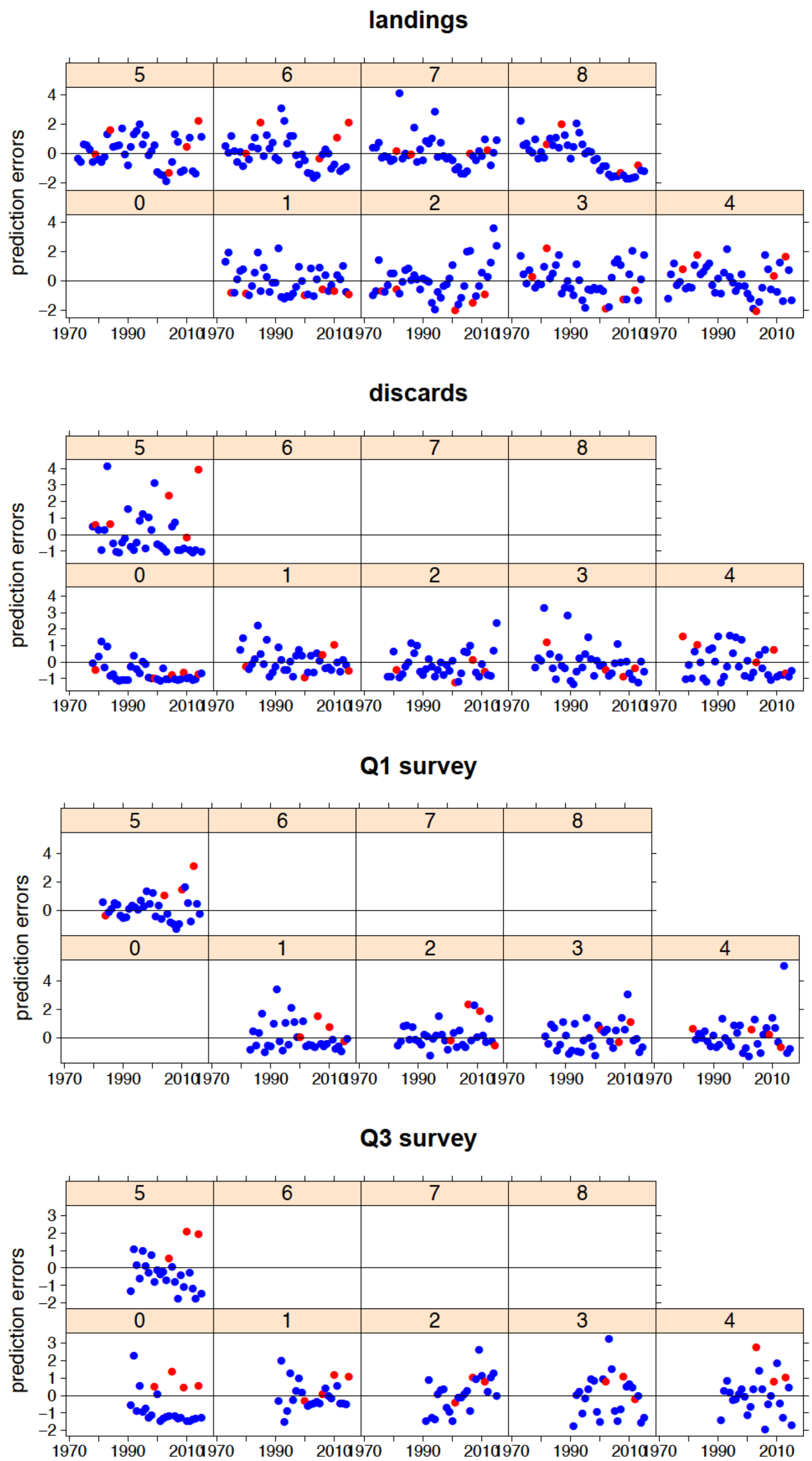

Figure 2.2.14. M3 prediction errors.

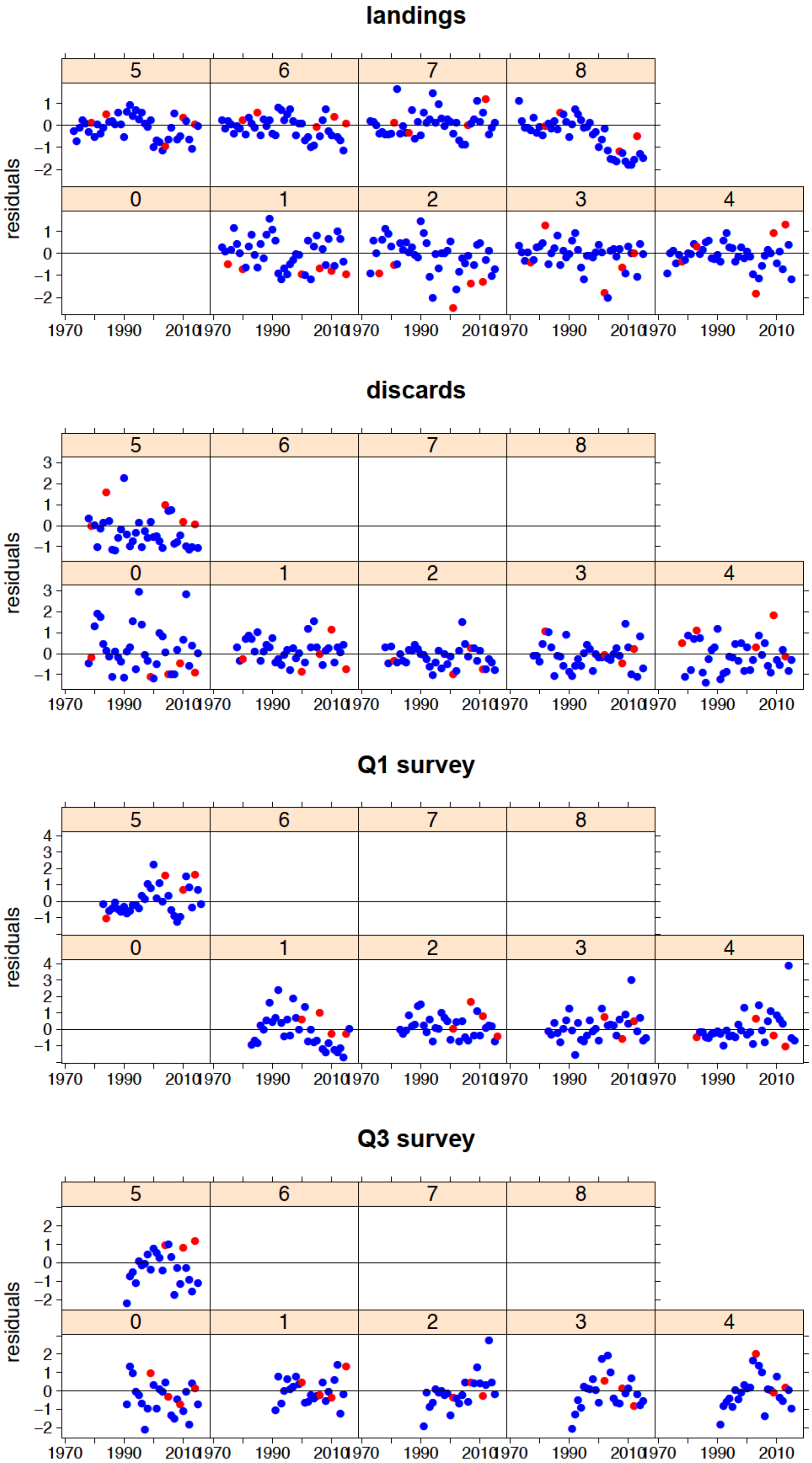

Figure 2.2.15. M3 residuals.

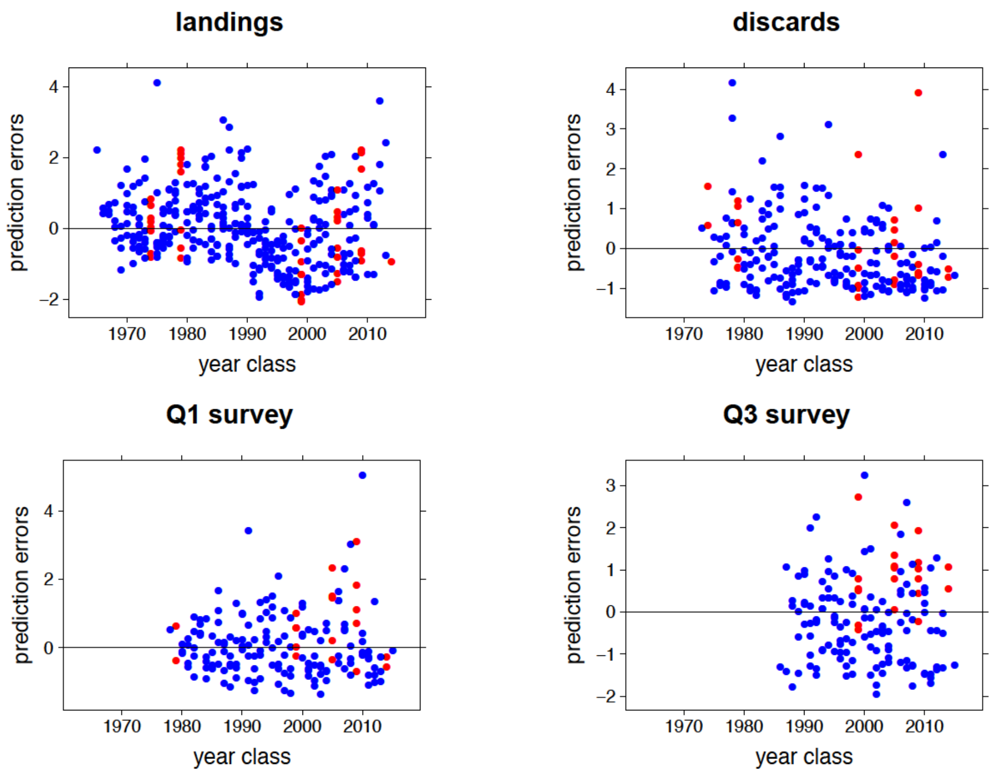

Figure 2.2.16. M3 prediction errors by year class.

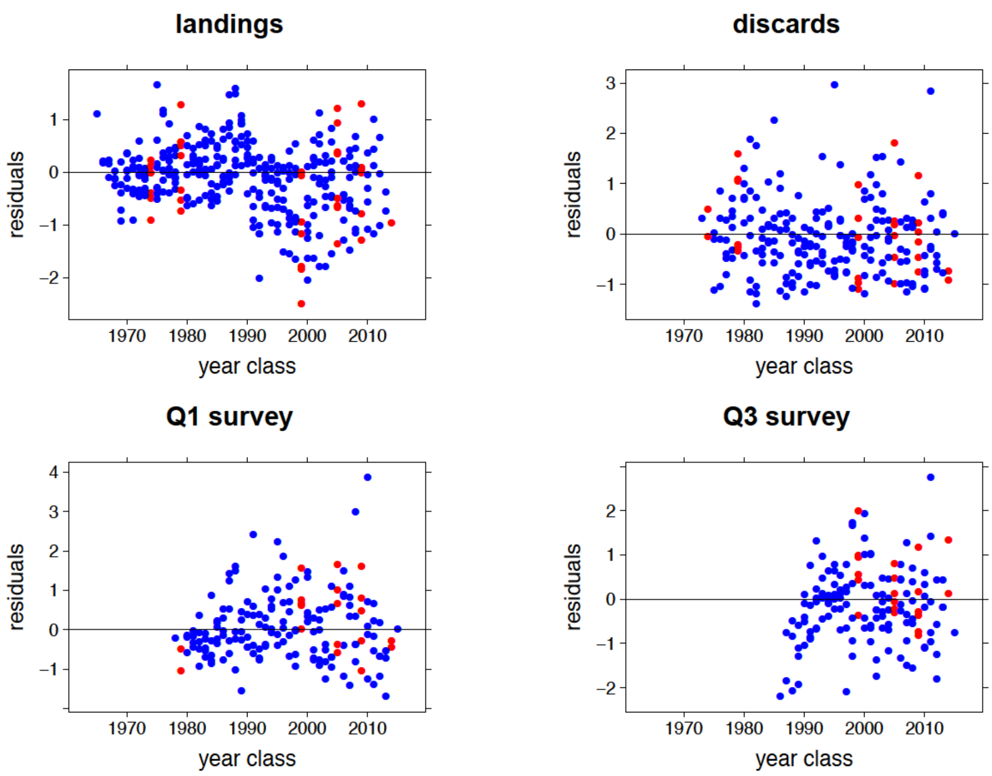

Figure 2.2.17. M3 residuals by year class.

2.3 Final proposed model

The final choice of model was M3. This was not much to choose between M2 and M3, but in the end, selection of M3 was on the basis of a slightly better deviance value (681.6 for M3 compared to 687.4 for M2), and the fact that there was no evidence from the diagnostics that 2005, 2009 and 2014 are particularly unusual year classes. Figure 2.3.1 provides the stock summary for the final model (M3) and Figure 2.2.13 the associated recruitment time-series compared to the median, while Figures 2.2.11–12 explore retrospective patterns, and Figures 2.14–17 prediction errors and residuals over time and by cohort for the various data sources.

