BENCHMARK WORKSHOP ON ROCKALL HADDOCK (*Melanogrammus aeglefinus*) in DIVISION 6.B (ROCKALL) (WKROCK; OUTPUTS FROM 2019 MEETING)

VOLUME 2 | ISSUE 2

ICES SCIENTIFIC REPORTS

RAPPORTS
SCIENTIFIQUES DU CIEM

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

ISSN number: 2618-1371 I © 2020 International Council for the Exploration of the Sea

ICES Scientific Reports

Volume 2 | Issue 2

BENCHMARK WORKSHOP ON ROCKALL HADDOCK (*MELANOGRAMMUS AEGLEFINUS*) IN DIVISION 6.B (ROCKALL) (WKROCK; OUTPUTS FROM 2019 MEETING)

Recommended format for purpose of citation:

ICES. 2020. Benchmark Workshop on Rockall haddock (*Melanogrammus aeglefinus*) in Division 6.b (Rockall) (WKROCK; outputs from 2019 meeting).
ICES Scientific Reports. 2:2. 69 pp. http://doi.org/10.17895/ices.pub.5547

Editors

Helen Dobby • Alexander Kempf

Authors

Michel Bertignac • Helen Dobby • Andrzej Jaworski • Alexander Kempf • Vladimir Khlivnoy • Daniel Ricard

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

Contents

i Executive summary

The Workshop on the Benchmark of Rockall haddock (*Melanogrammus aeglefinus*) (WKROCK), chaired by Helen Dobby (UK) for ICES and External Chair, Alexander Kempf (Germany), took place 2–5 April 2019. Other participants in the meeting included scientists from the UK and the Russian Federation, and external reviewers from Canada and France. This stock had never before been through the benchmarking process and therefore there were a significant number of both data and modelling issues to be addressed.

Given the limited time available and the time spent preparing input data at the meeting, the workshop focused more on a review of the current assessment rather than on advancing the assessment and subjecting it to a complete review of data inputs, modelling assumptions and an examination of the advice consequences of alternate model runs. The main focus of the WK was how to estimate stock weights from the very uncertain (and increasing) mean weight-at-age in the catch and in addition how best to deal with changes in the survey design (of the tuning index) including an expansion to cover a greater portion of the stock distribution. It was agreed that stock weights would be best modelled as a five year running average of catch weights. The WK agreed to maintain the approach adopted by the assessment WG for dealing with the survey data i.e. making use of a single series based on the original survey area, due to a much poorer retrospective when the new survey series was included as a separate index (most likely due to the short time-series).

A further outcome of the workshop was that XSA assessment, which had previously been conducted using the Lowestoft VPA suite, was implemented in FLXSA, facilitating the management strategy evaluation required in the second part of this benchmarking process.

Future research and data requirements were also identified in consultation with the external reviewers.

ii Expert group information

Expert group name	The Workshop on the Benchmark of Rockall Haddock (*Melanogrammus aeglefinus*) (WKROCK)
Expert group cycle	Annual
Year cycle started	2019
Reporting year in cycle	1/1
Chair(s)	Helen Dobby, UK
	Alexander Kempf, Germany
Meeting venue(s) and dates	2–5 April 2019, ICES HQ in Copenhagen, Denmark (six participants)

1 Introduction

A **Benchmark of Rockall haddock (*Melanogrammus aeglefinus*) in Division 6.b (Rockall)** (WKROCK), chaired by Helen Dobby, UK for ICES and External Chair Alexander Kempf, Germany and attended by two invited external experts, Daniel Ricard, Canada and Michel Bertignac, France, met at ICES, Copenhagen, Denmark 2–5 April 2019.

a) Examine data sources, as necessary
b) Update assessment, as necessary
c) Re-examine and update, if appropriate, MSY and PA reference points according to ICES guidelines (see Technical document on reference points); and,
d) Conduct an MSE for use in the ICES advice in 2019.

Stocks	**Stock leader**	**Stock assessor**
Rockall haddock (*Melanogrammus aeglefinus*) in Division 6.b (Rockall)	Vladimir Khlivnoi	Vladimir Khlivnoi

The Benchmark Workshop will report by 5 May 2019 for the attention of ACOM.

2 Description of the benchmark process

Prior to this benchmark, Rockall haddock (Division 6b) had not previously been benchmarked. The input data and stock assessment were agreed at an assessment WG over ten years ago, with a number of enforced *ad hoc* changes being made since then to account for changes in survey design and limited commercial sampling. (The last update to the stock annex prior to this benchmark was made at the assessment WG in 2017.)

In preparation for the benchmark WKROCK, a data call was issued for historical (2009–2018) landings, discards, numbers and weights-at-age, sample information, and effort data, to be imported into InterCatch.

No data evaluation workshop was held ahead of the meeting, as it was considered that there ought to be sufficient time at WKROCK for this to take place (given that it was a single-stock benchmark). However, a substantial amount of time at WKROCK had to be devoted to catch-at-age estimation in InterCatch due to uncertainty regarding the process which had been carried out. As a result, there was limited time at the benchmark meeting for a full evaluation of input data quality or consideration of model results. Explorations conducted as part of the benchmark were therefore limited to the current assessment model. Assessment model runs were conducted following the benchmark meeting and the final assessment settings were agreed by WebEx. The final agreed assessment and additional sensitivity analysis are documented in this report.

Two working document were prepared and presented to the benchmark describing and exploring some of the input data issues:

WD 1: Q3 Rockall Haddock Survey and index evaluation by Andrzej Jaworski

WD 2: Prediction of mean catch weight-at-age and estimation of mean stock weight-at-age for haddock in Division 27.6b (Rockall) by Andrzej Jaworski

The working documents can be found in Annex 2.

This benchmark meeting was part of the process of providing a response to a NEAFC special request for advice on a Rockall haddock long-term management plan. A further meeting to address this request is scheduled for August 2019 where the reference points for the stock will be re-evaluated on the basis of the assessment agreed at this benchmark.

3 Haddock (*Melanogrammus aeglefinus*) in Division 6.b (Rockall)

3.1 Stock ID and substock structure

No results were presented on the stock ID during the Inter-benchmark Protocol.

3.2 Issue list

The issue list (below) has been developed over a number of years and was finalised at the assessment WG in 2018 (ICES, 2018). The main issues identified for addressing at the benchmark included: the use of a revised survey index for 2011 onwards (following redesign of Scottish Q3 survey), uncertainties in historical catch data and documenting the catch estimation process, potential extension of the age range covered in the assessment (increasing plus group age), appropriate stock weights-at-age, and exploring potential alternative assessment methods (which allow for uncertainty in the input data).

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed to be able to do this: are these available / where should these come from?	External expertise needed at benchmark /type of expertise / proposed names
Tuning series	For tuning data the age range 1–6 was used. Plus group: 7+. Last years one strong year class dominate in haddock 6.b stock which appears once in 5–7 years. The determination of the fishing mortality for the strong year class often is uncertain because that year class is included in the plus group. For example in 2012-2013 year class 2005 dominated in SSB and that year class was included in the plus group.	An improved time-series of landings and discards for ages 7 and older is needed for this assessment. It is necessary for separate estimation of fishing mortality of haddock included in the age plus group.	Data for the landings and discards separated for ages 7, 8 and older.	Experts in the age-based assessment and the survey analysis experts (Helen Dobby and Andrzej Jaworski from MSS Aberdeen; Norman Graham and Colm Lordan from MI Galway; Vladimir Khlivnoi from PINRO Murmansk)
	The survey area coverage was reviewed and extended into deeper waters in 2011 and 2012. The indices obtained from the standard survey area used for the assessment on account of the heterogeneity distribution of the haddock in different parts of the bank.	Determine new survey indexes from whole area can be used for assessment or not. How long should be time-series of new indices for inclusion in the assessment. What to do with the surveys, which covered only part of the new whole area after.	Survey data 1991–2015	

Issue	**Problem/Aim**	**Work needed / possible direction of solution**	**Data needed to be able to do this: are these available / where should these come from?**	**External expertise needed at bench-mark /type of ex-pertise / proposed names**
Discards	A main uncertainty in the assessment concerns the estimates of discards in the EU fleets. Current discards estimates are based principally on Marine Scotland and hardly ever the Irish sampling trips. In years when the discards trips were not conducted the discards was calculated using fishery selectivity and discarding ogives by the theoretical method developed by V. Khhlivnoi. The last years the number of the discards samples and discards trips is very small what can cause leads to a distortion of information on the size and age composition of the discards.	It is necessary to organ-ize the collection of dis-cards data on fishing ves-sels in the amount nec-essary for the assess-ment. It possible through an extended discards tripe programme. It is necessary to analyse that in the case of a small amount of data will better correspond to the real discards data ob-tained by small number of samples or the results obtained from the theo-retical model proposed by V. Khlivnoi. It is advisable to do the analysis and determine for which other stocks (with the discards) the theoretical model can be applied with the aim of improving the quality of assessment.	Length distribu-tion and age composition for the landings, the discards and the survey.	Experts in the dis-cards analysis from MSS Aberdeen and MI Galway. Experts in the theo-retical modelling of discards (e.g. Vladi-mir Khlivnoi, PINRO Murmansk).
Biological Pa-rameters	There are doubts on the de-gree of age-reading agree-ment by international ex-perts. Results of age-reading of the identical otoliths differ.	it would be beneficial to develop and introduce standardization methods for reading the age for haddock.	Haddock otoliths which were col-lected at the Rockall.	The age-reading ex-perts from MSS Ab-erdeen, MI Galway and PINRO Mur-mansk.
	The mean weights-at-age in the stock are assumed to be the same as the catch weights.	Recalculate new the mean weights-at-age in the stock. Make an anal-ysis of the influence of new stock weights-at-age data on the results of assessment.	Data for this are the same as for the XSA assess-ment and the weights-at-age in the survey 1991–2015.	Experts in the age-based assessment and the survey anal-ysis experts (Helen Dobby and Andrzej Jaworski from MSS Aberdeen; Norman Graham and Colom Lordan from MI Gal-way; Vladimir Khlivnoi from PINRO Murmansk)

Issue	**Problem/Aim**	**Work needed / possible direction of solution**	**Data needed to be able to do this: are these available / where should these come from?**	**External expertise needed at benchmark /type of expertise / proposed names**
Assessment method	The current two models were used: the XSA and StatCam (model of Statistical catch-at-age analysis). The XSA is based model. Both Statistical catch-at-age analysis and VPA results show a similar tendency for the SSB dynamics but the absolute value of the stock biomass is slightly different.	Make an analysis of whether the use of StatCam to improve the quality of the assessment.	Data for this are the same as for the XSA assessment.	Experts in stock assessment for the models mentioned: Coby Needle, Rob Fryer from MSS Aberdeen; Colm Lordan from MI Galway; Vladimir Khlivnoi from PINRO Murmansk) Experts in Experts in StatCam model e.g. J. Brodziak from Northeast Fisheries Science Center USA
Biological Reference Points	Curent Blim = Bloss=6000 t is defined as the lowest observed spawning stock estimated in previous assessments. Since 2014 SSB is below the Blim. This stock is characterized by the following features, which affect the time value of the reference points (Fmsy, Bmsy and other): 1) No significant relationship between spawning biomass and the recruitment. 2) Stock state does not depend on the value of the SSB, and is determined by the number of recruitment. 3) Yield and biomass have increased when the level of fishing mortality was more than 0.6 in periods with high recruitment. 4) Currently recruitment is low and the stock decline below the Blim at the decreasing of Fbar to 0.2. 5) Striping long time periods with high and low recruitment, which is typical for that stock which have a significant impact on the change in f the reference points.	Generation of new reference points after a final assessment configuration is agreed, potentially through risk-based simulation. Determine the reference points for current stock state with low recruitment is needed. Suggest the reference points at high recruitment time period. Define the criteria for the identification of period with a low and with a high level of recruitment. The calculations should be made, including by a model which takes into account absence relationship between SSB and the recruitment.	Data for this are the same as for the assessment itself.	Experts in simulation risk analysis (e.g. Coby Needle from MSS Aberdeen, V. Khlivnoi from PINRO Murmansk PINRO, Chris Darby from Cefas).

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed to be able to do this: are these available / where should these come from?	External expertise needed at benchmark /type of expertise / proposed names
	Reference points will need to be revised taking into account the features of the stock.			

3.3 Scorecard on data quality

A scorecard was not used for this benchmark process.

3.4 Multispecies and mixed fisheries issues

No new information was presented as part of this benchmark process.

3.5 Ecosystem drivers

No new information was presented as part of this benchmark process.

3.6 Stock Assessment

3.6.1 Catch–quality, misreporting, discards

Given the nature of this fishery (distant waters/few vessels), the commercial data suffer from very low sampling levels, both for landings and discards and historically, missing discards have been imputed on the basis of survey data and estimated selectivity/discard ogives. There is also anecdotal evidence that in the past misreporting of Rockall haddock landings may have occurred which may affect the quality of the landings data (although the extent of misreporting is unknown).

Ages in historical commercial data were separated and included in the plus group (7+) to allow cohorts to be tracked for a greater number of years in the assessment. Also, 2010 discard data was re-estimated but due to time constrains it was not possible for the benchmark to look at that and that task was deferred to WGCSE.

Following the data call issued as part of this benchmarking process, landings and sample data were submitted to InterCatch. In recent years over 80% of landings are taken by UK (Scotland), with smaller proportions taken by Ireland, the Russian Federation and Norway (in descending order). Landings weights were submitted by all required countries. Sampled landings and discards age compositions were provided by fleet for UK (Scotland) and Ireland and landings age compositions provided by the Russian Federation.

For 2012 onwards, the catch-at-age data were re-estimated in InterCatch. The main fleets (UK(Sco) OTB_DEF_>=120 and Irish OTB_DEF_100-119) are typically sampled for both landings and discards. Discard rate allocation to other unsampled fleets consisted of:

i. Manually matching annual discards to available quarterly landings by country/fleet (where necessary);

ii. Using a weighted average discard rate for all unsampled fleets (weighted by CATON) with the exception of the Norwegian longline fleet and the Russian fleet for which discards are both assumed to be zero.

Landings age compositions were allocated to unsampled fleets using a weighted average of all sampled fleets (excluding the Russian fleet which is likely to be less applicable given they do not discard). The weighting algorithm used is 'Mean weight weighted by numbers-at-age or length'. Discards age compositions were allocated in a similar manner.

The resulting age compositions and mean weights-at-age show only minor differences to those compiled at previous assessment WG meetings.

3.6.2 Surveys

Only a single trawl survey covers this area, conducted by Scotland in Q3. The survey began in the mid-1980s and in general has been conducted on an annual basis with the exception of a period during the late 1990s/early 2000s when the survey was carried out every two years (and a vessel breakdown in 2010).

In 2011, there was a major change in survey design, with a move to a random stratified design (covering a larger spatial distribution than earlier years) along with a modification to the gear deployed on the survey. The increase in surveyed area followed concerns raised from other surveys undertaken on the Rockall Bank (for anglerfish) that a significant portion of the Rockall haddock stock could be found in depths out with the range of the survey. At the same time, the ground gear of the survey was altered (from type "C") to one that was more suited to the rough ground encountered in this area (type "D"). WD 1 provides elaboration on the survey design for 2011 onwards and further details can also be found in ICES (2017).

Studies conducted in 2006 and 2009, comparing the catchability of the two ground gears (WD 5 in ICES, 2012a) suggested no significant differences in the catch rate of haddock. The assessment WG in 2012 (ICES, 2012b) proposed an approach whereby only the subset of survey stations occurring within the depth range of the original (pre-2011) survey were included in the post 2011 index calculation. This allowed the survey to be treated as a continuous time-series in the assessment (known as the 'standard index'). The WG at this time also recommended that new survey indexes would be used for the assessment once the time-series for the whole area of haddock distribution is of sufficient length. Note that data from the early part of the time-series (1985–1990) were previously found to have poor consistency, possibly due to the survey being more exploratory in nature, and have not been considered since (ICES, 2004).

WD 1 provides the survey index calculated according to the random stratified survey design (full haddock distribution area) for 2011 onwards and conducts some exploratory analysis of the index which suggests good internal consistency (based on log catch curves and survey scatterplots). It was therefore considered that the new survey index (with greater coverage of the stock distribution) from 2011 onwards should be considered as a potential tuning index in the assessment either in addition to a shortened 'standard index' (pre-2011) or as a single index.

During the benchmark process an error in the ALK used in the calculation of the 2015 survey index was identified and corrected. The corrected index is given in Table 1. The correction results in small differences to the estimates at age 4 and 5 in 2015 (S1). Survey log catch curves and scatterplots can be found in the Supplementary Figures in Annex 3.

A number of additional approaches to calculating the survey index based on bathymetric strata and swept-area (rather than number per hour) were also proposed and although WKROCK had limited time to discuss the merits of these approaches they are included here for completeness. Three alternative stratifications were proposed:

1. by geographic strata of 15' latitude wide and 15' longitude long (S2);
2. by five bathymetric strata depending on depth: <150 m, 150–175 m, 176–200 m, 201–225 m and >225 м (S3)
3. the whole survey area is taken for one strata without substratification (S4).

In each case, the number of individuals within the survey area was determined according to expression 1.

$$\bar{N}_{st} = \sum_{i=1}^{L} S_i \bullet \bar{y}_i \quad , \qquad (1)$$

where L – number of strata, Si – area covered by survey within i-stratum, $\bar{y}_i$ - average density of individuals distribution in i-stratum which was calculated according to (2):

$$\bar{y}i = \frac{\bar{n}}{s} \quad , \qquad (2)$$

where $\bar{n}$ – average abundance of fish determined by haulings in i-stratum, s–covered by one hauling.

To account for annual changes in the survey area abundance indices were recalculated per mile². Total abundance trends and abundance per mile² are broadly similar across these three indices although confidence intervals are substantially greater for indices S3 and S4. The internal consistency also differs little across indices or compared to the 'standard' index. (See Annex 3 for further details).

Table 1. Survey index from the Scottish groundfish survey. 'Standard index' in number per ten hours using corrected 2015 ALK (Index S1).

SCOGFS
1991 2017
1 1 0.66 0.75
0 6

1	14458	16398	4431	683	315	228	37	64	3
1	20336	44912	14631	3150	647	127	200	4	32
1	15220	37959	15689	3716	1104	183	38	73	21
1	23474	13287	11399	4314	969	203	30	12	4
1	16923	16971	6648	5993	1935	483	200	16	-1
1	33578	19420	5903	1940	1317	325	69	6	1
1	28897	10693	2384	538	292	281	71	9	1
1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1	10178	9969	2410	708	279	172	90	64	32
1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1	31813	7455	521	284	154	39	14	12	14
1	11704	20925	2464	173	105	65	20	10	15
1	2526	10114	10927	1656	138	97	100	26	6
1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1	24452	4082	920	1506	2107	231	33	13	7
1	3570	18715	2562	256	1402	1694	349	16	6
1	558	2671	6019	570	254	516	367	28	2
1	85	560	966	3813	182	41	282	249	49
1	132	139	323	488	1651	40	9	54	17
1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1	13	17	96	22	42	88	607	4	4
1	39619	4	12	73	14	75	50	635	9
1	6035	14179	5	8	8	9	11	23	166
1	3044	7232	4692	5	0	13	0	11	10
1	1997	2908	5635	3357	0	0	16	2	20
1	67096	1576	1483	2064	1526	11	1	5	2
1	30130	29449	956	909	1389	663	5	1	2

3.6.3 Weights, maturities, growth, natural mortality

Weights-at-age in the catch and stock

The catch mean weights-at-age are shown in Figure 1. Since 2010, there appears to be an increasing trend in the mean weight-at-age in the catch (particularly for older ages) and also estimates have become more variable from year to year, with a particularly noticeable outlier in 2016 for ages five and above. The older age classes during this period are very weak year classes and hence density-dependent factors may have resulted in faster growth of these cohorts (accounting for the increasing trend in mean weights). Furthermore, the small size of these cohorts means that the number of fish sampled at these ages is likely to be extremely low resulting in increased variability in the estimates. (Further investigation of the sampling data indicates that only two individuals at age 5 and only one at age 6 were sampled).

In the past, stock weights-at-age have been assumed to be equal to the raw catch weights-at-age. However, the 2016 outlier is likely to have a significant impact on the estimates of stock biomass in the assessment and also consequences for the short-term forecast (if recent mean weights are used) in the coming years. Stock weights-at-age as derived from the survey data are shown in Figure 2. Although there is some evidence of an increase at ages 4–6 over the period 2011–2016, both the trend and variability in these data are much less pronounced than in the raw catch weights. For this reason, it was agreed that smoothed catch weights-at-age, using a five year moving average, should be used as stock weights.

Consideration was also given to the assumptions regarding the mean weight-at-age for use in the short-term forecast (WD 2). A number of different approaches (including taking a mean of recent years and using fitted linear growth models based on weight increment) were considered in a retrospective analysis, comparing predicted yield to actual yield over time (See WD 2 and Jaworski, 2011 for further methodological details). Using a three or five year average or a linear model for year class resulted in a relatively precise estimate of the true catch with some bias (underestimation). Over the full time-series (1991–2017) the bias is relatively low, but becomes more significant when the recent data are included and the results are likely to be particularly impacted by the 2016 outlier in the catch data. Other options based on growth models incorporating combinations of weight, age and year appeared unbiased, but were more imprecise. The benchmark concluded that in general a three year average was likely to be the most appropriate assumption in the long run but given the large outlier in 2016, a five year average should be used in order to reduce the influence of this datapoint.

Maturity and natural mortality

Changes to the maturity ogive and natural mortality were not considered as part of this benchmark process. Maturity is considered to be knife-edge at age 3:

Age	1	2	3	4	5	6	7+
Proportion mature	0	0	1	1	1	1	1

Natural mortality is assumed to be 0.2 at all ages in the assessment.

Table 2. Catch weights-at-age (kg).