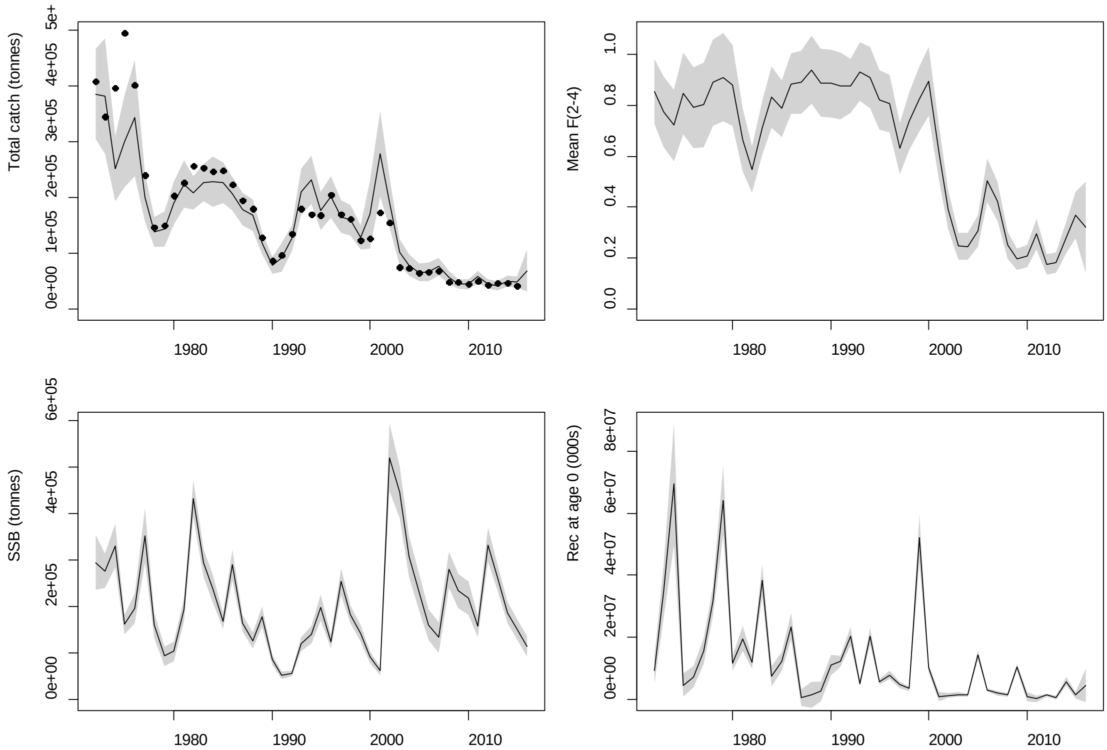

Figure 2.3.1. Stock summary for final proposed model (M3).

3 Tor b: Alternative models

A suitable model configuration for TSA was found (Section 2) and therefore there was no need to pursue ToR b. However, some comparisons with SAM and XSA were made.

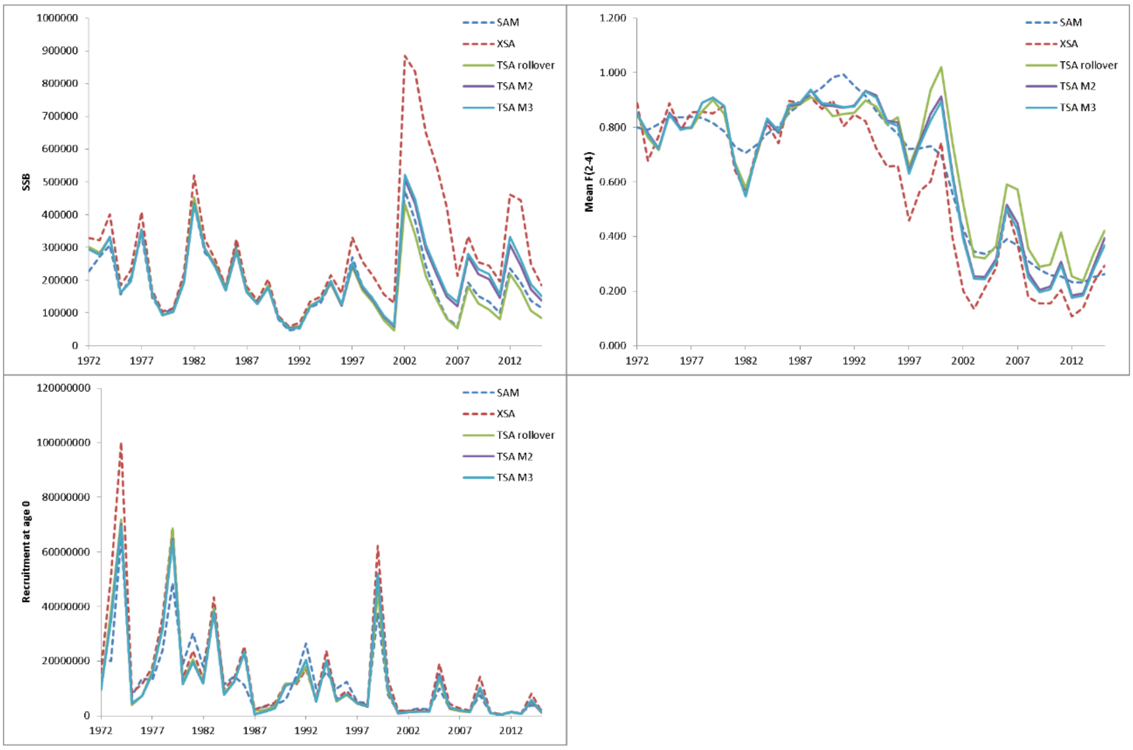

Figure 3.1. Comparisons of stock summary estimates from SAM, XSA and three TSA (rollover, M2 and M3) models.

The stock summary results from SAM, XSA and three TSA models all show similar trends over time in SSB, mean F and recruitment. XSA is notably more “optimistic” than the other models and has the highest estimates of SSB and recruitment and the lowest mean F. Meanwhile, the TSA rollover model is the most “pessimistic” and has the lowest SSB and recruitment estimates and highest mean F. However, the SSB and recruitment estimates from SAM are relatively close to those from the TSA rollover, though the mean F estimate from SAM is significantly much smoother than the other models. The summary results from the M2 and M3 TSA models both fall midway between the extremes of the TSA rollover and XSA and could be considered to be approximate to a “model mean”.

4 ToR c: Reference points and F_{MSY} ranges

4.1 Reference points

Reference points were derived following the ICES guidelines and using the software EqSim. Before running EqSim a decision was needed on which period of recruitment to use in the analysis. The TSA model accepted by this benchmark does not treat any year class after 1999 as an outstanding year class. To remain consistent with the absence of very large year classes in recent recruitment, it was decided that the period 2000+ would be used as the basis for deriving the reference points.

The re-estimated reference points are compared to the results obtained during WKHAD (2014) in Table 4.1.1 and proceeded as follows. Using the ICES guidelines for sporadic spawners, B_{lim} was revised to 94 kt (the estimated SSB for 1979, the smallest stock size to produce a good recruitment), and B(pa) was revised to 1.4 x B_{lim} = 132 kt (which was also used as the MSY $B_{trigger}$ value). An EqSim run with no advice error or rule generated F_{lim} = Fp50 = 0.38, and F_{pa} = F_{lim}/1.4 = 0.27. A second EqSim run with advice error but no advice rule produced an estimate of F_{MSY} = 0.24 with the range of 0.18 to 0.30 (Figure 4.2.1, top plot). However, an EqSim run with advice error and rule showed that Fp05 = 0.19 <F_{MSY} (Figure 4.2.1, bottom plot) so both F_{MSY} and the upper limit of the F_{MSY} range were constrained resulting in an F_{MSY} estimate of 0.19 and associated range of 0.18–0.19.

Table 4.1.1. Comparison of reference points derived from EqSIM to those derived during WKHAD (2014). Note that WKHAD did not follow the same estimation guidelines, and also based the values on a previous implementation of the TSA model.

Variable	WKHAD (2014)	IBPHaddock (2016)
B(lim)	63 kt	94 kt
B(pa)	88 kt	132 kt
F(lim)	n/a	0.38
F(pa)	n/a	0.27
F(MSY)	0.37	0.19

4.2 F_{MSY} ranges

Figure 4.2.1 illustrates the estimation of F_{MSY} and the F_{MSY} range (top plot), and the estimation of Fp05 (bottom plot) which, in this instance, constrains both the F_{MSY} and the upper limit of the F_{MSY} range.

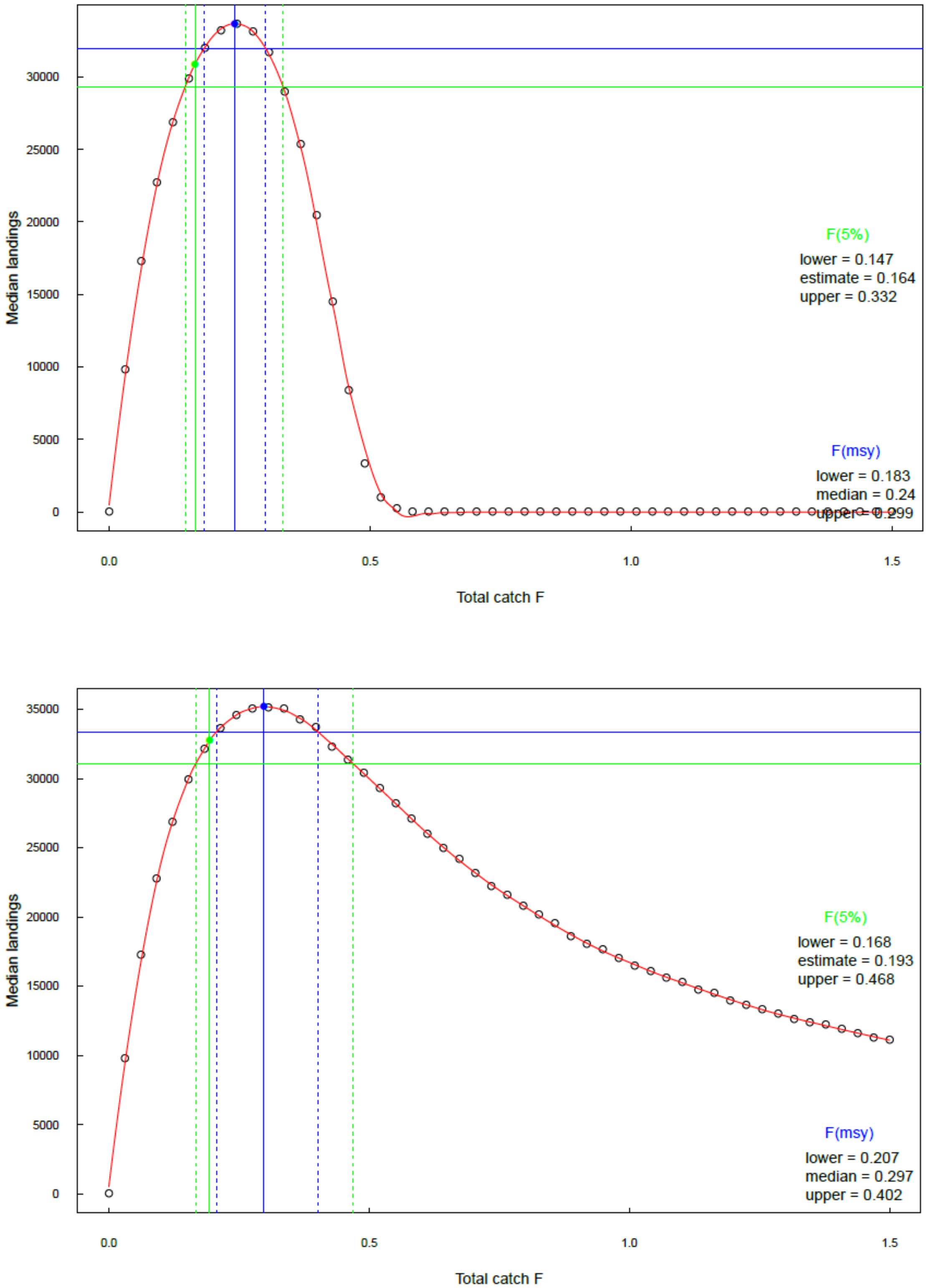

Figure 4.2.1. Results of EqSIM estimation of F(MSY) with the advice error but no rule (top) and of Fp05 with both advice error and rule (bottom).

5 Conclusions

5.1 General conclusions

- The retrospective pattern, which led to the rejection of the TSA model by the May 2016 WGNSSK meeting because of inconsistent advice from one year to the next, was due to the way the larger year classes are treated in the model.
- Alternative TSA models that no longer treat the larger year classes (post 1999) as outstanding, substantially reduce the retrospective pattern. The choice of one of these alternative models to be used for the assessment of the stock is supported by there being no evidence from the model diagnostics that the post-1999 year classes are particularly unusual. The TSA model that does not treat any of the post-1999 year classes as unusual (M3) was selected (on the grounds of a slightly better deviance value) over one that only treats the 2005 year class in the post-1999 period as unusual (M2).
- The selected TSA model (M3) is deemed adequate for assessment purposes (ToR a), and the inter-benchmark did not pursue models other than TSA (ToR b). Nevertheless, a comparison with other assessment models previously produced by WGNSSK (XSA and SAM) showed that the selected TSA model fell midway between the extremes of the rollover TSA assessment and XSA, and could be considered to approximate a "model mean".
- B_{lim} was selected as the lowest SSB to have produced an outstanding year class (1979), but all remaining reference points (ToR c) were estimated on the basis of the post-1999 period that lacked any outstanding year classes, and which is considered to represent current prevailing conditions for the stock (in terms of recruitment that can be expected).

5.2 Reviewers' comments and conclusions

The analysts' overview of the issues and responsiveness to concerns was excellent given the issues faced.

Whereas the diagnostics to choose between model approaches (i.e. TSA vs. SAM) were limited and lacked a strong statistically defensible basis, we conclude that the advice using the currently configured TSA method was reasonable. This is based on the comparisons of the modelling results that were presented and the fact that the currently configured TSA model appears intermediate to results arising from XSA and SAM. These alternative models have been reliably used in a similar manner for related stocks (e.g. North Sea cod). The clarification on how the retrospective patterns changed between years was appreciated. We note that these characteristics (i.e. the persistent pattern) are not unexpected given the propensity for this stock to have strong year classes, which by nature affect the expected value of recruitment. In this regard, the TSA model results seems sufficiently precautionary to proceed with its application for advice. We recommend that a more exhaustive comparison among the modelling approaches be conducted for the next benchmark.