YEAR	AGE					
	1	2	3	4	5	6	7
1991	0.142	0.240	0.291	0.378	0.469	0.414	0.681
1992	0.133	0.239	0.318	0.362	0.423	0.567	0.852
1993	0.137	0.238	0.335	0.400	0.493	0.503	0.882
1994	0.153	0.233	0.319	0.420	0.469	0.477	0.740
1995	0.118	0.222	0.309	0.401	0.501	0.460	0.870
1996	0.136	0.278	0.314	0.396	0.553	0.575	0.762
1997	0.136	0.240	0.322	0.381	0.512	0.634	0.940
1998	0.141	0.250	0.308	0.354	0.436	0.546	0.663
1999	0.138	0.208	0.272	0.334	0.379	0.483	0.619
2000	0.189	0.250	0.267	0.321	0.382	0.451	0.709
2001	0.133	0.264	0.326	0.447	0.427	0.520	0.683
2002	0.135	0.239	0.237	0.325	0.509	0.579	0.755
2003	0.153	0.203	0.256	0.349	0.384	0.424	0.604
2004	0.147	0.198	0.244	0.294	0.444	0.609	0.753
2005	0.114	0.197	0.235	0.311	0.459	0.600	1.062
2006	0.093	0.198	0.245	0.329	0.441	0.595	0.787
2007	0.114	0.186	0.265	0.294	0.386	0.496	0.578
2008	0.199	0.241	0.291	0.437	0.571	0.669	0.937
2009	0.248	0.288	0.339	0.391	0.668	0.513	1.012
2010	0.141	0.247	0.333	0.327	0.590	0.977	1.464
2011	0.198	0.280	0.596	0.449	0.695	0.603	0.748
2012	0.263	0.295	0.622	0.784	0.372	1.411	1.219
2013	0.211	0.368	0.236	0.704	0.423	0.827	1.261
2014	0.140	0.286	0.268	0.545	1.000	1.036	1.370
2015	0.104	0.254	0.601	0.354	1.178	0.948	1.439
2016	0.298	0.449	0.600	0.711	1.556	1.808	2.650
2017	0.219	0.430	0.586	0.691	0.944	0.780	1.270

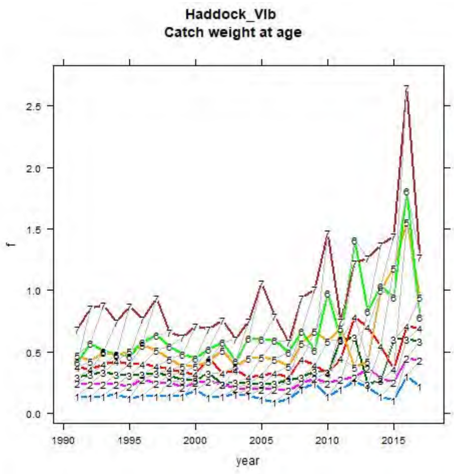


Figure 1. Haddock in Division 6.b. Mean weight-at-age in the catch over time for plus group at age 7.

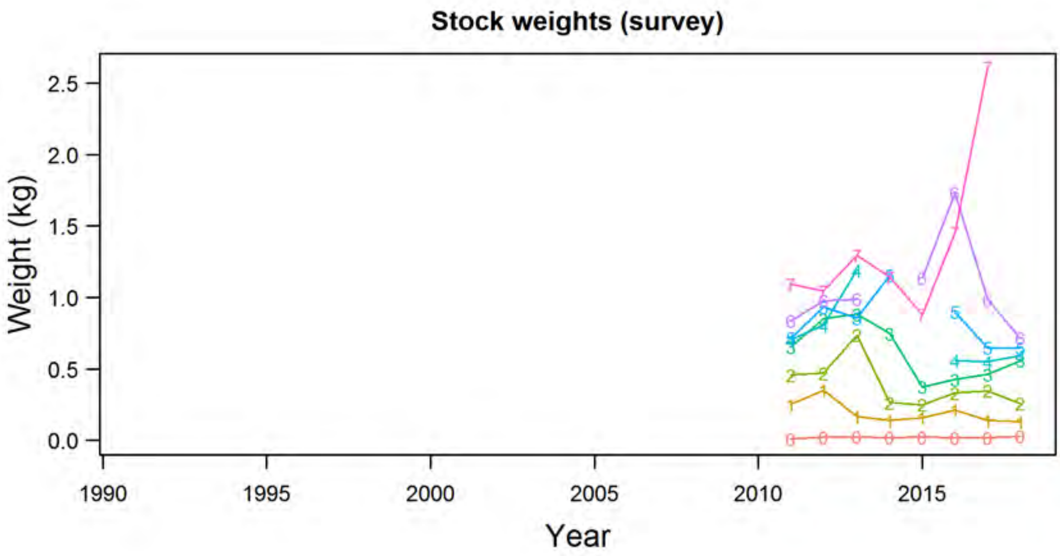


Figure 2. Haddock in Division 6.b. Mean weight-at-age in the catch over time for plus group at age 8.

3.6.4 Assessment Model

The model used to assess Rockall haddock is an extended survival analysis (XSA). No new assessment models were tested during this benchmark. Initial model runs were conducted using the Lowestoft suite of VPA programs while the final assessment was conducted using FLXSA.

Due to the limited time available, the XSA assessment settings were not reconsidered (with the exception of those described below) and are given in Table 3. All model runs (with the exception of those shown in Figure 3) make use of the catch-at-age data re-estimated as part of this benchmarking process (from the data call InterCatch submissions).

Alternative Survey Indices

First, a number of different runs were conducted to explore the sensitivity of the assessment results and diagnostics to different survey indices.

Run S1

The first run to be conducted used the 'standard' survey index with the corrected 2015 ALK with other model settings unchanged from WGCSE 2018 and compared model outputs to the baseline run using the 'standard' index from WGCSE 2018.

The correction to the survey data has little impact on the results of the stock assessment (Figure 3 shows a comparison between the 2018 assessment and the model run with corrected ALK but 2018 WG catch data). In terms of diagnostics (Figure 4 and 5), the fit to the 2015 survey data at age 5 shows significant improvement in Run S1 compared to the baseline: in the baseline model run there is a large positive residual which is no longer apparent (See Supplementary Figures, Annex 3). There is some evidence of blocks of positive residuals (early 1990s) followed by blocks of negative residuals (late 1990s) and negative residuals for the 2011 cohort (recruiting in 2012) over time. However, these are all of relatively low magnitude. Figure 4 also shows that the residuals at age 6 are lower than those at other ages indicating that the model fits the survey index better for this age class.

The retrospective plots (Figure 5) show that there is some uncertainty in the estimates for the final year, but there does not seem to be persistent under or over estimation of either SSB or F.

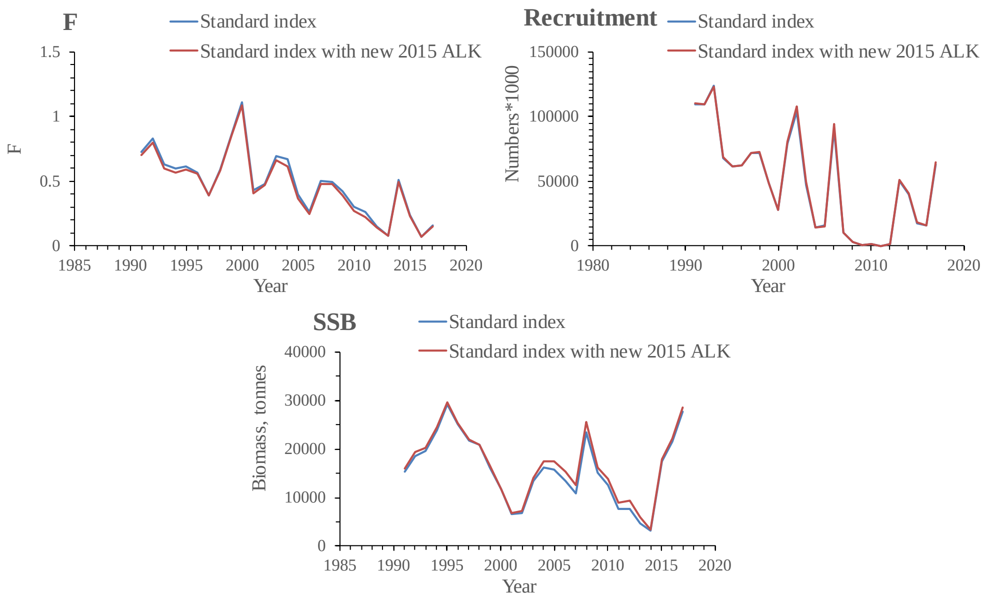


Figure 3. Haddock in Division 6.b. Comparison of XSA results. Index S1: survey Index (N/10 hours) based on 'standard' unstratified area with correct ALK for ages 4–5 in 2015. (2018 WG assessment compared to run using the same catch data with correct ALK).

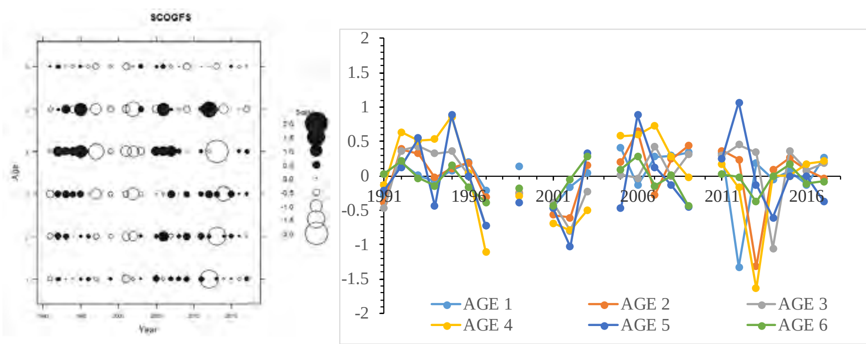


Figure 4. Haddock in Division 6.b. Run S1. Log catchability residual plots. Age plus group 7+. Index S1: survey Index (N/10 hours) based on 'standard' unstratified area with correct ALK for ages 4–5 in 2015.

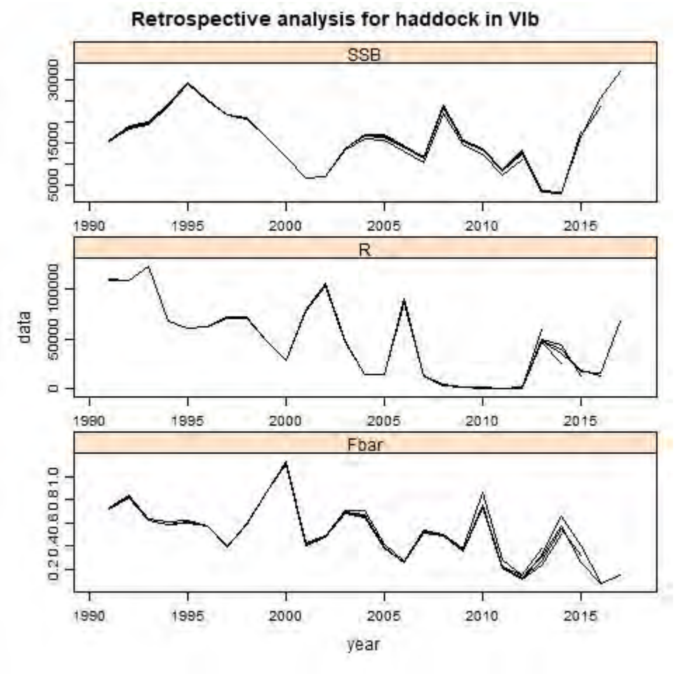


Figure 5. Haddock in Division 6.b. Run S1. Retrospective analyses. Age plus group 7+. Index S1: survey Index (N/10 hours) based on 'standard' unstratified area with correct ALK for ages 4–5 in 2015.

Runs S2 to S4

Instead of the 'standard' index, runs S2 to S4 make use of the swept-area indices (S2 to S4) described in Section 3.6.2. These runs differed little from each other in terms of diagnostics or results (See Supplementary figures). Compared to Run S1, residual plots are similar, with exception of age 6 which in runs S2 to S4 show residuals of a size consistent with the other age classes. The retrospective plots are marginally worse than those for Run S1, although there is no suggestion of bias.

Run S5

Run S5 uses two survey indices:

i. The 'standard' survey index as used in Run S1 up to 2009;
ii. The new survey index for 2011 onwards, worked up according to the random stratified design including all survey stations (including those in deeper water) as described in WD 1.

The model diagnostics are shown in Figures 6 and 7. Compared to Run S1, the residuals appear to be of smaller magnitude for the period 2011 onwards; in particular, in run S5, the large negative residuals between 2012 and 2015 are less apparent. However, there is a large negative residual at age 6 in 2016.

Run S5 exhibits a much poorer retrospective pattern (Figure 7) than Run S1 (Figure 5) with substantial revisions being made to both F and SSB with each successive retrospective peel. This suggests that the time-series of the second survey (2011–2017) is of insufficient length to provide a robust assessment.

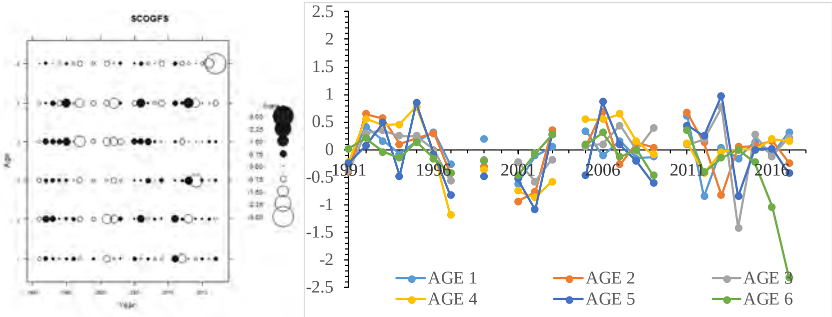


Figure 6. Haddock in Division 6.b. Run S5. Log catchability residual plots. Age plus group 7+. Two survey indices (N/10 hours): i) 'standard' index 1991–2009, and ii) new index 2011 onwards.

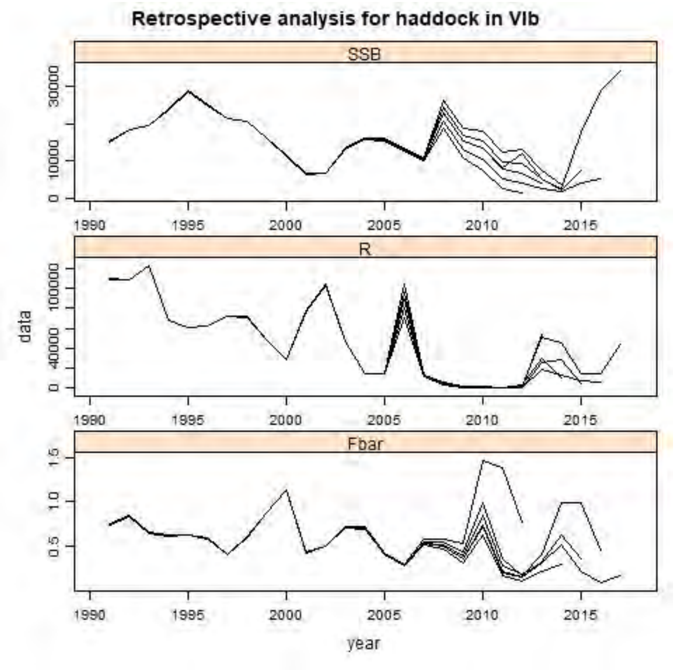


Figure 7. Haddock in Division 6.b. Run S5. Retrospective analyses. Age plus group 7+. Two survey indices (N/10 hours): i) 'standard' index 1991–2009, and ii) new index 2011 onwards.

Run S6

Run S6 differed from S5 in that it excluded the first survey index (up to 2009) and was run only including survey data from 2011 onwards. As would likely be expected for a short survey index, the residuals are generally small (Figure 8) and the retrospective analysis quite poor (Figure 9).

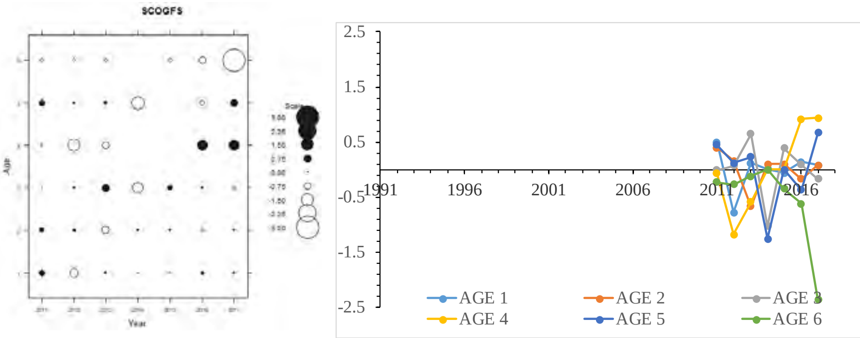


Figure 8. Haddock in Division 6.b. Run S6. Log catchability residual plots. Age plus group 7+. Single survey index (N/10 hours) 2011 onwards.

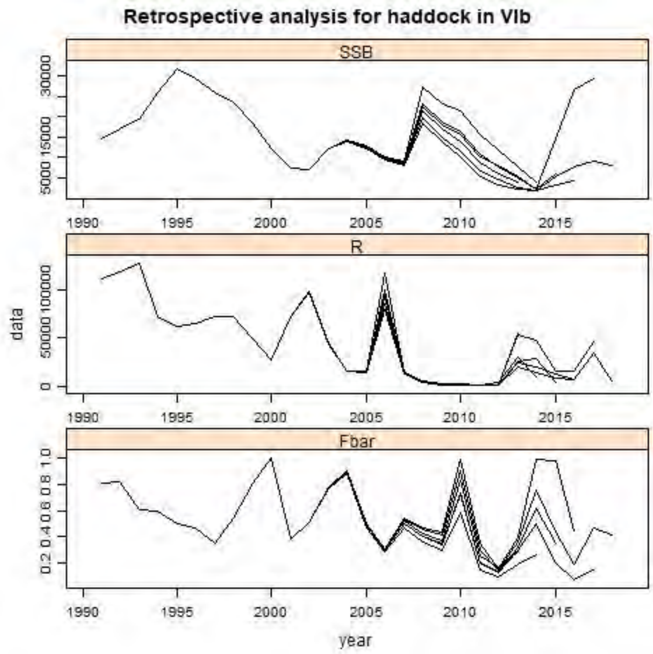


Figure 9. Haddock in Division 6.b. Run S6. Retrospective analyses. Age plus group 7+. Single survey index (N/10 hours) 2011 onwards.

Choice of Plus Group

One of the issues identified ahead of this benchmark was to explore the use of an extended age range (beyond 7+) in the stock assessment, the reason being that this would allow year classes to be tracked for a greater number of years and result in more reliable estimates when there was a period of persistent poor recruitment i.e. most of the catch in the plus group (as in 2012). Datasets were prepared with the plus group at age 7 (as previous assessment WGs) and a plus group of age 8. Higher ages were not considered due to noisier data in these age classes; lower numbers of individuals sampled and potentially more difficulties with age reading.

Model runs including alternative survey indices (runs S1 to S5 described above) were conducted with the age range extended to 8+ (and using the survey index to age 7) and are listed below:

Run P1: 'Standard' survey index (S1) and plus group at age 8.

Run P2: Swept-area index based on rectangle strata (S2) and plus group at age 8.

Run P3: Swept-area index based on bathymetric strata (S3) and plus group at age 8.

Run P4: Swept-area index with no stratification (S4) and plus group at age 8.

Run P5: Two survey indices with break in 2011 and plus group at age 8.

Diagnostic plots from Runs P1–P5 are similar to those from Runs S1–S5. Figures 10 and 11 show the log catchability residuals and retrospective from run P1. (Other runs are shown in the Supplementary Figures in Annex 3).

The residuals from Run P1 (Figure 10) show a similar pattern to those from Run S1 (Figure 4). Similar blocks of positive and negative residual, but generally small. The exception to this is the residuals at age 6 which are noticeably larger in this model run (with the plus group at age 8) than the same run with the plus group at age 7. Residuals at age 7 in Run P1 are very small across the full time-series (similar to age 6 in Run S1). The retrospective analysis for Run P1 shows slightly less consistency between retrospective peels than Run S1.

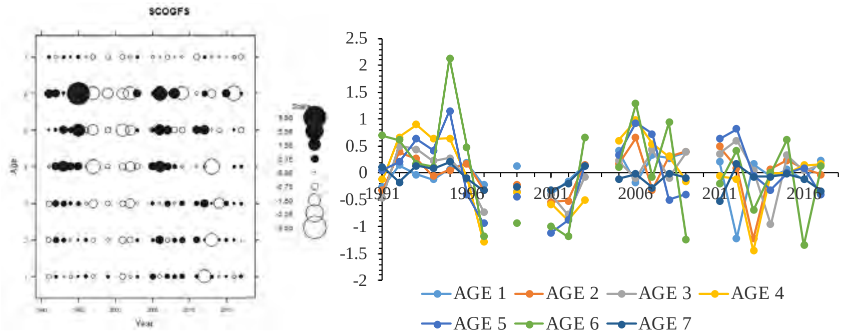


Figure 10. Haddock in Division 6.b. Run P1. Log catchability residual plots. Age plus group 8+. Index S1: survey Index (N/10 hours) based on 'standard' unstratified area with correct ALK for ages 4–5 in 2015.

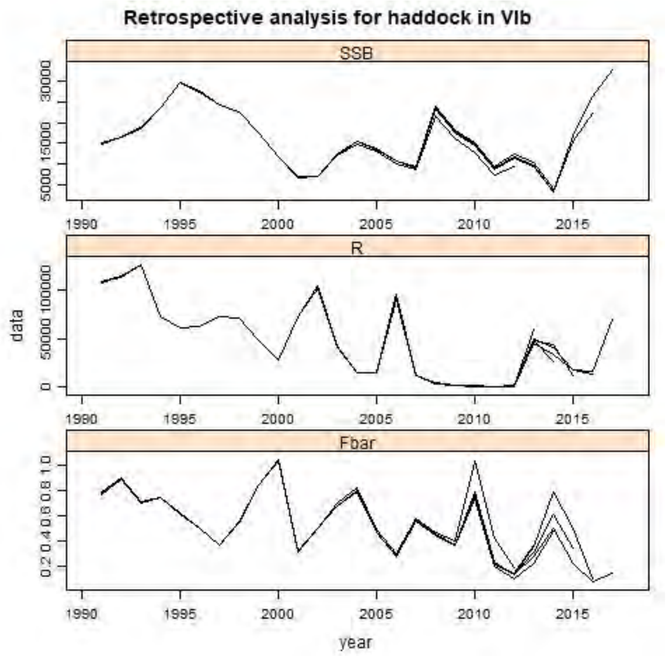


Figure 11. Haddock in Division 6.b. Run P1. Retrospective analyses. Age plus group 8+. Index S1: survey Index (N/10 hours) based on 'standard' unstratified area with correct ALK for ages 4–5 in 2015.

3.6.5 Final Assessment

The benchmark agreed that based on the model diagnostics, the 'standard' survey index with corrected ALK (S1) should continue to be used in the assessment. Although, the new survey index covering a greater extent of this stock distribution ought to provide a better index of full stock abundance it appears that this index is still too short to be used in the assessment. The benchmark therefore agreed that an additional exploratory assessment, making use of the two indices (with split at 2011), continue to be run by the assessment WG in order to monitor performance of the index.

Limited exploration of the current XSA settings was carried out with the exception of comparing the performance with alternative plus groups. The benchmark recommended that a plus group at age 7 be retained. Table 3 gives the settings for the final XSA assessment.

Figure 12 shows a comparison of the XSA stock summary from WGCSE 2018 (but with corrected survey ALK) with the final assessment agreed at this benchmark. The main difference is in the estimate of fishing mortality in 2010 which is due to the revision in estimated discards for that year. Minor differences in estimated SSB can also be seen which will also be partly attributed to the new assumptions about mean weight-at-age in the benchmark assessment.