Calculation of reference points was consistent with standard approach. The decision to base reference point calculations on recruitment post-1999 is consistent with the treatment of these recruitments in the revised TSA as being representative of prevailing conditions and that there was no evidence in model diagnostics that any of the 2000+ year classes were considered especially large. As EqSim lacks the capability to deal with occasional strong year classes, using recruitments from years with strong year classes in the period used to calculate reference points may be overly optimistic, given that there is no apparent evidence of a recent strong year class. The approach taken by the analysts therefore seems reasonable. We suggest exploration and investigation of alternative methods for calculating reference points that are able to incorporate and account for episodic strong year classes (e.g. via a mixture distribution or similar) and note that this work will likely be relevant for other stocks also.

6 References

ICES. WKHAD. 2014. Report of the ICES Benchmark Workshop for Northern Haddock Stocks. Marine Laboratory, Aberdeen, 27–29 January 2014 and ICES Headquarters, Copenhagen, 24–28 February 2014. ICES CM 2014/ACOM:41. 150 pp.

Annex 1: List of participants

Name	Address	Phone/Fax	E-mail
Harriet Cole	Marine Laboratory 375 Victoria Road Aberdeen AB11 9DB United Kingdom		H.Cole@marlab.ac.uk
Gavin Fay Invited Expert	Universite of Massachusetts Dartmouth School for Marine Science and Technology Dept of Fisheries Oceanography 200 Mill Road Suite 325 Fairhaven, MA 02719 US		gfay@umassd.edu
Robert Fryer	Marine Laboratory Scotland Marine Laboratory 375 Victoria Road ABERDEEN AB11 9DB UK	+44 1224 295502	R.Fryer@MARLAB.AC.UK Rob.Fryer@scotland.gsi.gov.uk
José De Oliveira Chair	Cefas Lowestoft Laboratory Pakefield Road Lowestoft Suffolk NR33 0HT UK	+44 1502 527 7 27 Fax +44 1502 524 511	jose.deoliveira@cefas.co.uk
Jim Ianelli Invited Expert	NOAA Alaska Fisheries Science Center Bldg. 4 7600 Sand Point Way Seattle, WA 98115 US	+1 206 526 6510	jim.ianelli@noaa.gov
Alexander Kempf	Institute of Sea Fisheries Palmaille 9 22767 Hamburg Germany	+49 40 38905 194	Alexander.kempf@thuenen.de
Scott Large	ICES H.C. Andersens Blvd. 44–46 1553 Copenhagen V. Denmark		scott.large@ices.dk
Coby Needle	Marine Laboratory 375 Victoria Road Aberdeen AB11 9DB United Kingdom	+44 1224 295456 / 87 6544	needlec@marlab.ac.uk

Annex 2: Working paper to ADGNS_Northern Shelf haddock (24 May 2016)

Proposed revisions to the WGNSSK assessment for Northern Shelf haddock

Coby Needle, Marine Scotland Science

Introduction

The ICES Working Group for the Assessment of Demersal Stocks in the North Sea and Skagerrak (Chairs Alexander Kempf and José de Oliveira) met during 26th April–5th May 2016 in Hamburg, Germany. As usual, one of the stocks covered was haddock in Subarea 4 and Divisions 6.a and 3.a (also known as Northern Shelf haddock); this has been assessed as a combined stock since the relevant benchmark meeting in 2015 (ICES, WKHAD 2014).

During the WGNSSK meeting, problems with retrospective bias became apparent in the haddock assessment (conducted using TSA since the benchmark). There proved to be insufficient time at the meeting itself to address this issue, and the decision was taken to take the extant assessment through to the advice stage. Work carried out since the end of the WGNSSK meeting has not been successful in rectifying the retrospective problem, and an alternative assessment using XSA has consequently been carried out.

This short note summarises the problem with the TSA assessment, presents the alternative XSA assessment, and provides the outcomes for the advice.

The WGNSSK TSA assessment

WGNSSK carried out the TSA assessment according to the specifications laid out in the relevant Stock Annex. As a time-series smoother (Fryer, 2001), TSA tends to provide relatively consistent stock estimates from year to year, and is not considered prone to retrospective bias. However, this was not the case this year: Figure 1 illustrates considerable retrospective bias in estimates of fishing mortality and SSB (although not recruitment). The bias is significant, in that the retrospective runs lie outwith the approximate pointwise 95% confidence interval of the full-year run, and they lead to considerable revisions in stock perception; the estimates for SSB and mean F(2–4) in 2015 change by -62 kt and +0.192, respectively. Retrospective revisions of this kind are problematic as they lead to very inconsistent advice, and must be addressed if possible. Exploratory runs carried out by WGNSSK show that such bias does not occur to the same extent for the SAM (Figure 2) or XSA (Figure 3) models. Model specifications in these cases were as used by the benchmark meeting (for SAM), or for the last XSA North Sea haddock assessment (ICES, WGNSSK 2013).

Due to work and resource pressures, data collation and assessment preparation was delayed this year, and the TSA assessment model was not run until well into the WGNSSK meeting. The retrospective problem therefore only became apparent around the half-way stage of the meeting, and considerable efforts went into attempting to understand and rectify it. These were not successful, however, and the decision was taken to revisit the problem during the week following the WGNSSK meeting. The post-WG work focused on two main hypotheses as potential drivers of the bias: variance partition, and selectivity doming.

Variance partition

First, there was the suggestion that the retrospective bias may have been caused by an unusual series of fishing mortality estimates in combination with the way that TSA partitions variance between transitory and persistent effects (Fryer, 2001). Figure 4 shows the log total catch data for year classes 2007–2012 (catch curves), with simple linear regressions fitted through ages 2–4. The slopes of these lines are very rough approximations to the mean fishing mortality experienced by these cohorts. We can see that, according to the catch data, the 2009 and 2010 cohorts suffered very low mortality over ages 2–4 (in 2011–2013 and 2012–2014 respectively): in fact, the regression slopes are positive, suggested negative mortality. The corresponding mortalities experienced by the 2011 and 2012 cohorts are much higher, with a strong negative slope for the 2011 cohort in particular. The mean F estimates in the assessment are of course collated from three different cohorts in each year, but the upshot from the catch data is that the fishing mortalities in 2012 and 2013 were probably extremely low, and that these were followed by two relatively high mortality years (2014 and 2015).

In the assessment carried out by WGNSSK in 2015, the two low Fs were interpreted by TSA as a persistent trend in fishing mortality, while the following single year of higher F was interpreted as a transitory deviation from that persistent low trend. In this year's assessment, there has been one further year of higher F, and TSA now interprets the two low F years has having been transitory deviations from a persistent F trend at a higher level. The model therefore revises the underlying trend in F upwards, and (correspondingly) revises SSB downwards. This is one plausible explanation for why the addition of a single year has led to the sudden appearance of a retrospective bias, and also why this has not been seen before (the two low F estimates were at the historical minimum).

Selectivity doming

The second suggested reason for the retrospective problem was possible doming in the fishery selectivity pattern. The TSA model stipulated in the Stock Annex specifies a selectivity plateau at-age 5; this is largely a legacy from the original implementation by Gudmundsson (1994), but in the past it has been seen as a reasonable assumption for haddock. Figure 5 compares the estimated exploitation patterns (averaged over decades) from exploratory SAM and XSA assessments, which do not have a corresponding assumption about selectivity at-age. Both SAM and XSA indicate that exploitation rates for older fish (age 6 and above) fell in the 2000s and 2010s. TSA is constrained so that estimated F plateaus at-age 5, so cannot show the same kind of decline; if anything, TSA shows that exploitation rates for older fish increase in the 2000s and 2010s. This is very different from the conclusion from both XSA and SAM, and is difficult to understand. It is also not clear how this feature would lead to retrospective bias, and work during the post-WG period was unable to determine whether this was influential or not. We did produce TSA runs without the selectivity plateau assumption, which did show some doming (although the TSA runs were probably non-optimal), but there was very little impact on the stock summaries which would suggest that doming is not in itself a causal factor.

Conclusion

The TSA assessment was conducted following the specification in the Stock Annex, and led to significant retrospective bias in both fishing mortality and SSB, to the extent that advice based on the assessment would be inconsistent from that provided last year. The bias was not present (at least not to the same extent) in exploratory SAM and

XSA runs. Two of the differences between TSA on the one hand, and SAM and XSA on the other, were a) the way that TSA partitions variance, which could in theory have led to the retrospective revisions, and b) assumptions and results on selectivity at-age. However, while WGNSSK were unable to reach a firm conclusion on the causes of the bias, it was clear that the bias was present and would have a significantly deleterious effect on the efficacy of the advice. Therefore, WGNSSK recommend that the update TSA assessment be **rejected** pending an inter-benchmark protocol on suitable assessment methods for Northern Shelf haddock.

The proposed XSA assessment

If TSA is rejected for the reasons given above, an alternative assessment must be provided in order to generate fisheries advice for 2017. Two exploratory runs were carried out by WGNSSK and subsequently. The SAM model run was provided by Anders Nielsen (DTU-Aqua) and has not been subject to sufficient WG scrutiny to enable it to be used as an alternative assessment, although it should certainly be considered in any future inter-benchmark exercise. The XSA model has also not been considered in depth by WGNSSK, but it was the model used for the last assessment of the North Sea haddock stock and thus has history as an accepted model for this stock. WGNSSK suggests that the addition of the Division 6.a component does not invalidate the use of the model as an alternative assessment, as this part of the stock is relatively small compared with the much larger North Sea component.

The XSA run was carried out using the settings from the 2013 WGNSSK run (ICES-WGNSSK 2013), namely:

Q plateau:	Age 6	
F shrinkage:	SE = 2.0	
Tuning indices:	IBTS Q1	Ages 1–5 Years 1983–2015
	IBTS Q3	Ages 0–5 Years 1991–2015
No power model		
No time-weighting		

Figures 6 and 7 show respectively the residuals and stock summaries from single-fleet XSA runs, using each of the available survey series (IBTS Q1 and Q3) separately. Residuals in both cases are small and without obvious pattern. Estimates of recruitment are similar, although the run using only IBTS Q3 does indicate higher SSB and lower mean F(2–4) overall (and particularly in the last year).

Table 1 gives the XSA diagnostic output from the full XSA run. Residuals from the full run (Figure 8) are reasonable. The stock summary from the full run (Figure 9) shows a lower fishing mortality, although rising to just above the new estimate of $F_{(MSY)}$ (see Table 8) in the final years; a higher though steeply declining SSB; and generally low recruitment since 2000; there have been three larger year classes in that time, although the size of these also appears to be declining.

Figure 10 compares the stock summaries from three assessments: the final TSA run from the 2016 WGNSSK meeting, the exploratory SAM run provided to the same meeting by Anders Nielsen, and the XSA run presented here. The XSA estimates of SSB are much higher than the TSA and SAM estimates, particularly since the appearance of the strong 1999 year class, and the mean F(2–4) estimates are correspondingly lower in most years since the early 1990s. The XSA estimates of the larger year classes are also higher. We also note that the survey-based assessment model SURBAR (results presented in the 2016 WGNSSK report) also indicates that SSB in 2002 was by far the highest in the available time-series.

Figure 7 suggested that the IBTS Q3 survey led to a more optimistic perception of the stock than the IBTS Q1 survey. Figure 11 tests this by comparing the WGNSSK TSA run, the full XSA run, and the single-fleet (IBTS Q1 only) XSA run. We see from this that the inclusion of IBTS Q3 does increase the SSB estimates in recent years slightly, but the Q3 survey does not explain the discrepancy with TSA.

Conclusions on model selection

In this short note, we have discussed the potential problems of retrospective bias in the TSA run, along with two potential hypotheses for why it has appeared this year. However, we are not in a position to determine which of these (if either) is the more influential, and we have not been able to modify the TSA model sufficiently to correct for the bias.

We considered two alternative models. The first is SAM, which was kindly provided for us by Anders Nielsen (DTU-Aqua) and which does not present significant retrospective or residual patterns; however, it has never been used as the update model for haddock and would need to be evaluated in a benchmark (or in an inter-benchmark) before it could be taken as the basis for advice. The second is XSA, which was the model used for North Sea haddock prior to the benchmark of 2014 and which thus has history as an assessment model for this stock. Similarly to SAM, XSA does not show retrospective or residual patterns. Also like SAM, it indicates some doming in the selection pattern in recent decades, which the update TSA was not able to model (although enabling doming in TSA does not bring the models much closer together).

Given that it appears to be internally consistent, and that it has previously been used for this stock, **WGNSSK recommend that XSA be taken forward as the basis for forecast and advice this year**. However, we also emphasise that this should only be considered as an interim measure, pending a benchmark or inter-benchmark protocol. The results of the XSA model are very different from the results of the TSA or SAM models, and without more detailed analysis is it difficult to know why this should be.

Appendix 1: Reference point estimation

Assessment WGs were asked this year to provide new estimates of appropriate fishing mortality and biomass reference points, following the EqSim-based protocol specified by ACOM. This was done by WGNSSK for both TSA and XSA assessments, and the results are compared in Table 2 with the previous reference-point values. The table indicates the settings of EqSim that were used for each estimate. The key difference with the previous estimate of F(MSY) (produced by WKHAD in 2014) is that the values are now derived from a segmented regression model fitted to stock–recruitment data from **2000-present only;** previously, the full 1972-present time-series had been used for this. Given the history of recruitment since the large 1999 year class (see Figures 1–3, for example), WGNSSK considers that a model which assumes that pre-2000 year classes are at all likely in future would be too optimistic and therefore not appropriate. It also transpires that the use of the longer recruitment time-series leads to very anomalous estimates for $F_{(lim)}$ and $F_{(pa)}$, with both being above 1.0 (see accompanying PDF files). Therefore, the recruitment model is limited to the 2000–2015 time-series.

The relevant EqSim runs are summarised in the following PDF files which accompany this note. Each PDF contains the results of four runs: no advice error or rule, advice error but no rule, advice rule but no error, and both advice error and rule.