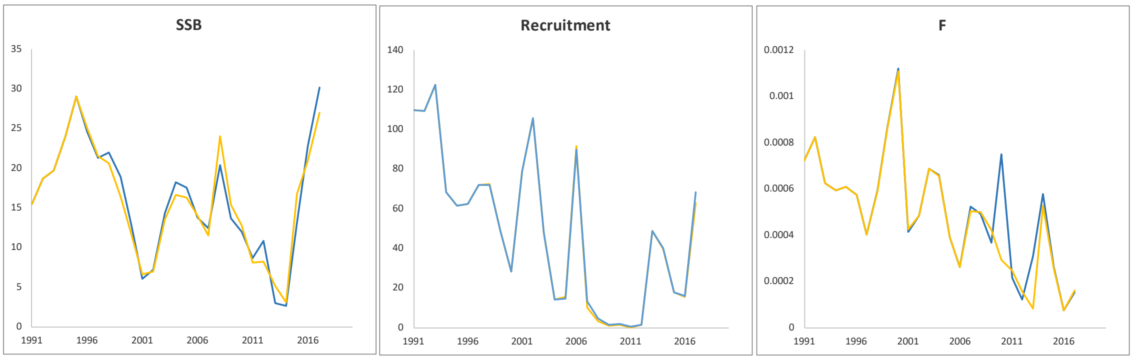


Figure 12. Haddock in Division 6.b. Run S1. Comparison of results of assessment with pre-benchmark catch and weight data (corrected 2015 survey ALK) (yellow) and final benchmark assessment using the new catch and weight data and corrected 2015 survey ALK (dark blue).

Table 3. Summary of settings for final XSA assessment.

Assessment model/software	FLXSA (Lowestoft VPA package used in exploratory assessments)		
First data year	1991		
Last data year	2017 (current year–1)		
First age Last age	1 7+		
Survivors estimates shrunk towards mean F	4 years / 3 ages		
s.e. of the means	1.0		
Min s.e. for pop. Estimates	0.3		
Prior weighting	None		
Fbar	Ages 2–5		
Tuning index	One survey (Sco Rock-IBTS-Q3)	Years: 1991 –2017	Ages: 1–6
Time-series weights	None		
Model	Power	Catchability dependent on year class strength for age <4	
Regression type	C	Min number of datapoints for regression = 10	
Q plateau	Age 5		

3.7 Reference points

To be completed at the management strategy evaluation meeting (August, 2019).

3.8 Future research and data requirements

A number of issues were identified for addressing either intersessionally or in preparation for a future benchmark:

- Current assessment model

  The final XSA presented at this benchmark meeting was deemed to be adequate for the provision of advice. However, as noted in Section 3.6.4, there remain some odd residual patterns, with the smallest residuals being apparent for the oldest true age (age 6) in the assessment which coincides with the oldest age in the survey index. The same pattern is generally true for all of the assessments presented with the exception of those using a swept-area index and plus group at age 7 (Runs S2 to S4). Further investigation is required (which could be conducted within the WG) to understand why this would happen and whether it undermines the assessment.

- Alternative assessment methods

  This stock continues to suffer from very poor sampling levels due to the distant nature of the fishery (and small number of vessels operating there). The age compositions and discard estimates, are likely to be highly uncertain (both historically and more recently), with age compositions also impacted by uncertainty in age readings. It would therefore make sense, in future, to explore alternative assessment approaches which allow for uncertainty in these data. Potential methods could include SAM (Nielsen and Berg, 2014) and a4a (Jardim *et al*., 2015). These models also have the advantage in that they provide estimates of uncertainty for fishing mortality and SSB.

- Biological data

  Concerns have been raised, regarding differences in age reading of Rockall haddock between readers in different laboratories. A haddock ageing exchange convened under the recommendation of WGBIOP (ICES, 2018) and coordinated by Marine Scotland Science is currently ongoing to explore the issue and recommend a standardized approach. This process is considering haddock from Rockall, ICES Division 6.a, Division 4.a and subareas I and II. How the results of this study impact the age composition data going into the assessment will need to be considered ahead of the next benchmark.

Other biological parameters such as maturity and natural mortality have remained fixed for many years. Consideration should be given to updating them ahead of future benchmark working groups. Most other demersal stocks assessed by WGCSE now make use of age dependent natural mortality estimates based on, for example, Lorenzen (1996). A similar approach could potentially be adopted for Rockall haddock.

The current maturity ogive assumes knife-edge maturity at age 3. A number of maturity studies based on both Russian and Scottish sampling suggest a high proportion of individuals at age 2 may actually be mature. There is therefore a need to collate all the information on maturity and evaluate whether the current ogive requires updating.

3.9 External reviewers comments

WKROCK turned out to be a challenging benchmark workshop due to insufficient preparation of data inputs, model runs and thorough documentation of the assessment process before the workshop. While a working document on survey design and indices and another one on different ways to predict mean weight-at-age for the short-term projections/forecast were available before the workshop, catch-at-age raising in InterCatch had to be done during the benchmark workshop after doubts appeared regarding the raising procedure used so far. This stresses the importance of carrying out a complete data compilation workshop before the benchmark workshop itself like is usually done for other ICES benchmarks. No test runs with alternative input had been prepared. Because of the limited preparation status, only little could be achieved and work focused more on a review of the current assessment rather than on advancing the assessment and subjecting it to a complete review of data inputs, modelling assumptions and an examination of the advice consequences of alternate model runs.

3.9.1 Following topics were discussed during the benchmark workshop

3.9.1.1 Catch-at-age

For recent years it was unclear what method has been used to raise wanted and unwanted catch. Data were uploaded to InterCatch but the extraction of data in a format suitable to be input into the model was not finalized. Therefore, data from 2012 onwards were reprocessed in InterCatch to ensure a proper raising of wanted and unwanted catch. This is a step forward to a more transparent process of creating input data to the assessment, but it also means that a significant amount of time was spent on this task and not on other important aspects of the review.

3.9.1.2 Mean weights in the catch

Mean weight-at-age is especially uncertain for older age groups of small cohorts. An outlier is obvious for 2016 where the mean weight for ages 5+ show a serious jump upwards. Only very few fish of ages 5+ were sampled during this year. This outlier has important consequences on the assessment and it is important to ensure that it does not get a high weight in short term forecasts for the next years. It is unclear whether it is a sampling artefact or the consequence of very low year classes that experienced fast growth through density-dependent mechanisms. In general, there is a trend towards higher mean weight-at-age and higher interannual variability in recent years. Because of this and because samples collected during the survey do not show such pronounced variations in weight-at-age between years, it was concluded to use smoothed stock weights-at-age (five years moving average?) instead of the raw catch-at-age values.

Different methods to estimate mean weight-at-age for the short-term projections were tested in a retrospective analysis (WD 2). This analysis shows that, in recent years, a three or five years average shows some bias but less variation around the “truth”. Taking only the last available year gives an unbiased estimate but much more variation around the “truth”. When analysing the full time-series from 1991 onwards, the bias for the three and five year average is much lower. The working group considered that a three-year average would be the best option provided that the 2016 outlier be either excluded or down-weighted in the short-term forecast.

3.9.1.3 Survey index

The survey index for the Scottish trawl survey has been reanalysed. The survey design changed from a fixed station design to a random stratified survey in 2011. The depth range covered also increased after 2011 (from 250 to 350 m). So far, only the stations inside the original grid and depth range are used for the years after 2011 and only one time-series from 1991 to 2017 is provided. During the benchmark an error in the 2015 ALK used for the survey has been corrected.

Seven versions of tuning indices were tested and the diagnostics of the model runs were analysed:

Runs 1 and 2: Time-series used so far from 1991 to 2017 (baseline) with and without the corrected ALK for 2015.

Runs 3 to 5: Three versions of swept-area indices with different stratifications (5 strata based on bathymetry, geographic strata of 15' latitude wide and 15' longitude long and no stratification).

Run 6: Split the survey index time-series in 2011 and use all stations conducted after 2011 including the stations from deeper areas.

Run 7: Only use the new time-series after 2011.

The working group concluded that an assessment with the currently used survey time-series and the corrected ALK is the best option for now. Results were similar for all versions tested, but the retrospective pattern was considerably worse in the runs using the new time-series from 2011 onwards (Runs 6 and 7). The likely reason is that the time-series is still not long enough to produce robust assessment results. The three swept-area indices could have been also an option. Due to the limited preparatory work carried out before the workshop, there was however not enough time left during the workshop for a proper analysis and a review of the estimation process behind the indices calculations.

3.9.1.4 Plus group

Age 7 has been treated as plus group thus far. There is a potential to increase the plus group to higher ages. This would help to follow strong and weak cohorts over more years. However, the sampling level deteriorates at higher age groups and age reading problems become more apparent. Therefore, only an assessment with age 8 as plus group has been considered as plausible alternative.

A test assessment was thus conducted with age 8+. XSA results were similar for both options (plus group set at age 7 or at age 8) and the retrospective pattern was slightly better for the 7+ run. The sampling level deteriorates above age 7. In addition, apart from two years (where only stronger cohort are left in the plus group and all other cohorts are very weak) the proportion of catch numbers in the 7+ group is rather small (<10%). Therefore, the working group concluded that the assessment with a plus group set at age 7 is still appropriate. It is worth noting that the residuals were smaller for the highest age used in the survey regardless whether a 7+ or 8+ group was used. This indicates that the model closely follows the survey index of the oldest age in the tuning data (age 6 or 7 dependent on the plus group). It was not possible for the WG, due to time constraints, to analyse/ examine this feature but further investigations could be carried out until the next benchmark or during an inter-benchmark (see also recommendations).

3.9.1.5 FLXSA

The assessment was conducted using the XSA software from Lowestoft. The benchmark group felt that, while the software robustness and appropriateness were not in doubt, relying on legacy software not being developed anymore was a hindrance to further development of the assessment and to achieving a higher level of transparency and reproducibility. As such, effort was made during the benchmark to replicate the XSA assessment run with Lowestoft software in FLXSA. Although some more details have to be adapted in the code, the results of the Lowestoft XSA and FLXSA were found to be reasonably similar to each other. The FLXSA assessment has been included in the ICES TAF framework, therefore enabling a more transparent and collaborative assessment in future years. Furthermore, the FLXSA implementation will facilitate the planned MSE simulations.

3.9.2 Final conclusions

The external reviewers agree with the conclusions of the benchmark and consider that the final benchmark assessment can be used as basis for advice. It considers however that more work is still needed in the next years to improve the input data and the assessment methodology. Reference points need to be determined based on the final assessment according to ICES reference point guidelines.

3.9.3 Recommendations

Near future until the next advice

It is strongly recommended to check and update the stock annex including a description of the raising of catch data done in InterCatch.

Longer-term

A data compilation workshop is needed prior to the next benchmark. Otherwise a successful benchmark workshop cannot be guaranteed. The level of preparation needs to be checked more carefully before the final benchmark workshop. Meetings by correspondence well before the benchmark workshop may help to detect problems well in advance.

As it stands, the compilation of the Rockall haddock stock assessment input files from InterCatch is based on numerus manual data processing. It is thus error-prone and not readily reproducible. Further automatisation of the input data compilations is highly recommended to avoid unnecessary sources of error and to improve transparency. It is worth noting that this is a general issue for all stocks currently using InterCatch. A new system based on the regional databases and creating input data files using R scripts would be beneficial.

XSA is a deterministic model and therefore results are directly impacted by uncertainties in age readings and the amount of catch reported. Moving to a statistical model-like SAM or SS3 would permit to better account for such uncertainties in the model fits. These models also provide confidence regions on the parameter estimates and stock dynamics indicators outputs (F, SSB, R) associated with those uncertainties.

The survey index is derived by applying mean density estimates for different depth strata. However, other factors may also influence the spatial distribution of haddock potentially biasing the survey index. In addition, a large part of the area on Rockall is currently protected, and fishing is not allowed. It is assumed for the index calculation that densities inside and outside these zones are similar. It may be tested with historical information whether this assumption is likely correct or not.

The survey from which the index is derived changed its sampling design in 2011. It is recommended to further explore whether the currently used survey index could be split into two separate indices once more years with the new design (and better coverage of depth strata) are available.

The expert that presented the analyses conducted to generate the survey index also mentioned the presence of noticeable diurnal trends in the catches (i.e. haddock seem to be more catchable during the day than at night). The effects that this diurnal variation in trawl catch efficiency can have on the survey index should be further investigated.

When forecasting mean weight-at-age, cohort strength effects seem to play a role. It is recommended to analyze whether methods including cohort effects for mean weight-at-age could improve the short-term forecasts of yield and biomass.

Currently a knife-edge maturity ogive at age 3 is used. A smoother maturity ogive may be derived from further analyses of available data.

Natural mortality is currently kept constant over ages and years. Alternatives may be tested for the next benchmark. At least age-dependent natural mortalities could be derived based on life-history analyses.

The stock coordinator pointed out that discrepancies exist in the ageing of Rockall haddock, noting that different ages were estimated by different laboratories using the same otoliths. The accuracy and precision of ages obtained from haddock otoliths should be further investigated so as to establish whether ageing biases are present, and to determine whether/how they will affect the generation of catch-at-age matrices derived from age–length keys.

The residuals in the final benchmark assessment are smallest for the oldest age in the survey time-series. This indicates that the assessment model follows the index for the oldest age more closely than the indices for the younger ages. Runs with the proposed swept-area indices in combination with age 7 as plus group do not show this feature. Further investigations on the estimation procedure behind the proposed swept-area indices are needed next to analyses of whether it is problematic if the assessment follows the indices for the oldest age in the survey to a larger extent.

4 References

ICES. 2005. Report of the Working Group on the Assessment of Northern Shelf Demersal Stocks (WGNSDS) 4–13 May 2004 ICES Headquarters, Copenhagen. ICES CM 2005/ACFM:01.

ICES. 2012a. Report of the International Bottom Trawl Survey Working Group (IBTSWG), 27–30 March 2012, Lorient, France. ICES CM 2012/SSGESST:03. 323 pp.

ICES. 2012b. Report of the Working Group for the Celtic Seas Ecoregion (WGCSE), 9–18 May 2012, Copenhagen, Denmark. ICES CM 2012/ACOM:12.

ICES. 2017. Manual of the IBTS North Eastern Atlantic Surveys. Series of ICES Survey Protocols SIPS 15. 92pp. http://doi.org/10.17895/ices.pub.3519.

ICES. 2018. Working Group on Biological Parameters (WGBIOP), 1–5 October 2018. Ghent, Belgium. ICES CM 2018/EOSG:07. 186 pp.

Jardim, E., Millar, C. P., Mosqueira, I., Scott, F., Osio, G. C., Ferretti, M., Alzorriz, N., and Orio, A. 2015. What if stock assessment is as simple as a linear model? The a4a initiative. ICES Journal of Marine Science, 72: 232–236.

Jaworski, A. 2011. Evaluation of methods for predicting mean weight-at-age: an application in forecasting yield of four haddock (*Melanogrammus aeglefinus*) stock in the Northeast Atlantic. Fisheries Research, 109: 61–73.

Lorenzen K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49, 627–647.

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

Annex 1: List of participants

Participant	Institute	Country	Email
Michel Bertignac Invited Expert	Ifremer Centre de Brest	France	Michel.Bertignac@ifremer.fr
Helen Dobby ICES Chair	Marine Laboratory Marine Science Scotland	Scotland, UK	h.dobby@marlab.ac.uk
Andrzej Jaworski	Marine Laboratory Marine Science Scotland	Scotland, UK	jaworskia@marlab.ac.uk
Alex Kempf External Chair	Institute of Sea Fisheries Thünen Institute	Germany	alexander.kempf@thuenen.de
Vladimir Khlivnoy	Knipovich Polar Research Institute of Marine Fisheries and Oceanography PINRO	Russia	khlivn@pinro.ru
Daniel Ricard Invited Expert	Fisheries and Oceans Canada Moncton	Canada	daniel.ricard@dfo-mpo.gc.ca

Annex 2: Working Documents

WK1: Q3 Rockall Haddock Survey and index evaluation

Andrzej Jaworski, Marine Scotland – Science

March 2019

Survey design

Three regular groundfish surveys are being conducted at present off the West Coast of Scotland: in Q1 and Q4 in ICES Subarea 6a (SCOWCGFS-Q1 and Q4) and in Q3 in ICES Subarea 6b (SCOROC-Q3). The surveys are described in detail in "Manual of the IBTS North Eastern Atlantic Surveys" (ICES, 2017). The Scottish Rockall Haddock Survey (SCOROC) is primarily a juvenile haddock survey. Since 2011, the SCOROC has utilised a random stratified survey design with station positions being randomly distributed within a series of "*a priori*" sampling strata. Prior to this, the survey employed a fixed station format with a bathymetric limit of 240 m. Evidence from other surveys undertaken on Rockall Bank (notably the Q2 MSS monkfish survey) raised concerns that the existing survey was missing significant components of the Rockall haddock stock found in depths greater than 240 m. The observed haddock distribution raised the idea of increasing the precision of the survey through stratification. The survey design was thus revised in 2011, and the upper limit has now been set to 350 m resulting in four sampling strata: 0–150, 151–200, 201–250, and 251–350 m denoted as "R1", "R2", "R3" and "R4", respectively (Figure 1). Figure 2 further demonstrates that only a fraction of the population was adequately sampled in the past and that the extension of the sampling area to 350 m was appropriate. All the three surveys in ICES Area 6 conducted since 2011 have been undertaken using the standard GOV research trawl, but with a modified ground gear of the type "D" (replacing the ground gear "C"), the new ground gear being more suited to the hard and often undulating topography encountered within ICES Area 6 (ICES, 2017).

To maximise the precision of the fish density estimates for the total survey area, survey effort was allocated among strata in such a way that the proportion of the samples in each stratum (n_i/n) was given by:

$$n_i/n = \frac{A_i s_i}{\sum_{i=1}^{S} A_i s_i} \qquad (1)$$

where A_i = area (m^2) of stratum i, s_i = standard deviation within stratum i, S = number of strata (Gunderson, 1993). Thus, more sampling effort was allocated to larger strata and those with a higher within-stratum variance. The selection of stations was carried out randomly in each stratum (given the number of hauls per stratum), and with the constraint that the minimum distance between two nearest stations was 7 nmi. This ensured that (a) each possible sample point had an equal chance of being selected; and (b) that there was an even coverage of samples throughout the strata (avoiding clustering of samples and concomitant large open spaces without samples). On agreeing the above points, and assuming a similar number of stations to previous Rockall surveys (n = 40), the allocation for each stratum should be as shown in Table 1. In addition to the selected stations, a number of alternative positions (Table 1) were proposed for each stratum in the event of a sample not being obtainable at the originally identified location.

To choose the shortest route among stations, an optimising algorithm ("Travelling Salesman") was also created (Figure 1).

Table 1. Depth strata used on Rockall surveys and number of stations per stratum.

Strata	Depth range (m)	Number of stations*
R1	0–150	5 (1)
R2	151–200	21 (5)
R3	201–250	10 (7)
R4	251–350	4 (9)

*** Number of stations in 2018. The number of alternative positions is given in the brackets.**

Figure 3 shows haddock distribution by age in Subarea 6.b as observed in the recent period (2011–2018), that is, with the new survey design. From 2012, we have been observing relatively strong year classes (with an above-average strength). We note the 2016 year class being comparable in strength with the very strong 2005 year class (partly seen in this figure). The catch rates of haddock at age 1+ in the last two years were particularly high and comparable with those in 2006 (not shown in this figure).

Index calculation

The index calculation presented in this section is applied to all the three Scottish West Coast Groundfish Surveys in both ICES subareas 6a and also 6b. However, formally, the current abundance index has been used up to now in the assessment of haddock in Subarea 6b (ICES, 2018).

In the "new" index calculation, numbers at length (the length frequencies, LF) per haul are standardised to numbers per hour towing. In previous years, all otoliths from all hauls in a given demersal sampling area were combined to create an age–length key (ALK) for that area (Holmes, 2008). With the new design, all otoliths taken within each of the four strata are combined to form an ALK. This ALK is applied to all LFs in the stratum individually to produce age frequencies for each haul. Finally, for each stratum, the age frequencies are summed and the values divided by the number of valid hauls to provide numbers-at-age per hour. This procedure can be summarised as:

$$CPUE_{i,a} = \frac{\sum_{h=1}^{H_i} \sum_{l=l_{\min}}^{l=l_{\max}} N_{i,a,l,h}}{H_i}$$

where $N_{i,a,l,h}$ is the number of fish at age a and length l caught during haul h, H_i is the number of valid hauls in stratum i and $CPUE_{i,a}$ is the catch per unit effort of fish at age a in stratum i.

For each age, the age frequency for each stratum is raised by the stratum area. These raised frequencies are then summed and the result divided by the total area in the assessment region. The final index value for each age is given by:

$$I_a = \frac{\sum_{i=1}^{S} CPUE_{i,a} A_i}{\sum_{i=1}^{S} A_i}$$

where A_i = area (m^2) of stratum i and S = number of strata The same procedure as described above was applied to the SCOWCGFS-Q1, SCOWCGFS-Q4 and SCOROC-Q3 data from 2011 onwards.

The calculated abundance index (with its variance) in the Rockall haddock survey for the years 2011–2018 is shown in Table 2.

Table 2. Abundance index with its variance in the Rockall haddock survey in 2011–2018. Effort unit = 1 h.

Index									
Year	Age								
	0	1	2	3	4	5	6	7	8
2011	0.534	1.586	13.760	1.792	6.795	10.145	81.659	0.261	0.271
2012	1477.860	0.215	0.847	5.582	0.959	5.930	3.203	41.267	0.528
2013	335.879	1206.143	0.804	2.192	3.617	2.291	2.808	7.224	26.934
2014	192.584	614.609	527.452	0.384	0.000	0.882	0.000	0.661	0.638
2015	121.167	223.797	539.005	419.488	0.000	0.000	0.860	0.054	0.642
2016	3344.108	115.450	140.312	244.432	170.292	1.355	0.076	0.353	0.076
2017	1858.348	2385.274	61.522	96.659	159.560	69.167	0.071	0.019	0.092
2018	611.872	287.879	1039.564	24.922	53.229	85.683	32.510	0.000	0.000
Variance									
Year	Age								
	0	1	2	3	4	5	6	7	8
2011	0.029	0.278	16.974	0.191	3.410	5.267	259.216	0.003	0.014
2012	1089596.850	0.007	0.054	2.733	0.067	2.776	0.450	61.601	0.012
2013	6251.967	33465.290	0.252	1.525	8.448	0.861	3.605	26.251	141.346
2014	1959.608	4033.312	6206.531	0.050	0.000	0.079	0.000	0.030	0.051
2015	651.239	614.547	1715.185	3143.844	0.000	0.000	0.127	0.001	0.016
2016	5494577.522	213.677	294.217	1301.138	889.363	0.069	0.003	0.023	0.003
2017	1843547.855	49111.905	68.745	143.322	414.159	61.842	0.000	0.000	0.002
2018	12048.123	1167.847	13412.498	54.894	148.671	622.305	148.162	0.000	0.000

Index evaluation

In the assessment of haddock in Subarea 6b, one tuning series has been used up to now that covers two periods, 1991–2009 (before the survey design change in 2011) and from 2011–present (ICES, 2018). In the year 2010, the survey was not carried out due to an engine breakdown of the research vessel. For the period since 2011, only the subset of stations which occur within the depth range of the original survey (pre-2011) have been included in the index calculation and hence the survey index has been treated as a continuous time-series. The rationale for using the same time-series was that it covers the same "standard survey area" (pre-2011 survey coverage) with depths up to 240 m. At present, six age groups are used in the VPA assessment with ages 1–6. No documented evaluations have been made as to how the new survey design affects the index.

In this section, three tuning series are explored:

- The current assessment time-series (1991–2017);
- The trimmed assessment time-series (1991–2009); and
- The proposed new time-series (2011–2018).