File	Assessment model	Recruitment time-series
had346a EqSim_short_segreg3_tsa.pdf	TSA	2000–2015
had346a EqSim_long_segreg3_tsa.pdf	TSA	1972–2015
had346a EqSim_short_segreg3_xsa.pdf	TSA	2000–2015
had346a EqSim_long_segreg3_xsa.pdf	TSA	1972–2015

Appendix 2: Forecast settings and draft outcomes

Although not appropriate as the basis for choosing between assessment models, it is instructive to consider the potential consequences of the choice, and the following is given here for completeness. The forecast settings, reference point revisions and draft advice produced by WGNSSK, on the basis of the update TSA model, are given in Tables 3 (forecast settings), 4 (forecast inputs), and 5 (draft catch options table). The equivalents for the proposed XSA assessment are given in Table 6–8. Ultimately, both approaches (using the new estimates of $F_{(MSY)}$ as the basis for advice) would lead to a reduction in total catch: TSA leads to a 53% cut, while XSA gives a 10% cut. The decline appears to be an unavoidable consequence of the series of generally low recruitments that the stock has produced since the strong 1999 year class. However, the draft catch option tables (Tables 5 and 8) also include a column comparing the total catch in 2017 with the wanted catch in 2015; haddock discarding rates have been very low in recent years, so this comparison is not without merit. The fact that the haddock quota is very rarely fully utilised, along with the 30% increase in the total catch quota in 2016, means that the comparison of total catch 2017 with wanted catch 2015 indicated just a 4% cut (TSA) or an 85% increase (XSA). So, in relation to the fish that skippers have been catching, the forecast for 2017 is not particularly harsh for either assessment. However, the fact remains that a 30% increase in 2016 followed by a 53% decrease in 2017 would be presentationally very difficult to justify, and would suggest very inconsistent advice.

References

Fryer, Rob F. 2001. TSA: Is It the Way? Working Document to the ICES Working Group on Methods of Fish Stock Assessments, Copenhagen, December 2001.

Gudmundsson, G. 1994. Time-series Analysis of Catch-at-Age Observations. *Applied Statistics*, 43, 117–126.

ICES. WGNSSK. 2013. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK), 24–30 April 2013, ICES Headquarters, Copenhagen. ICES CM 2013/ACOM:13.

ICES. WKHAD. 2014. Report of the ICES Benchmark Workshop for Northern Haddock Stocks. Marine Laboratory, Aberdeen, 27–29 January 2014 and ICES Headquarters, Copenhagen, 24–28 February 2014. ICES CM 2014/ACOM:41. 150 pp.

Table 1. Northern Shelf haddock. XSA diagnostic output.

FLR XSA Diagnostics 2016-05-17 09:11:19

cpue data from x.idx

Catch data for 44 years. 1972 to 2015. Ages 0 to 8.

	fleet	first age	last age	first year	last year	alpha	beta
1	North Sea IBTS Q1	1	5	1983	2015	0	0.25
2	North Sea IBTS Q3	0	5	1991	2015	0.5	0.75

Time-series weights :

Tapered time weighting not applied

Catchability analysis :

Catchability independent of size for all ages

Catchability independent of age for ages >= 6

Terminal population estimation :

Survivor estimates shrunk towards the mean F
of the final 5 years or the 3 oldest ages.

S.E. of the mean to which the estimates are shrunk = 2

Minimum standard error for population
estimates derived from each fleet = 0.3

prior weighting not applied

Regression weights

age \ year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
all	1	1	1	1	1	1	1	1	1	1

Fishing mortalities

age \ year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
0	0.001	0.001	0.004	0.001	0.004	0.011	0.002	0.003	0.001	0.005
1	0.045	0.031	0.035	0.036	0.044	0.024	0.039	0.044	0.052	0.017
2	0.393	0.177	0.130	0.094	0.151	0.069	0.053	0.133	0.200	0.319
3	0.531	0.483	0.134	0.146	0.192	0.281	0.116	0.082	0.286	0.378
4	0.592	0.464	0.269	0.226	0.123	0.262	0.154	0.200	0.217	0.184
5	0.373	0.591	0.215	0.174	0.179	0.199	0.113	0.103	0.237	0.255
6	0.123	0.238	0.395	0.136	0.119	0.168	0.063	0.069	0.073	0.186
7	0.072	0.071	0.089	0.258	0.094	0.135	0.082	0.030	0.059	0.097
8	0.072	0.071	0.089	0.258	0.094	0.135	0.082	0.030	0.059	0.097

XSA population number (Thousand)

year \ age	0	1	2	3	4	5	6	7	8
2006	4375064	5416192	120480	58329	28949	22131	158956	525863	3866
2007	2685809	1287317	1347207	51049	25329	12172	11842	111545	118931
2008	1857299	821688	333259	699411	23240	12072	5215	7398	52360
2009	14345891	587726	218759	178840	449420	13383	7483	2782	5321
2010	1393320	4698234	161322	119882	113021	268370	8563	5165	6039
2011	479042	467226	1313451	82243	71776	74001	168564	5992	4046
2012	1428904	163048	135729	714018	44532	40443	44878	111843	7509
2013	746077	501250	47298	73479	450526	27566	26268	32921	147651
2014	8080939	267187	146172	23169	47372	262543	17762	19059	41135
2015	1379571	2900194	77304	66923	12182	27138	147923	12833	21250

Estimated population abundance at 1st Jan 2016

year \ age	0	1	2	3	4	5	6	7	8
2016	530376	493208	868838	31420	32082	7211	15008	95479	9360

Table 1 cont. Northern Shelf haddock. XSA diagnostic output.

Fleet: North Sea IBTS Q1

Log-catchability residuals.

age	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
1	-0.330	-0.252	-0.295	0.238	0.058	0.258	0.600	0.219	0.406	0.967	0.607	-0.082	0.259	0.023
2	-0.095	-0.184	-0.101	0.295	0.019	0.034	0.506	0.432	-0.008	-0.103	0.377	-0.131	-0.048	-0.107
3	-0.038	-0.139	0.180	-0.078	-0.239	0.128	0.241	0.603	0.026	-0.915	0.201	-0.182	-0.205	-0.254
4	-0.067	0.050	0.128	-0.074	-0.095	0.098	0.139	0.159	0.109	-0.485	0.011	-0.139	-0.135	-0.169
5	0.175	-0.246	-0.033	0.129	0.163	0.088	-0.069	0.046	-0.056	-0.354	0.084	-0.228	-0.178	0.097

age	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
1	0.748	0.486	0.159	0.271	0.397	-0.320	0.110	-0.200	-0.176	0.250	-0.677	-0.810	-0.277	-0.162
2	0.402	0.244	0.218	-0.292	-0.135	-0.105	-0.428	0.199	-0.253	-0.338	0.263	-0.538	0.156	-0.218
3	0.211	-0.060	0.057	-0.308	0.427	-0.072	-0.235	0.116	0.203	-0.105	0.256	-0.500	0.014	0.006
4	0.132	-0.026	0.235	-0.157	-0.111	-0.510	-0.168	0.190	-0.030	-0.134	0.397	0.455	-0.302	0.099
5	0.061	0.168	0.146	0.121	-0.093	0.153	-0.174	0.064	-0.224	-0.256	-0.103	-0.779	-0.424	0.163

age	2011	2012	2013	2014	2015
1	-0.542	-0.680	-0.379	-0.632	-0.240
2	0.074	-0.121	-0.050	0.274	-0.235
3	0.797	-0.037	-0.250	0.355	-0.204
4	0.212	0.149	-0.670	0.955	-0.247
5	0.814	0.256	-0.205	0.525	0.168

Mean log-catchability and standard error of ages with catchability independent of year-class strength and constant w.r.t. time

	1	2	3	4	5
Mean_Logq	-13.1366	-12.0305	-12.0224	-12.3456	-12.7005
S.E_Logq	0.4406	0.2583	0.3168	0.2965	0.2809

Fleet: North Sea IBTS Q3

Log-catchability residuals.

age	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
0	0.097	1.003	0.465	0.177	0.326	0.089	-0.467	0.080	0.545	0.240	-1.190	0.159	0.260	0.435
1	-0.464	0.504	0.085	-0.054	0.149	0.100	0.051	0.358	0.186	0.051	-0.325	-0.203	0.074	0.024
2	-0.915	0.156	-0.089	0.001	0.051	-0.020	0.025	0.093	0.134	-0.337	-0.245	-0.285	-0.061	0.186
3	-0.980	-0.232	-0.056	-0.194	0.253	0.038	0.023	0.183	0.174	-0.124	0.488	-0.104	0.310	0.427
4	-0.924	0.063	-0.166	-0.035	-0.156	0.073	-0.250	0.046	-0.143	0.034	-0.023	0.420	0.217	0.201
5	-0.592	0.170	0.188	-0.562	0.308	0.069	-0.034	0.019	-0.161	-0.370	0.059	-0.028	-0.004	0.090

age	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
0	0.114	-0.433	-0.511	-0.037	-0.130	-1.215	-0.145	-0.458	0.416	0.191	-0.008
1	0.028	-0.280	-0.164	-0.279	-0.072	-0.191	0.114	0.274	-0.237	0.207	0.064
2	0.274	-0.139	-0.040	-0.120	0.171	0.216	-0.156	-0.031	0.616	0.373	0.142
3	0.208	-0.048	-0.067	-0.071	-0.221	0.105	0.313	-0.361	-0.078	-0.052	0.065
4	0.412	-0.144	0.305	0.246	-0.085	0.161	-0.016	-0.105	0.041	0.095	-0.267
5	0.173	0.278	-0.077	0.395	-0.043	0.450	0.002	-0.174	-0.344	0.507	-0.320

Mean log-catchability and standard error of ages with catchability independent of year-class strength and constant w.r.t. time

	0	1	2	3	4	5
Mean_Logq	-13.447	-12.6609	-12.3150	-12.5514	-12.9109	-13.2886
S.E_Logq	0.502	0.2280	0.2845	0.2946	0.2691	0.2879

Terminal year survivor and F summaries:

```
 Age 0 Year class =2015

source
                  scaledWts survivors yrcls
North Sea IBTS Q3     0.938    489120  2015
fshk                  0.062    559682  2015

 Age 1 Year class =2014

source
                  scaledWts survivors yrcls
North Sea IBTS Q1     0.306    683128  2014
North Sea IBTS Q3     0.679    925943  2014
fshk                  0.016    346457  2014
```

Table 1 cont. Northern Shelf haddock. XSA diagnostic output.

```
 Age 2 Year class =2013

source
                  scaledWts survivors yrcls
North Sea IBTS Q1     0.492     24827  2013
North Sea IBTS Q3     0.492     36212  2013
fshk                  0.015     90129  2013

 Age 3 Year class =2012

source
                  scaledWts survivors yrcls
North Sea IBTS Q1     0.458     26162  2012
North Sea IBTS Q3     0.525     34242  2012
fshk                  0.017     69108  2012

 Age 4 Year class =2011

source
                  scaledWts survivors yrcls
North Sea IBTS Q1     0.492      5633  2011
North Sea IBTS Q3     0.495      5521  2011
fshk                  0.013      6863  2011

 Age 5 Year class =2010

source
                  scaledWts survivors yrcls
North Sea IBTS Q1     0.493     17759  2010
North Sea IBTS Q3     0.493     10898  2010
fshk                  0.014     23964  2010

 Age 6 Year class =2009

source
     scaledWts survivors yrcls
fshk         1    187776  2009

 Age 7 Year class =2008

source
     scaledWts survivors yrcls
fshk         1      4069  2008
NULL
```

Table 2. Northern Shelf haddock. Comparison of previous reference point values with new estimates (based on TSA and XSA assessments). Note that all new points are derived using a segmented regression model fitted to stock–recruit data from 2000–2015 only.

Reference point	Previous value	TSA-based estimate	XSA-based estimate	Basis for new estimate
B(lim)	100 kt	96 kt	105 kt	Lowest parental SSB (1979) for large year class
B(pa)	140 kt	135 kt	147 kt	B(lim) * 1.4
B(trigger)	140 kt	135 kt	147 kt	B(pa)
F(lim)	Not defined	0.38	0.44	F(50) from EqSim run with no advice error or rule
F(pa)	Not defined	0.27	0.32	F(lim) / 1.4
F(MSY)	0.37	0.20	0.26	F(MSY) from EqSim run with advice error but no rule
F(MSY) range	Not defined	0.14 to 0.26	0.21 to 0.36	EqSim run with advice error but no rule

Table 3. Northern Shelf haddock. Basis for forecast settings (using update TSA assessment).

Weights-at-age	Linear extrapolation following method of Jaworski (2011)
Fishing mortality	2016 TSA estimate (partitioned into catch components using three-year mean proportions at-age)
Recruitment at-age 0 in 2016	2016 TSA estimate (based on underlying random walk recruitment model)=1140 million
Recruitment at-age 0 in 2017	Set equal to 2016 value

Table 4. Northern Shelf haddock. MFDP input data (using update TSA assessment).

2016							2017						
Age	N	M	Mat	PF	PM	SWt	Age	N	M	Mat	PF	PM	SWt
0	1140658	1.02	0	0	0	0.036	0	1140658	1.020	0	0	0	0.036
1	359712	1.19	0	0	0	0.152	1	.	1.190	0	0	0	0.152
2	533734	0.58	0	0	0	0.425	2	.	0.580	0	0	0	0.425
3	28465	0.36	1	0	0	0.580	3	.	0.360	1	0	0	0.583
4	23261	0.34	1	0	0	0.760	4	.	0.340	1	0	0	0.771
5	3820	0.34	1	0	0	0.958	5	.	0.340	1	0	0	0.944
6	4388	0.25	1	0	0	0.936	6	.	0.250	1	0	0	1.148
7	21572	0.22	1	0	0	0.763	7	.	0.22	1	0	0	1
8	4216	0.2	1	0	0	0.948	8	.	0.2	1	0	0	1

Catch					Catch				
Age	Sel	CWt	DSel	DCWt	Age	Sel	CWt	DSel	DCWt
0	0.000	0.000	0.006	0.036	0	0.000	0.000	0.006	0.036
1	0.002	0.356	0.046	0.141	1	0.002	0.356	0.046	0.141
2	0.164	0.523	0.098	0.267	2	0.164	0.523	0.098	0.267
3	0.320	0.620	0.048	0.369	3	0.320	0.620	0.048	0.348
4	0.409	0.654	0.016	0.479	4	0.409	0.642	0.016	0.482
5	0.516	1.016	0.019	0.497	5	0.516	0.726	0.019	0.587
6	0.525	0.857	0.010	0.551	6	0.525	1.197	0.01	0.588
7	0.530	0.708	0.005	0.491	7	0.53	0.957	0.005	1
8	0.533	0.83	0.002	0.714	8	0.533	0.87	0.002	1

2016				2017		
IBC				IBC		
Age	Sel	CWt		Age	Sel	CWt
0	0.000	0		0	0.000	0
1	0.000	0.3562		1	0.000	0.3562
2	0.000	0.5228		2	0.000	0.5228
3	0.000	0.6205		3	0.000	0.6205
4	0.000	0.6425		4	0.000	0.6425
5	0.001	0.6619		5	0.001	0.6619
6	0.001	0.7576		6	0.001	0.7576
7	0.001	0.7728		7	0.001	0.7728
8	0.001	1.41		8	0.001	1.41

2018						
Age	N	M	Mat	PF	PM	SWt
0	1140658	1.02	0	0	0	0.036
1	.	1.19	0	0	0	0.152
2	.	0.58	0	0	0	0.425
3	.	0.36	1	0	0	0.583
4	.	0.34	1	0	0	0.632
5	.	0.34	1	0	0	0.963
6	.	0.25	1	0	0	1
7	.	0.22	1	0	0	1

2016							2017		
8	.	0.2	1	0	0	1			
Catch							IBC		
Age	Sel	CWt	DSel	DCWt			Age	Sel	CWt
0	0	0	0.006	0.036			0	0	0.000
1	0.002	0.356	0.046	0.141			1	0	0.356
2	0.164	0.523	0.098	0.267			2	0	0.523
3	0.32	0.62	0.048	0.348			3	0	0.621
4	0.409	0.642	0.016	0.372			4	0	0.643
5	0.516	0.662	0.019	0.596			5	0.001	0.662
6	0.525	0.799	0.01	0.696			6	0.001	0.7576
7	0.53	1.379	0.005	0.679			7	0.001	0.7728
8	0.533	0.966	0.002	0.888			8	0.001	1.41

Table 5. Northern Shelf haddock. Draft advice table (using update TSA assessment).

Rationale	Total catch 2017	Wanted catch 2017	Unwanted catch 2017	IBC 2017	Basis	Total F 2017	F(land) 2017	F(disc) 2017	F(IBC) 2017	SSB 2018	% SSB change	% TAC change	% change from 2015 landings
MSY	33.741	29.904	3.820	0.021	New F(msy) estimate	0.200	0.169	0.031	0.000	136.556	-20%	-53%	-4%
Management plan	48.600	43.011	5.567	0.020	MP target F	0.300	0.254	0.046	0.000	122.852	-28%	-33%	38%
IBC only	0.024	0.000	0.000	0.024	No HC fishery	0.000	0.000	0.000	0.000	168.380	-2%	-100%	-100%
Other options	43.368	38.402	4.946	0.020	0.75 * F(sq)	0.264	0.223	0.041	0.000	127.665	-25%	-40%	23%
	55.681	49.238	6.424	0.019	Fsq	0.352	0.298	0.054	0.000	116.309	-32%	-23%	58%
	67.098	59.248	7.827	0.018	1.25 * F(sq)	0.440	0.372	0.068	0.000	105.965	-38%	-7%	90%
	64.156	56.674	7.460	0.018	15% TAC decrease (full)	0.416	0.352	0.064	0.000	108.634	-37%	-15%	82%
	73.827	65.127	8.679	0.017	Rollover TAC (full)	0.496	0.420	0.076	0.000	99.885	-42%	0%	109%
	82.641	72.793	9.830	0.016	15% TAC increase (full)	0.576	0.487	0.088	0.000	91.993	-46%	15%	134%
	44.865	39.721	5.123	0.020	F(pa)	0.274	0.232	0.042	0.000	126.286	-26%	-38%	27%

Table 6. Northern Shelf haddock. Basis for forecast settings (using proposed XSA assessment).

Weights-at-age	Linear extrapolation following method of Jaworski (2011)
Fishing mortality	Three-year mean exploitation pattern, scaled to mean F in last historical year (partitioned into catch components using three-year mean proportions at-age)
Recruitment at-age 0 in 2016	Geometric mean of five lowest XSA-estimated recruitments from 1994–2013=1051 million
Recruitment at-age 0 in 2017	Set equal to 2016 value

Table 7. Northern Shelf haddock. MFDP input data (using proposed XSA assessment).

2016							2017						
Age	N	M	Mat	PF	PM	SWt	Age	N	M	Mat	PF	PM	SWt
0	1050561	1.02	0	0	0	0.036	0	1050561	1.020	0	0	0	0.036
1	493996	1.19	0	0	0	0.152	1	.	1.190	0	0	0	0.152
2	846011	0.58	0	0	0	0.425	2	.	0.580	0	0	0	0.425
3	37854	0.36	1	0	0	0.580	3	.	0.360	1	0	0	0.583
4	43144	0.34	1	0	0	0.760	4	.	0.340	1	0	0	0.771
5	7099	0.34	1	0	0	0.958	5	.	0.340	1	0	0	0.944
6	17478	0.25	1	0	0	0.936	6	.	0.250	1	0	0	1.148
7	107307	0.22	1	0	0	0.763	7	.	0.22	1	0	0	1
8	26871	0.2	1	0	0	0.948	8	.	0.2	1	0	0	1
Catch							Catch						
Age	Sel	CWt	DSel	DCWt			Age	Sel	CWt	DSel	DCWt		
0	0.000	0.000	0.004	0.036			0	0.000	0.000	0.004	0.036		

2016						2017				
1	0.003	0.356	0.056	0.141		1	0.003	0.356	0.056	0.141
2	0.177	0.523	0.106	0.267		2	0.177	0.523	0.106	0.267
3	0.263	0.620	0.040	0.369		3	0.263	0.620	0.04	0.348
4	0.282	0.654	0.011	0.479		4	0.282	0.642	0.011	0.482
5	0.247	1.016	0.009	0.497		5	0.247	0.726	0.009	0.587
6	0.139	0.857	0.003	0.551		6	0.139	1.197	0.003	0.588
7	0.077	0.708	0.001	0.491		7	0.077	0.957	0.001	1
8	0.078	0.83	0	0.714		8	0.078	0.87	0	1
IBC						IBC				
Age	Sel	CWt				Age	Sel	CWt		
0	0.000	0				0	0.000	0		
1	0.000	0.3562				1	0.000	0.3562		
2	0.000	0.5228				2	0.000	0.5228		
3	0.000	0.6205				3	0.000	0.6205		
4	0.000	0.6425				4	0.000	0.6425		
5	0.000	0.6619				5	0.000	0.6619		
6	0.000	0.7576				6	0.000	0.7576		
7	0.000	0.7728				7	0	0.7728		
8	0	1.41				8	0	1.41		

2018						
Age	N	M	Mat	PF	PM	SWt

2016							2017
0	1050561	1.02	0	0	0	0.036	
1	.	1.19	0	0	0	0.152	
2	.	0.58	0	0	0	0.425	
3	.	0.36	1	0	0	0.583	
4	.	0.34	1	0	0	0.632	
5	.	0.34	1	0	0	0.963	
6	.	0.25	1	0	0	1	
7	.	0.22	1	0	0	1	
8	.	0.2	1	0	0	1	

Catch					IBC		
Age	Sel	CWt	DSel	DCWt	Age	Sel	CWt
0	0	0	0.004	0.036	0	0	0.000
1	0.003	0.356	0.056	0.141	1	0	0.356
2	0.177	0.523	0.106	0.267	2	0	0.523
3	0.263	0.62	0.04	0.348	3	0	0.621
4	0.282	0.642	0.011	0.372	4	0	0.643
5	0.247	0.662	0.009	0.596	5	0	0.662
6	0.139	0.799	0.003	0.696	6	0	0.7576
7	0.077	1.379	0.001	0.679	7	0	0.7728
8	0.078	0.966	0	0.888	8	0	1.41

Table 8. Northern Shelf haddock. Draft advice table (using proposed XSA assessment).

Rationale	Total catch 2017	Wanted catch 2017	Unwanted catch 2017	IBC 2017	Basis	Total F 2017	F(land) 2017	F(disc) 2017	F(IBC) 2017	SSB 2018	% SSB change	% TAC change	% change from 2015 landings
MSY	65.442	57.996	7.446	0.000	New F(msy) estimate	0.260	0.214	0.046	0.000	282.210	-23%	-10%	85%
Management plan	74.372	65.895	8.477	0.000	MP target F	0.300	0.246	0.054	0.000	273.897	-25%	3%	111%
IBC only	0.000	0.000	0.000	0.000	No HC fishery	0.000	0.000	0.000	0.000	343.791	-6%	-100%	-100%
Other options	56.150	49.771	6.379	0.000	0.75 * F(sq)	0.220	0.181	0.039	0.000	290.876	-20%	-22%	59%
	72.785	64.491	8.294	0.000	Fsq	0.293	0.241	0.052	0.000	275.336	-24%	1%	106%
	88.494	78.379	10.115	0.000	1.25 * F(sq)	0.366	0.301	0.065	0.000	260.785	-28%	22%	151%
	64.311	56.995	7.316	0.000	15% TAC decrease (full)	0.255	0.209	0.046	0.000	283.263	-22%	-15%	82%
	75.253	66.674	8.579	0.000	Rollover TAC (full)	0.304	0.250	0.054	0.000	273.077	-25%	0%	113%
	85.739	75.945	9.795	0.000	15% TAC increase (full)	0.353	0.290	0.063	0.000	263.339	-28%	15%	143%
	78.723	69.743	8.980	0.000	F(pa)	0.320	0.263	0.057	0.000	269.852	-26%	9%	123%

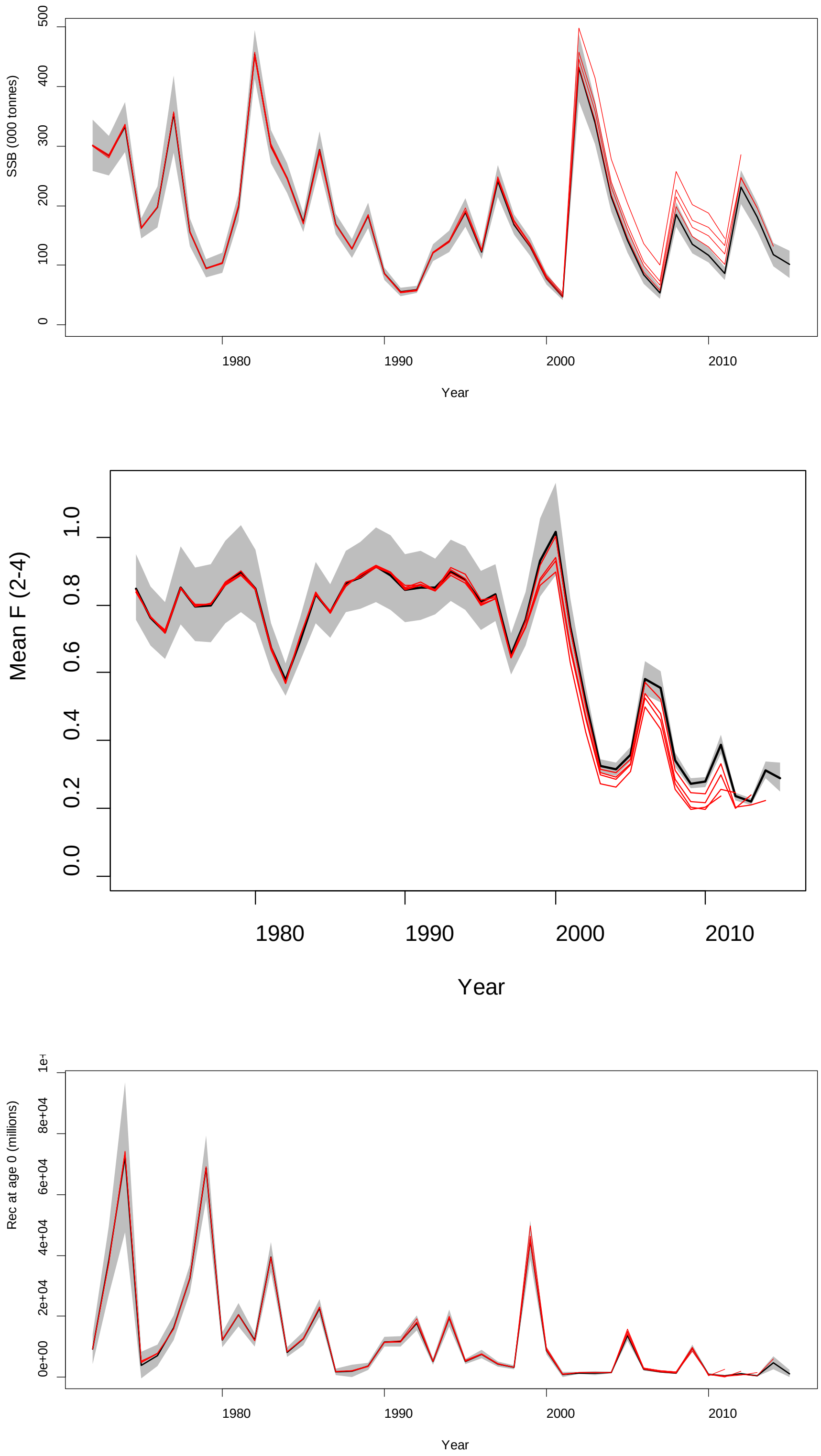

Figure 1. Northern Shelf haddock. Retrospective TSA runs (SSB, mean F(2–4), recruitment). The red line gives the full time-series estimates with the grey band indicating the approximate pointwise 95% confidence interval, while the black lines show the median estimate from retrospective plots.