The three tuning series are compared with the aim of providing one reliable index for each of the two time-series, 1991–2009 and 2011–2018, for future assessments.

The mean standardised catch proportions-at-age per year are similar in the first two tuning series for years before 2011 (Figure 4). The calculated index differ in the first and third time-series from 2011 with a tendency of lower values in the new time-series, which is understandable as it was calculated for a larger area. However, this did not affect largely the catch proportions. The three plots indicate strong year classes (notably the 2005 year class, and to a lesser extent the 2012 year class), but also consistently weak year classes. The high value at age 5 in 2011 shows some abnormality compared to the other ages of the 2006 year class, and this can be seen in the first and third plots. Apart from this outlier, year-class tracking is reasonably consistent up to age 7.

Figure 5 shows the log mean standardised indices in the three tuning series by year class and year. As before, there are almost no differences between the first two tuning series before 2011 and some differences can be seen between the first and third from the year 2011. Overall, year classes are tracked relatively well in the three tuning series with no obvious year effects.

Figure 6 shows the survey scatterplots. For the three tuning series, there is a general consistency in the estimates of the year-class strength across age groups, but the points are more scattered for old age groups. The second time-series is the least consistent, but it is shorter than the first one. For the third tuning series, the index values show high consistency, but there are still few years in this time-series.

Figure 7 shows log catch curves for the survey. Overall, the year classes are well tracked. The differences for the first and third time-series from the year 2011 are mainly due to the difference in the area for which the index was calculated. Generally, the decline in numbers has been slower in recent years compared to the earlier years of the time-series. For strong year classes born after 2011, the curve shapes are rather regular. For weak year classes seen in the recent period (from 2011 till present), the index was noisier.

Since the abundance index calculated before 2011 does not adequately represent the haddock population in Subarea 6b, it is justifiable to discontinue this time-series in 2009, but retaining it in the assessment. With a better coverage of the population in the new survey, the new index is a better representation of the stock and its respective age groups. Using seven age groups (ages 1–7) in the new tuning series (rather than six, as it has been the case till now) from 2011 onwards will provide more insight on exceptionally strong year classes to effectively conduct the assessment. It will be also consistent with using seven age groups for catch data.

References

Gunderson, D. R. 1993. Surveys of Fisheries Resources. John Wiley & Sons, Inc. New York, 248 pp.

Holmes, S. J. 2008. ROAME MF0170: Alternative Survey Index Estimation. FRS Production of Scottish West Coast Survey Indices. Fisheries Research Services Internal Report No 10/08.

ICES. 2017. Manual of the IBTS North Eastern Atlantic Surveys. Series of ICES Survey Protocols SISP 15. 92 pp. http://doi.org/10.17895/ices.pub.3519.

ICES. 2018. Report of the Working Group on Celtic Seas Ecoregion (WGCSE), 9–18 May 2018, Copenhagen, Denmark. ICES CM 2018/ACOM:13. 1251 pp.

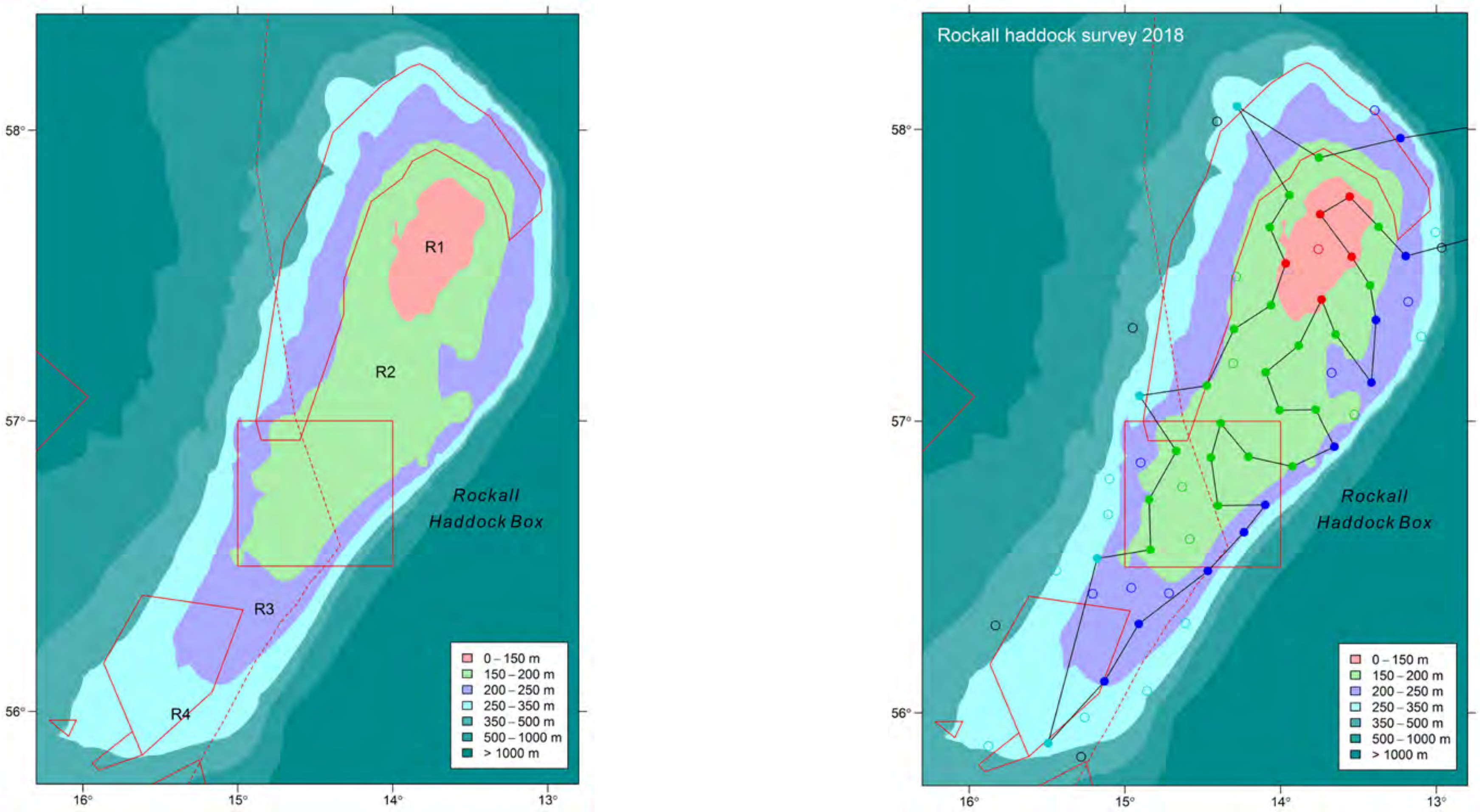


Figure 1. Strata in UK-SCOROC-Q3 (left panel). The red polygons show protected areas (including the Rockall Haddock Box). The dashed line shows the border between the NEAFC Regulatory Area and EU waters. The Rockall haddock survey in 2018 (right panel). Strata and allocation of sampling effort among strata with the optimised route. Filled circles show the selected sampling stations and empty circles show alternative stations. The black empty circles show alternative positions outside the four strata.

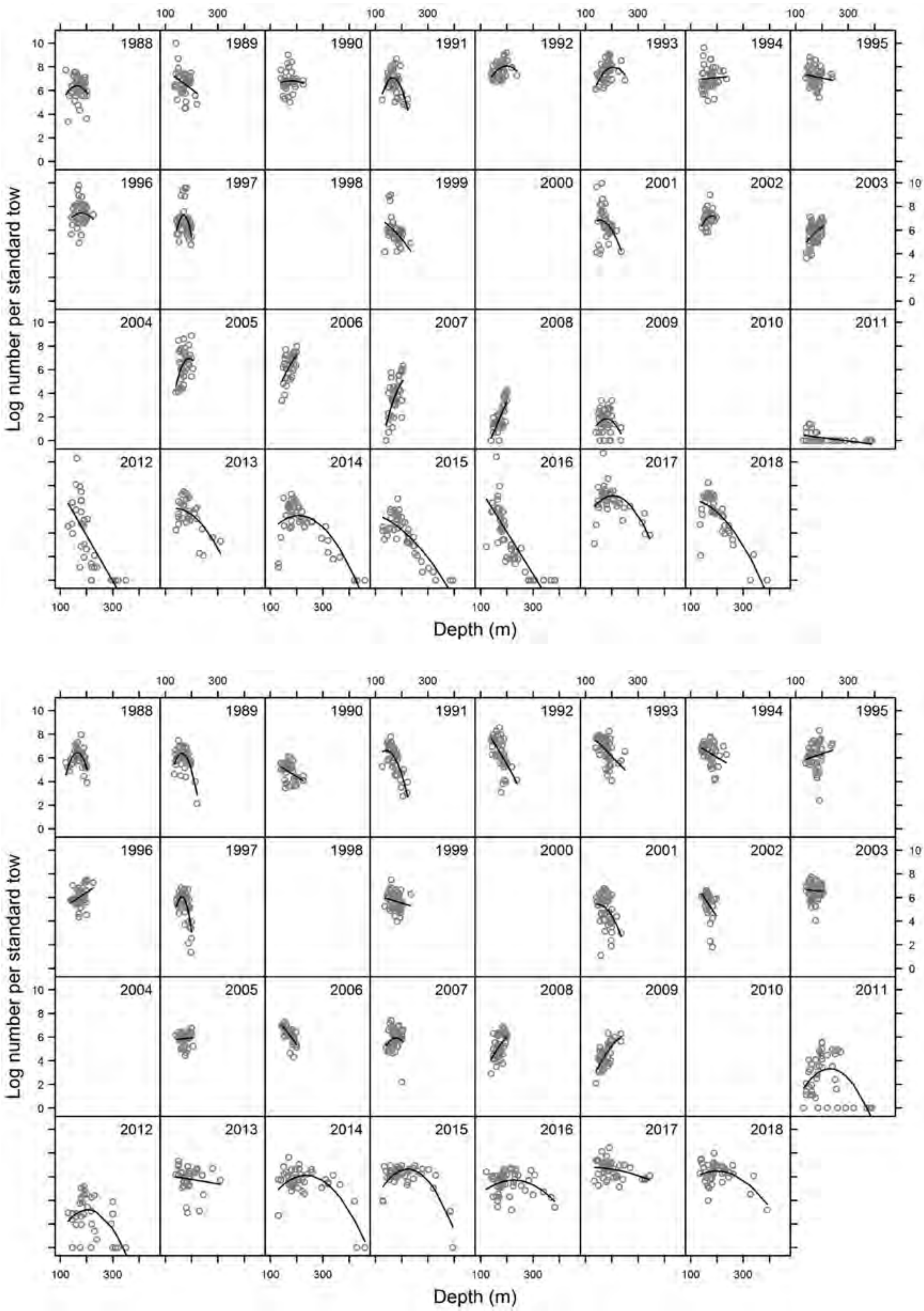


Figure 2. Observed juvenile (fish below 25 cm, upper panel) and adult (fish over 25 cm, lower panel) haddock densities at depth as observed in the Rockall Haddock Survey in 1988–2018 (grey points). The standard haul duration = 30 min. A proxy for the linear model or the quadratic polynomial is fitted to the data (black line).