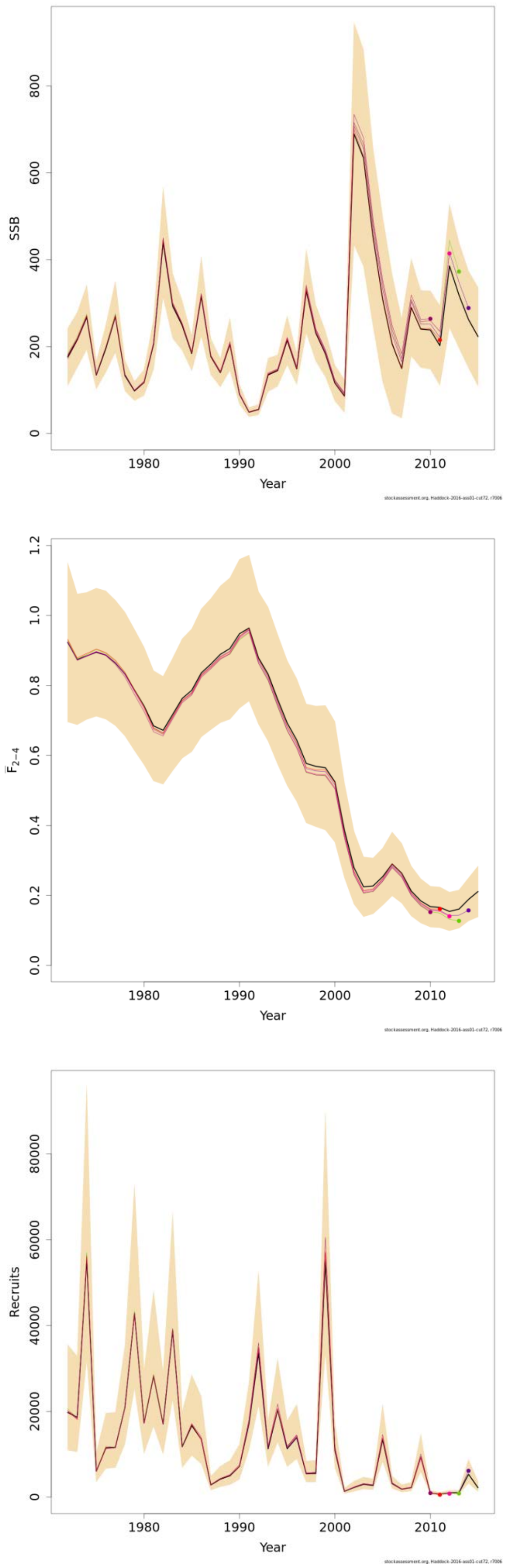

Figure 2. Northern Shelf haddock. Retrospective SAM runs (SSB, mean F(2–4), recruitment). Bands show the approximate pointwise 95% confidence intervals.

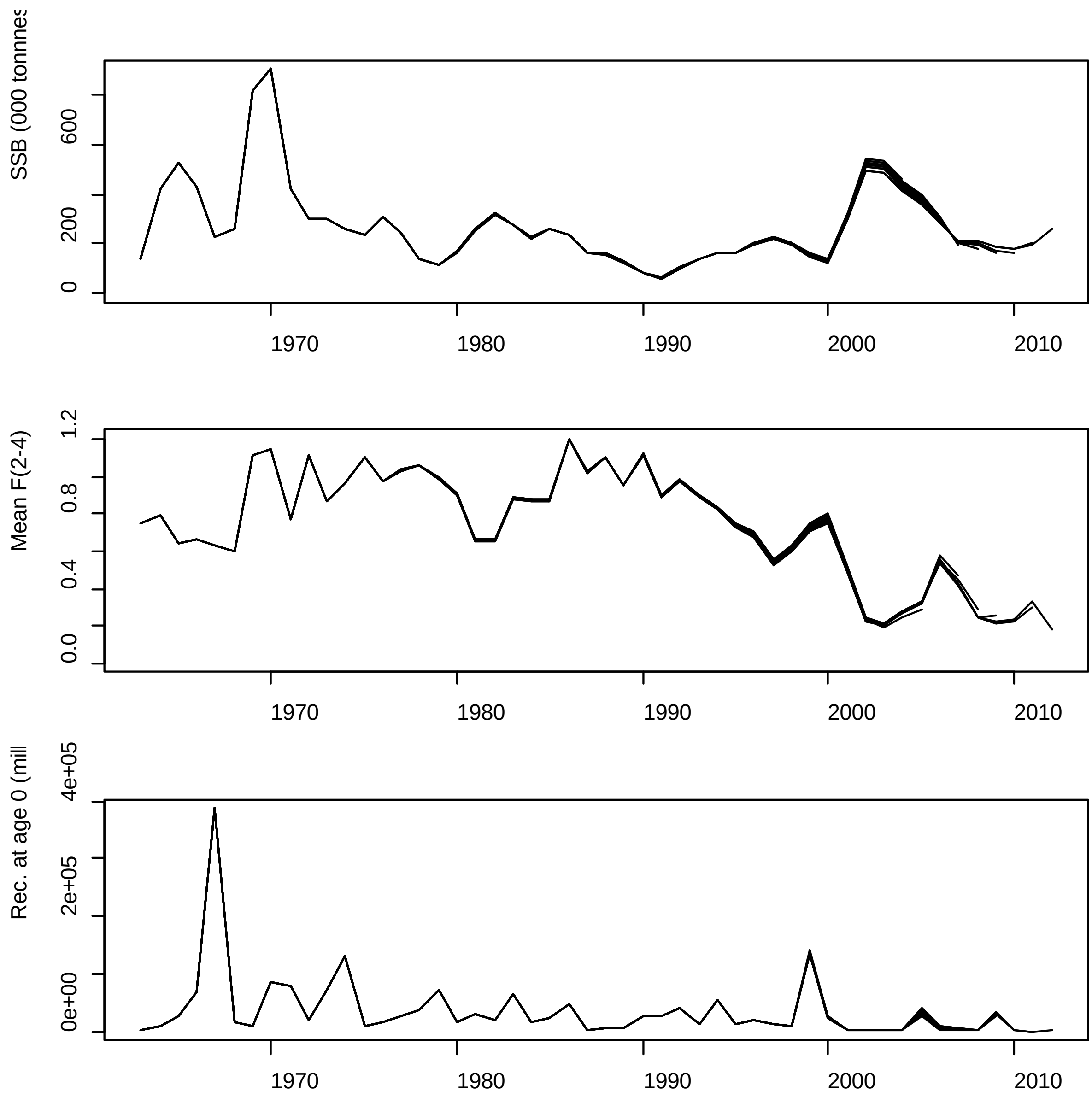

Figure 3. Northern Shelf haddock. Retrospective XSA runs (SSB, mean F(2 –4), recruitment).

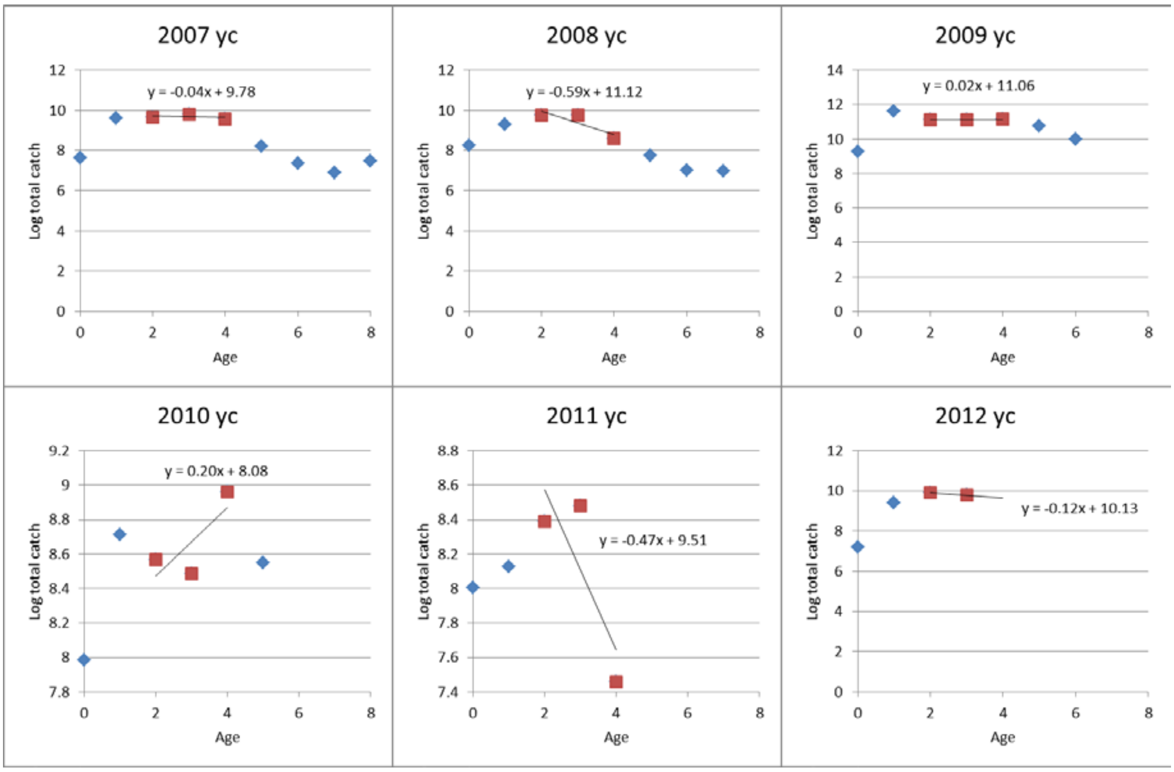

Figure 4. Northern Shelf haddock. Log total catch by age for year classes 2007–2013. Ages 2 –4 for each year class have been highlighted, and a simple linear regression has been fitted through the values for these ages.

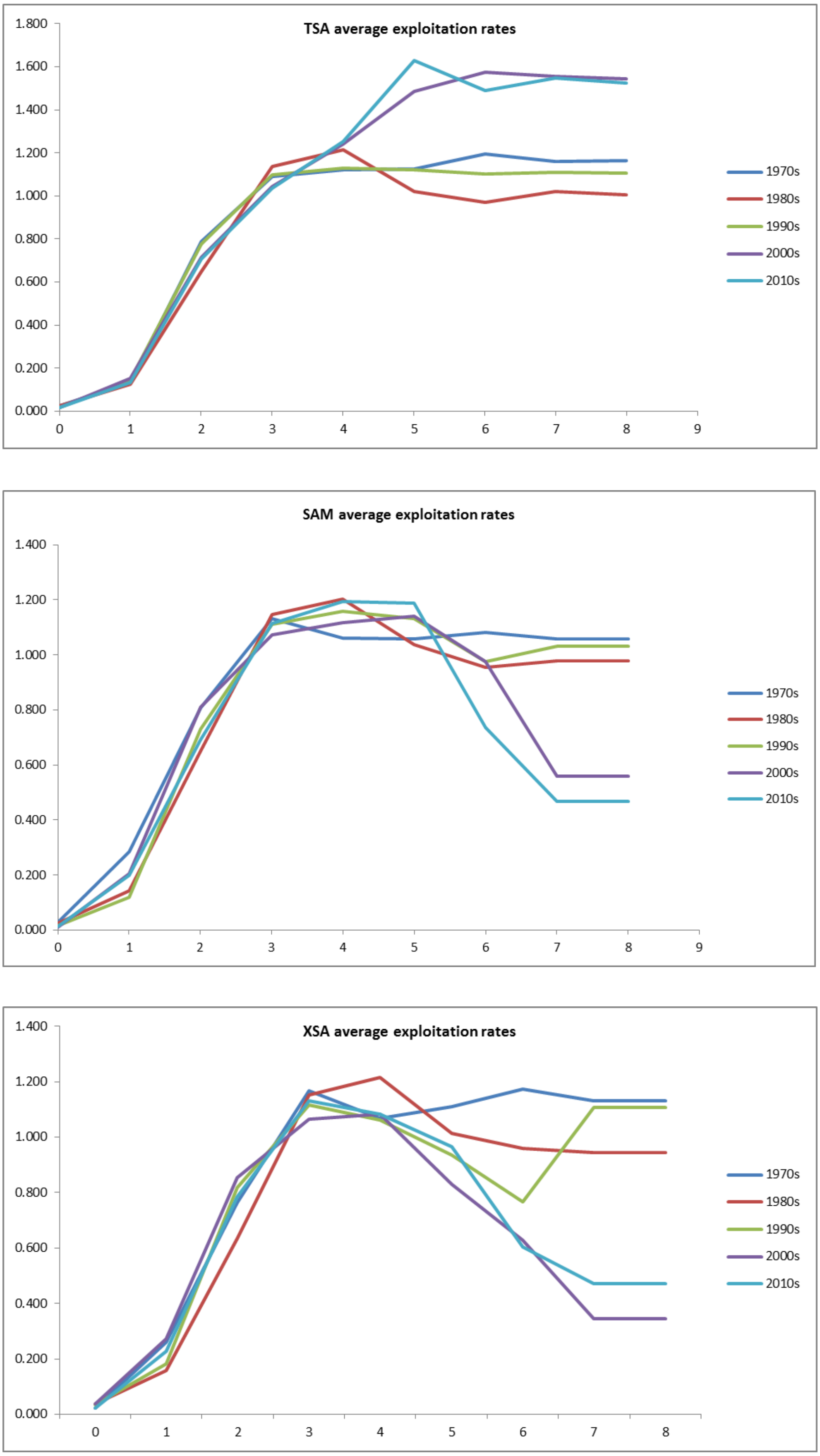

Figure 5. Northern Shelf haddock. Exploitation rates ($F_{a,y}$ / mean $F_{2-4,y}$) for TSA (upper), SAM (middle) and XSA (lower) exploratory assessments. Rates are averaged by decade.

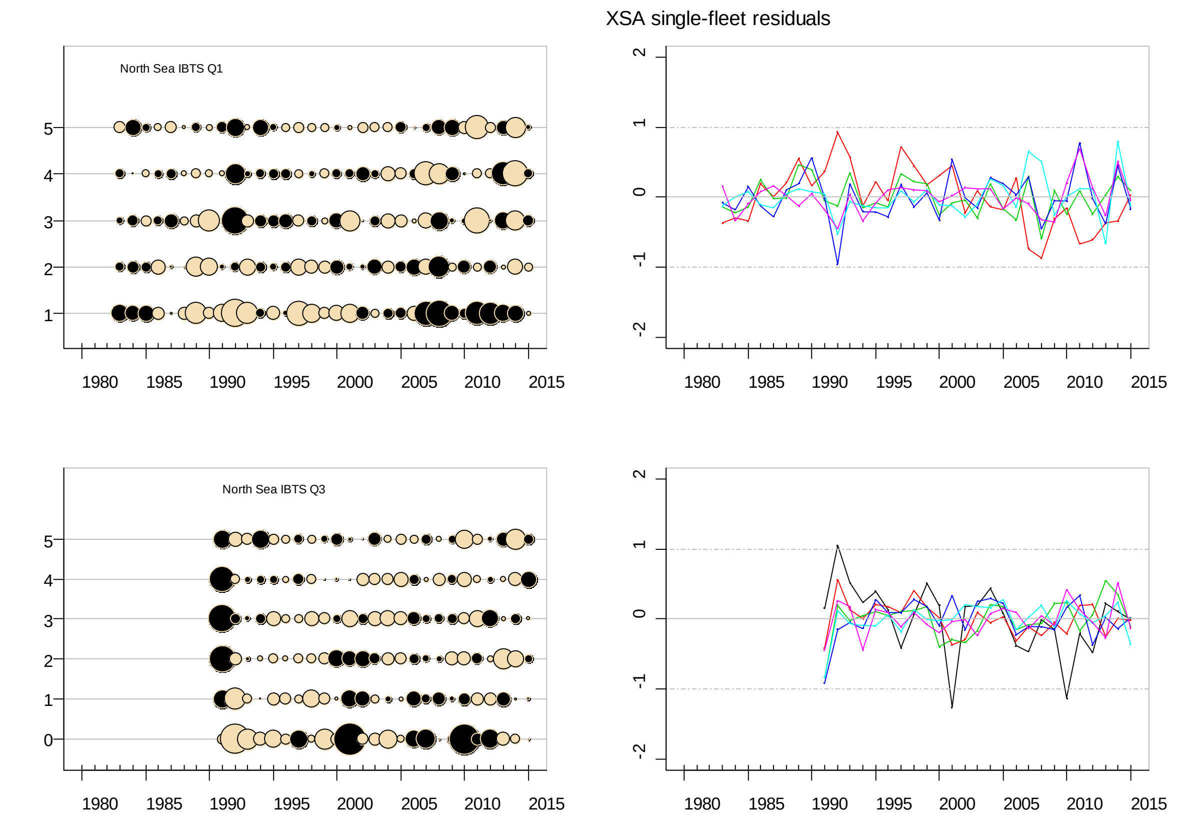

Figure 6. Northern Shelf haddock. Residuals from single-fleet XSA runs.

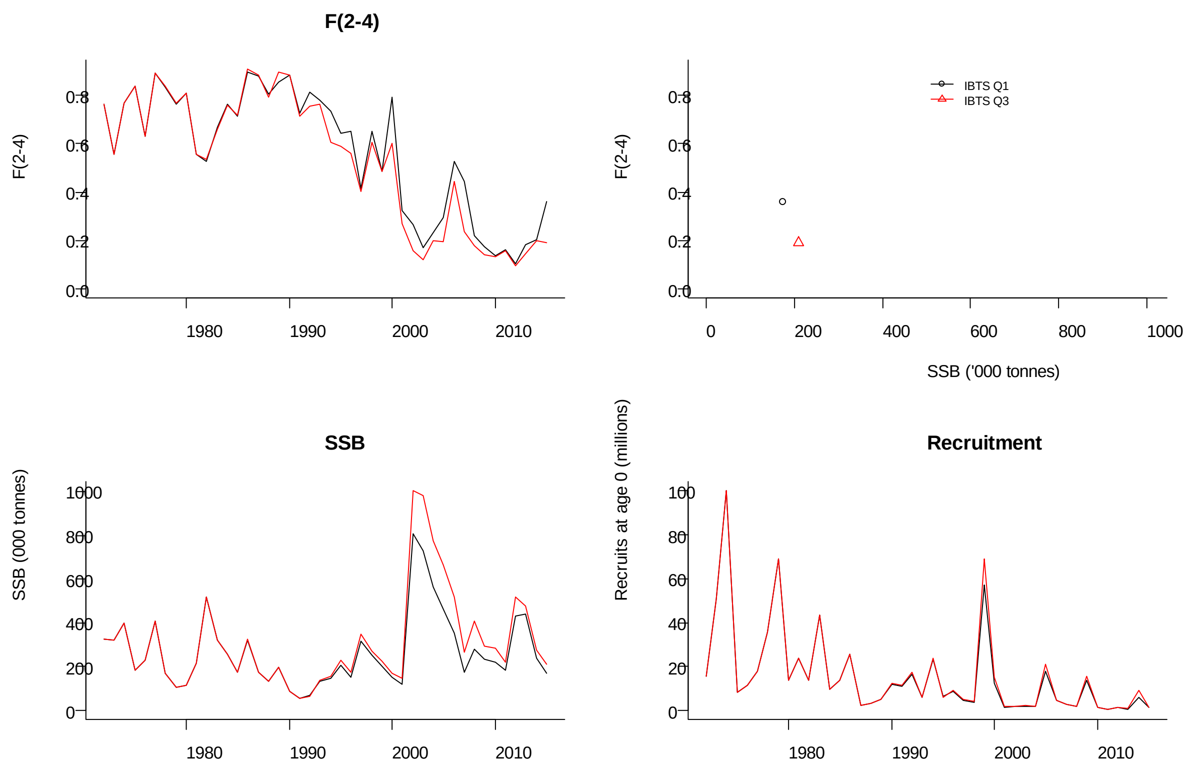

Figure 7. Northern Shelf haddock. Stock summaries from single-fleet XSA runs.

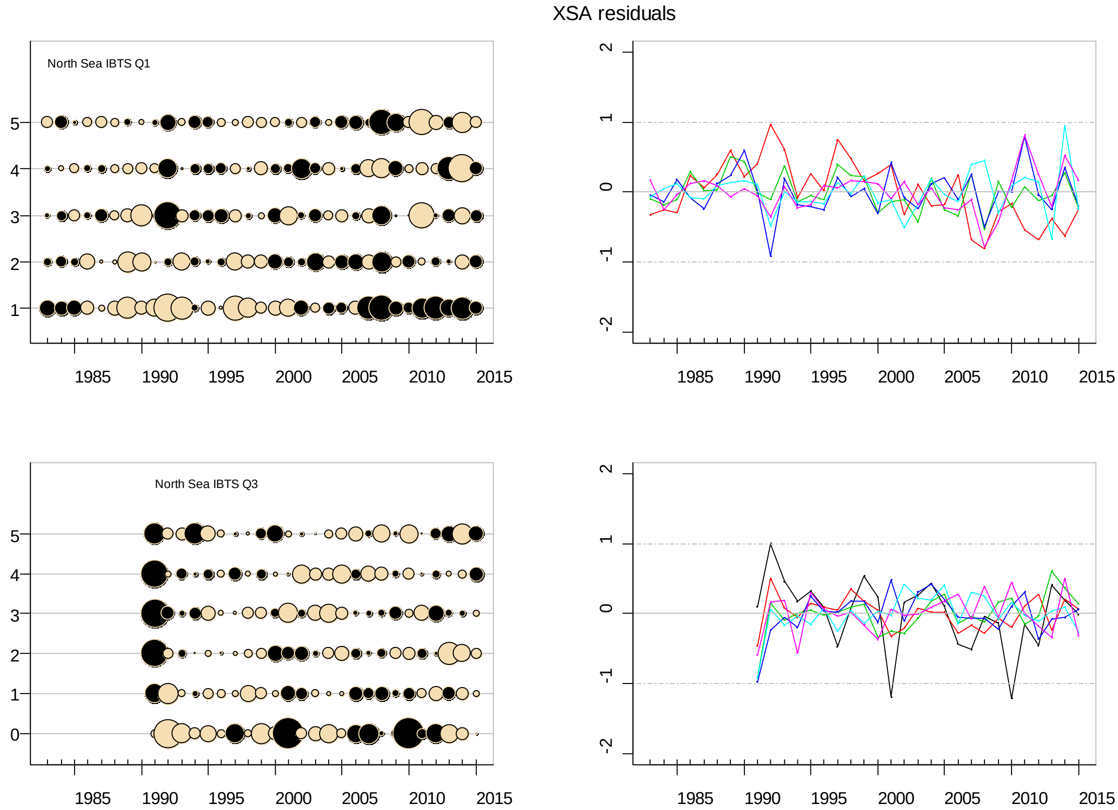

Figure 8. Northern Shelf haddock. Residuals from proposed final XSA run.

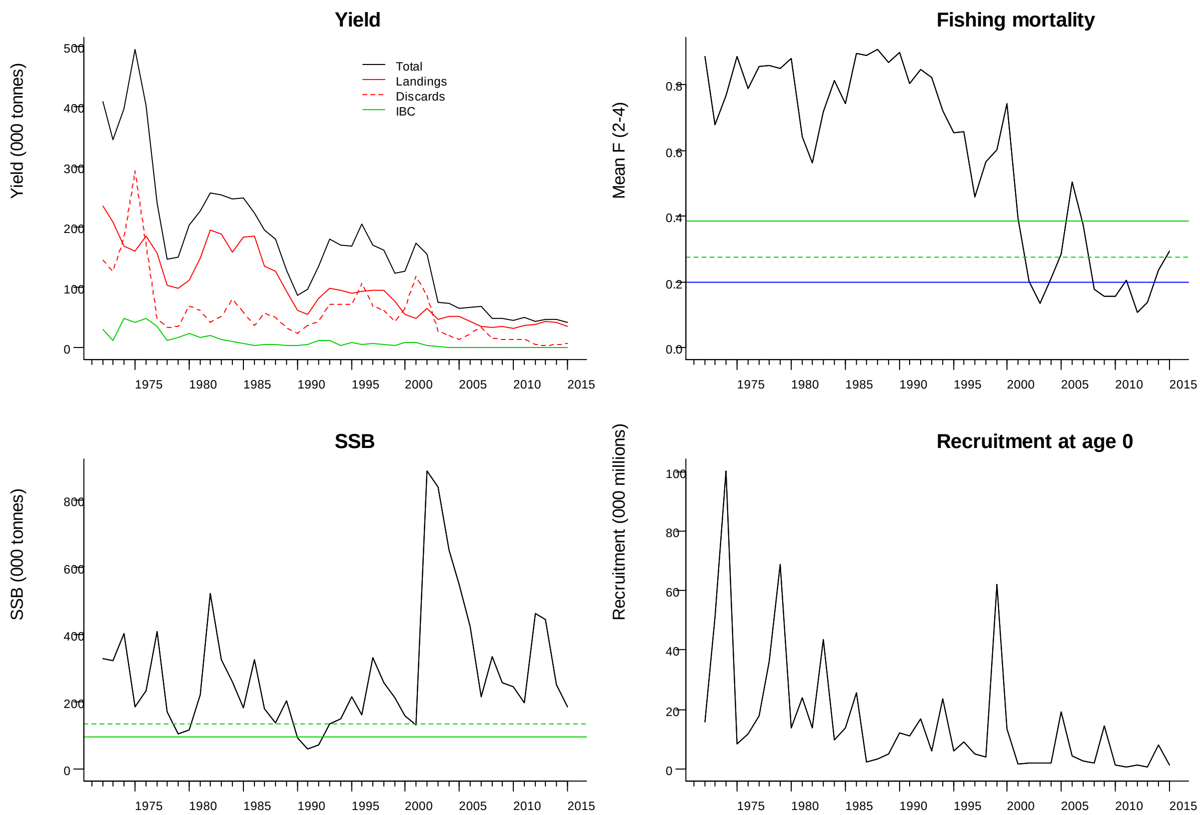

Figure 9. Northern Shelf haddock. Stock summary from proposed final XSA run.

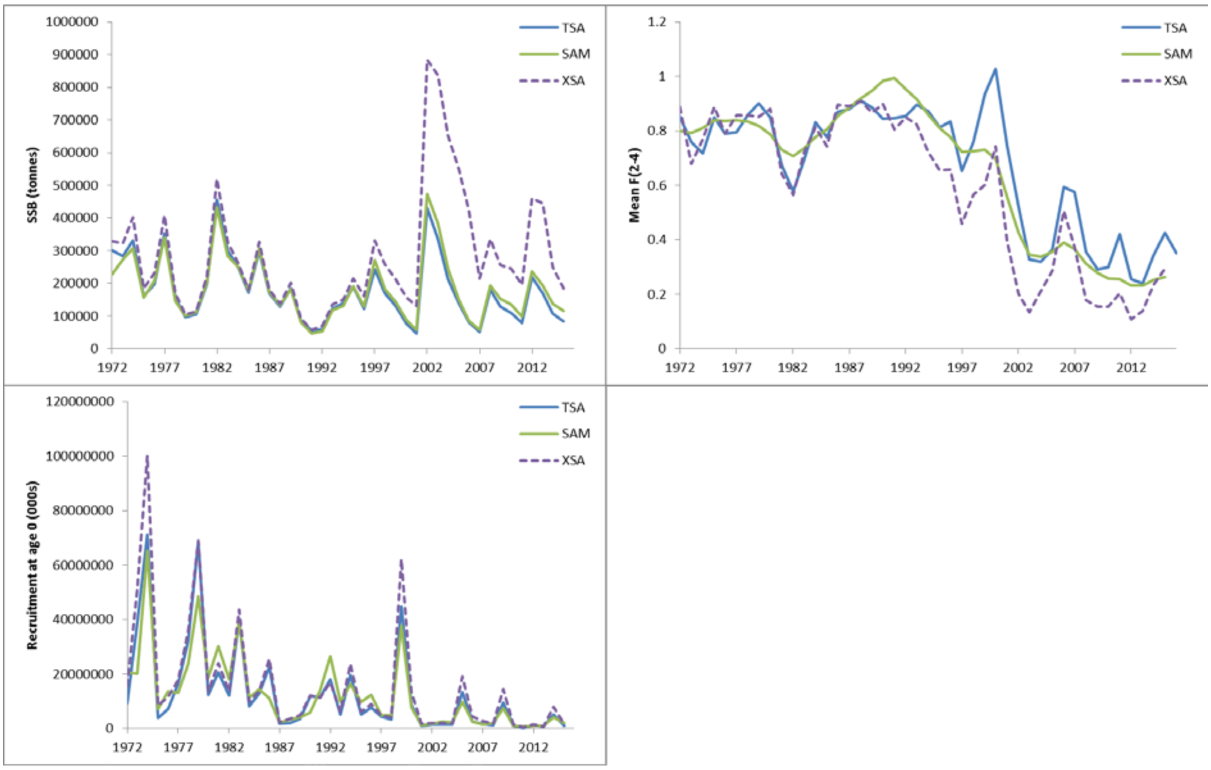

Figure 10. Northern Shelf haddock. Comparison between TSA (WGNSSK 2016), SAM and XSA.

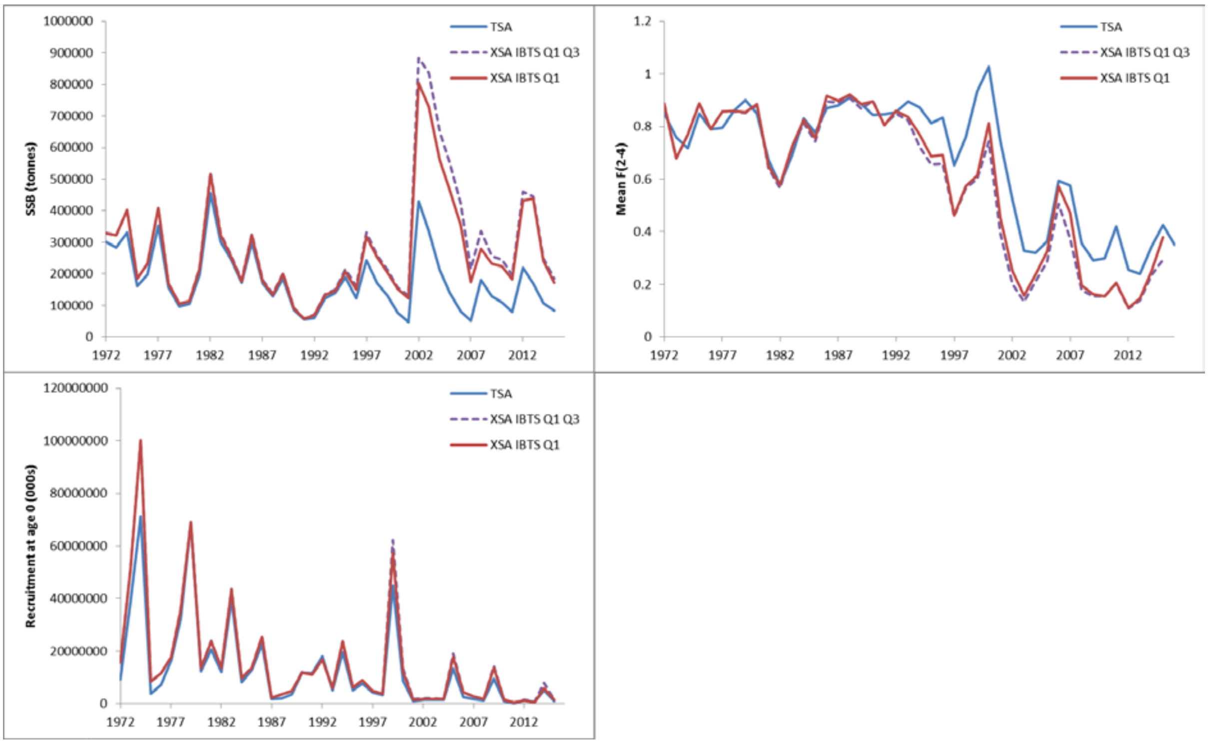

Figure 11. Northern Shelf haddock. Comparison between TSA, XSA (both IBTS Q1 and Q3), and XSA (IBTS Q1 only).

Annex 3: Technical Minutes from RGHaddock

Revised assessment for Northern shelf haddock

WGNSSK 2016

Review by Asgeir Aglen and Bjarte Bogstad, 31 May 2016

The requested review should focus on whether the WGNSSK proposal to reject the benchmarked model is warranted (Working Paper by Coby Needle), and to review the XSA assessment that WGNSSK has proposed.

Concerning the XSA, we have the following comments: Missing tuning data for some of the age groups contributing significantly to catches and SSB may often lead to strange results. For Northern shelf haddock there are tuning data for ages 0–5 and the model runs for 0–8+. The only tuning information available for estimating abundance of ages 6 and older in the last year is therefore derived from the rather uncertain “extended survivors” from observations at younger ages. The corresponding effects on SSB may get particularly large in cases when a strong year class becomes older than age 5. We have seen such issues for NEA haddock and saithe in recent years and those are part of the reason for not using XSA for those stocks anymore.

Another point, which probably explains much of the historical discrepancy between XSA and the other models, is the way XSA handles the plus group (stock number in the plus group is a direct result of catch number in the plus group and the estimated F in “oldest true age”). This may lead to a number of cases where the abundance of the plus group in year y+1 may be inconsistent with the abundance of the oldest true age group in year y, abundance of plus group in year y and fishing mortality on oldest true age group and plus group in year y. For this stock we have really big plus groups in some years!

Examples for strong cohorts (stock numbers and F):

2006: Age 7: 525 863, age 8+: 3866, F age 7 and 8+: 0.072

2007: Age 8+: 118931 – about 22% of age 7+ in previous year, inconsistent with F and M

2012: Age 7: 111 843, age 8+: 7509, F age 8+: 0.082

2013: Age 8+: 147 651, larger than 7+ in previous year!

Thus XSA gives a stock history affected by year-to-year inconsistencies in biomass of older fish, and the resulting SSB is not a good descriptor of the stock history.

Thus we can’t recommend using the XSA assessment presented for this stock.

We also compared 6+ numbers in all models to IBTS Q1 and Q3 numbers for 6+ (which are not used in the tuning), possibly not a relevant comparison, but it is important to identify how much is left of a strong year class at that age. The TSA model gives a decent comparison to the surveys while XSA and SAM show large discrepancies in some years including the last ones, see Figures 1 and 2.

The retro-runs for the three models are presented in separate plots with different scaling, and are, therefore, not easy to compare. TSA isn’t as bad as it may look; The SSB retro is not bad the last 3–4 years, and the recruitment retro seems decent, while for F the TSA has the worst retro among the three models.

Using ages 2–4 as reference ages for F does not seem to be a good measure of exploitation rate in several years when the fishery is dominated by older fish. One should look into using a wider age range, or even better, a harvest rate (Catch/exploitable biomass (which here could be 2+) like the Icelanders do, or alternatively a weighted F like NSS herring? A retro for TSA on those measures of exploitation rate would probably look better than the F2–4 retro.

We would also like to note that the text and tables on forecast results (in Section 13.6) could have been more informative, e.g. the SSB in 2017 is only given in the text and not in any tables. Detailed output on stock composition at the start of 2017 would have been useful (like Tables 4.19 and 4.20 in AFWG 2015). Also in the last line of the text it should be 2017 not 2016 as far as we can see.

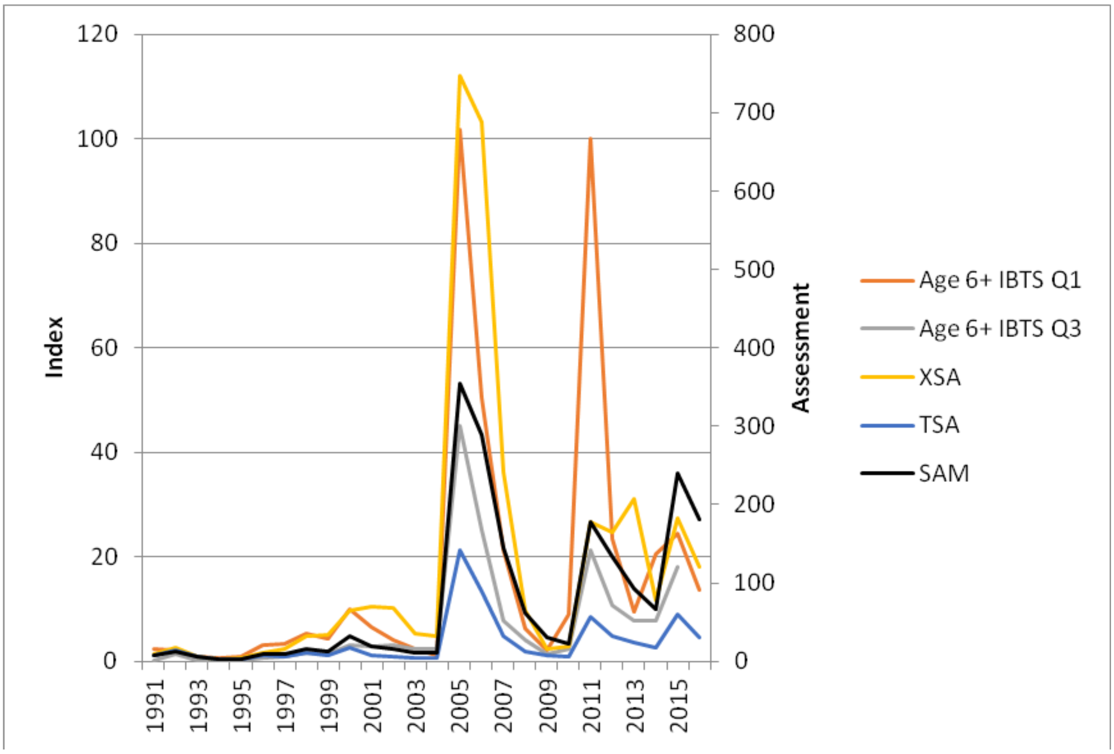

Figure 1. Comparison of age 6+ abundance for IBTS Q1 and Q3 to results from the XSA, TSA and SAM assessment models.

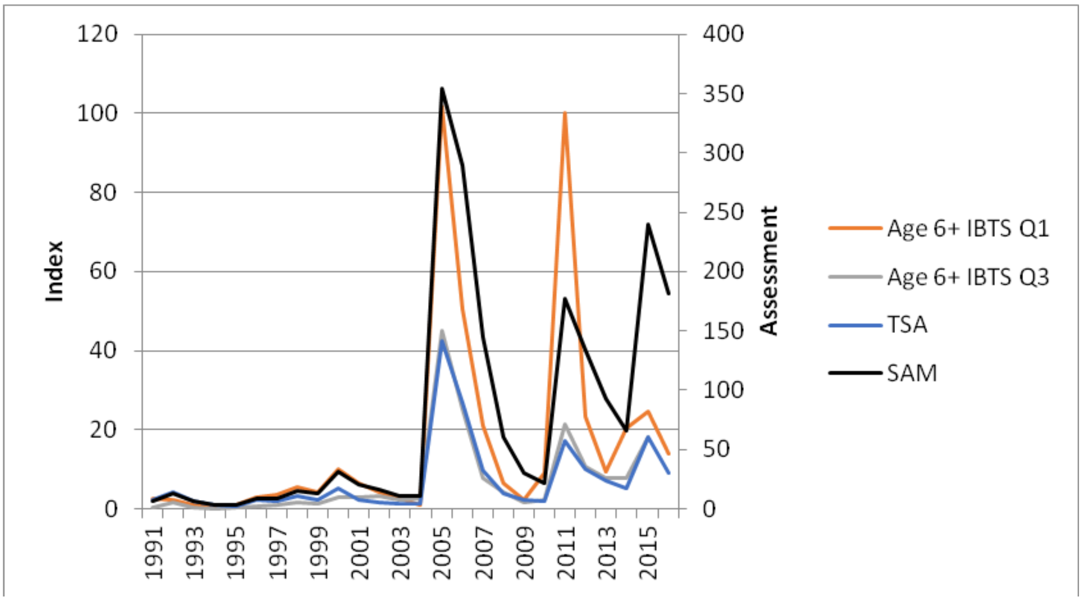

Figure 2. Comparison of age 6+ abundance for IBTS Q1 and Q3 to results from the TSA and SAM assessment models. (Note that right-hand axis is changed compared to Figure 1).

Revision of reference points

The software and calculations seem to be ok, but the choice of time period is not well justified. By excluding data prior to 2000, the recruitment pattern is radically changed to a situation where strong recruitment never occurs. "Spasmodic" strong year classes are typical for haddock stocks. If there is no firm evidence of regime shifts or severely increased mortality at early life stages, we think a longer time-series would be a more realistic basis for estimating reference points.

Also the F_{MSY} value of 0.20 calculated using the short time-series is below all observed values in the period 1972–present, and assumption about growth, maturation etc. may not be valid at stock sizes corresponding to such low fishing mortalities.

Conclusion

We see two options: Either give advice according to the TSA assessment presented by WGNSSK (and old reference points), or postpone the advice.

Annex 4: Stock Annex Haddock in Subarea 4 and Divisions 3.aW and 6.a (North Sea, Skagerrak and West of Scotland)

The table below provides an overview of the updated WGNSSK Stock Annex. 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
had-346a	Haddock in Subarea 4 and Divisions 3.aW and 6.a (North Sea, Skagerrak and West of Scotland)	September 2016	Haddock in 346a