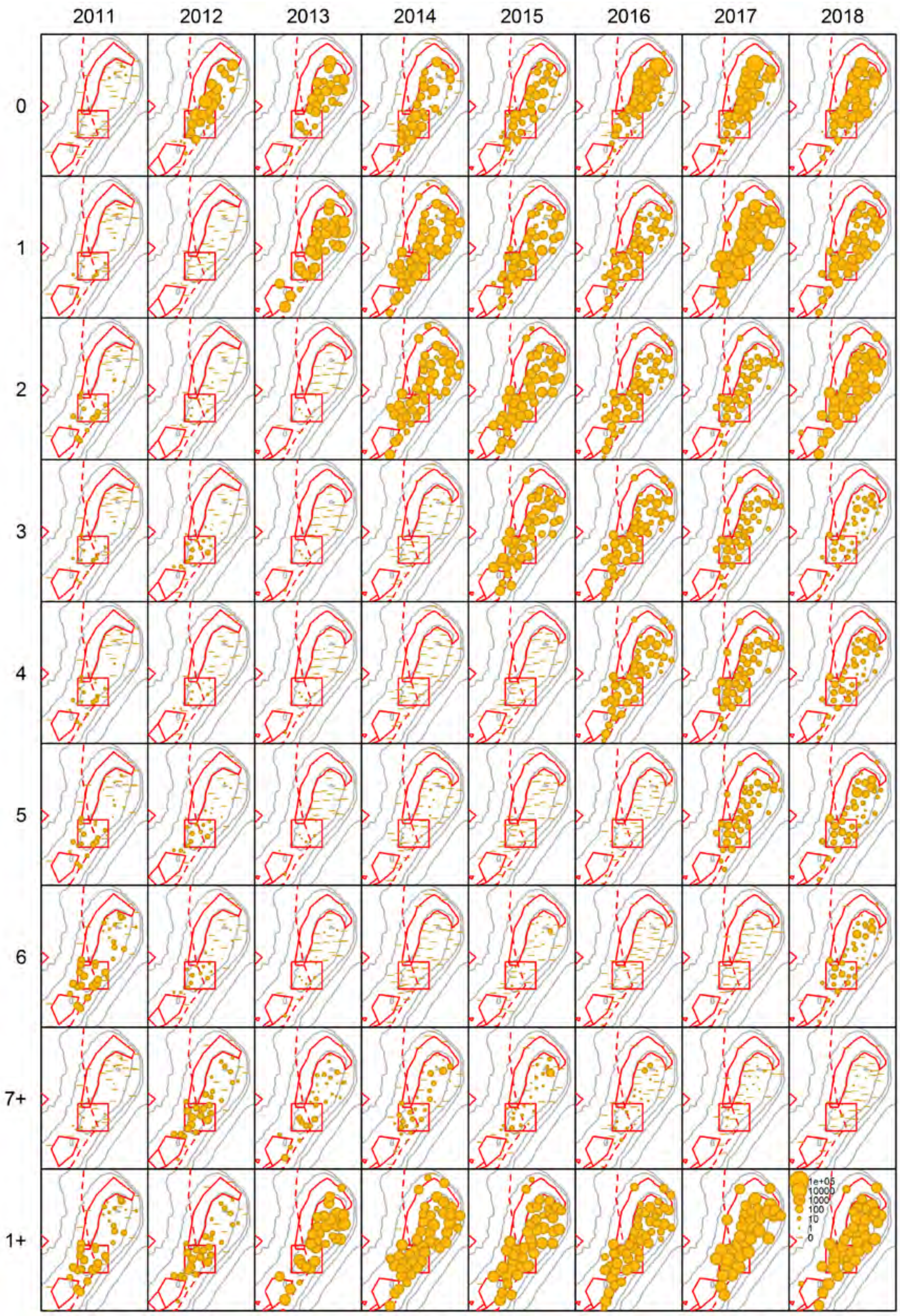


Figure 3. Haddock distribution by age on the Rockall Bank in 2011–2018 as observed in the Rockall Haddock survey. The densities (numbers of fish per 30 min) are represented by circles. The red polygons show the protected areas. The red rectangle in the centre shows the Haddock Box. The dashed line shows the NEAFC Regulatory Area.

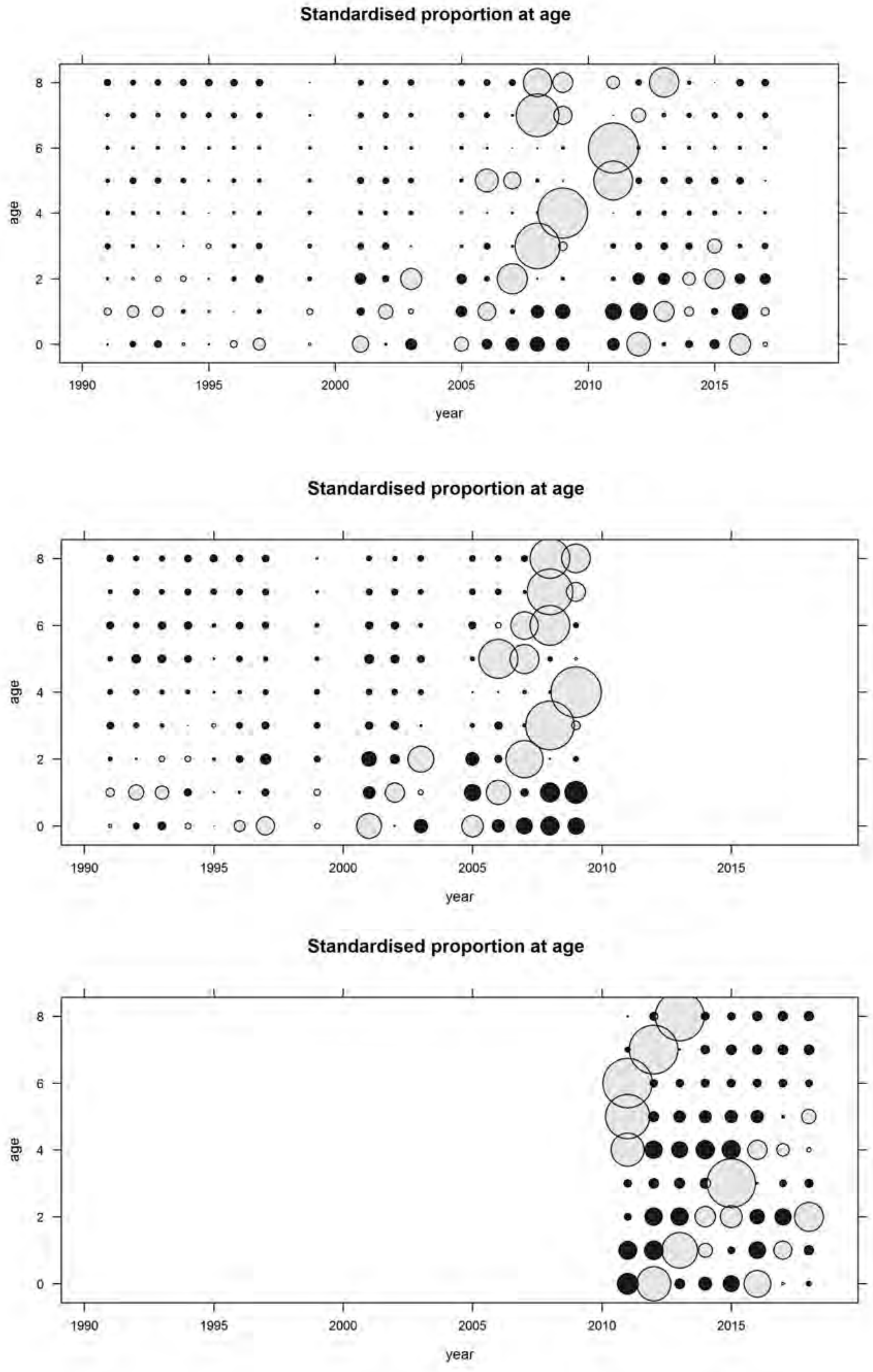


Figure 4. Standardised proportions at age per year ("spay"), three tuning series, the current assessment time-series (1991–2017, upper panel), trimmed assessment time-series (1991–2009, middle panel) and proposed new time-series (2011–2018, lower panel).

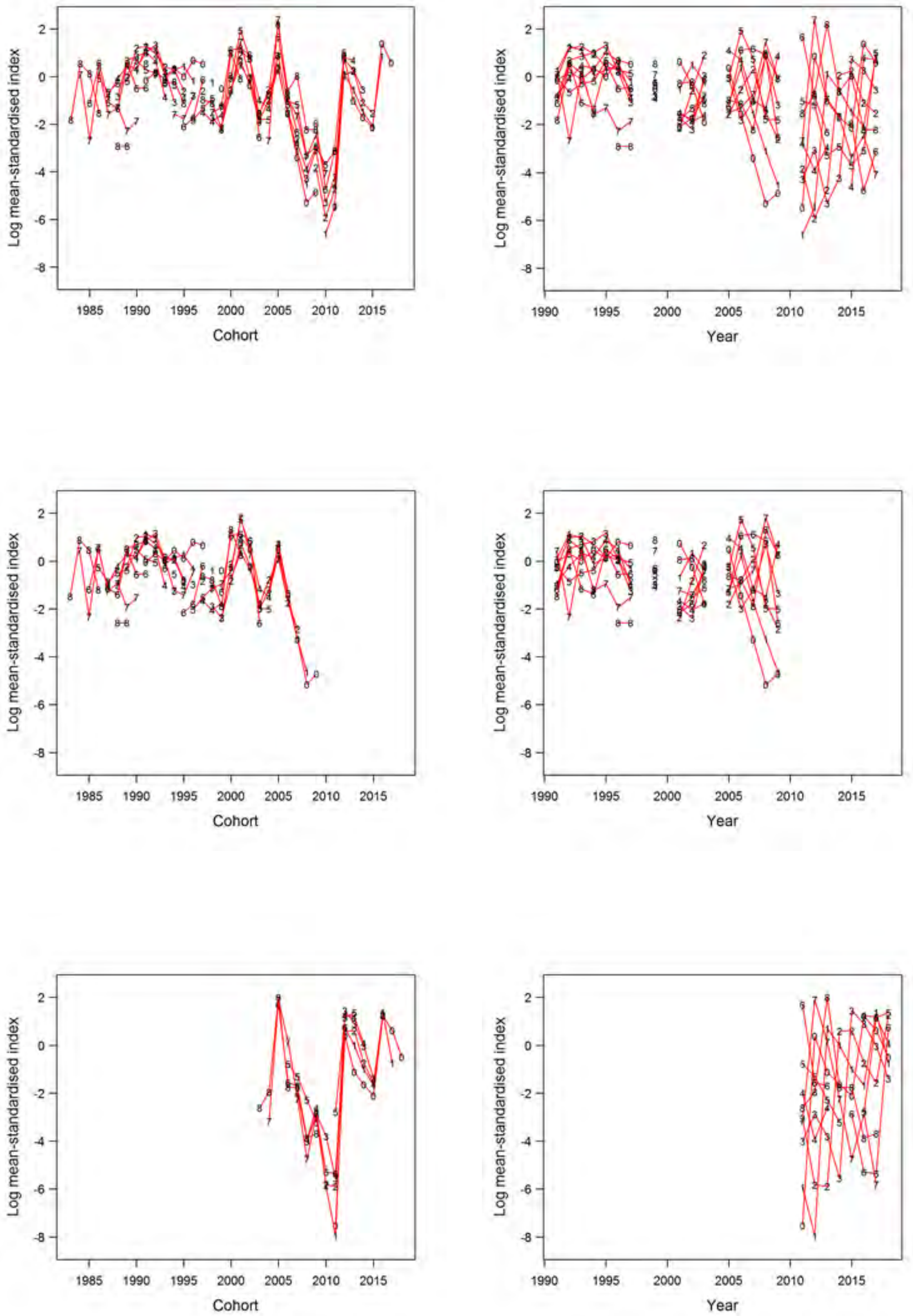


Figure 5. Log mean standardised index values, by year class and year, three tuning series, the current assessment time-series (1991–2017, upper panels), trimmed assessment time-series (1991–2009, middle panels) and proposed new time-series (2011–2018, lower panels).

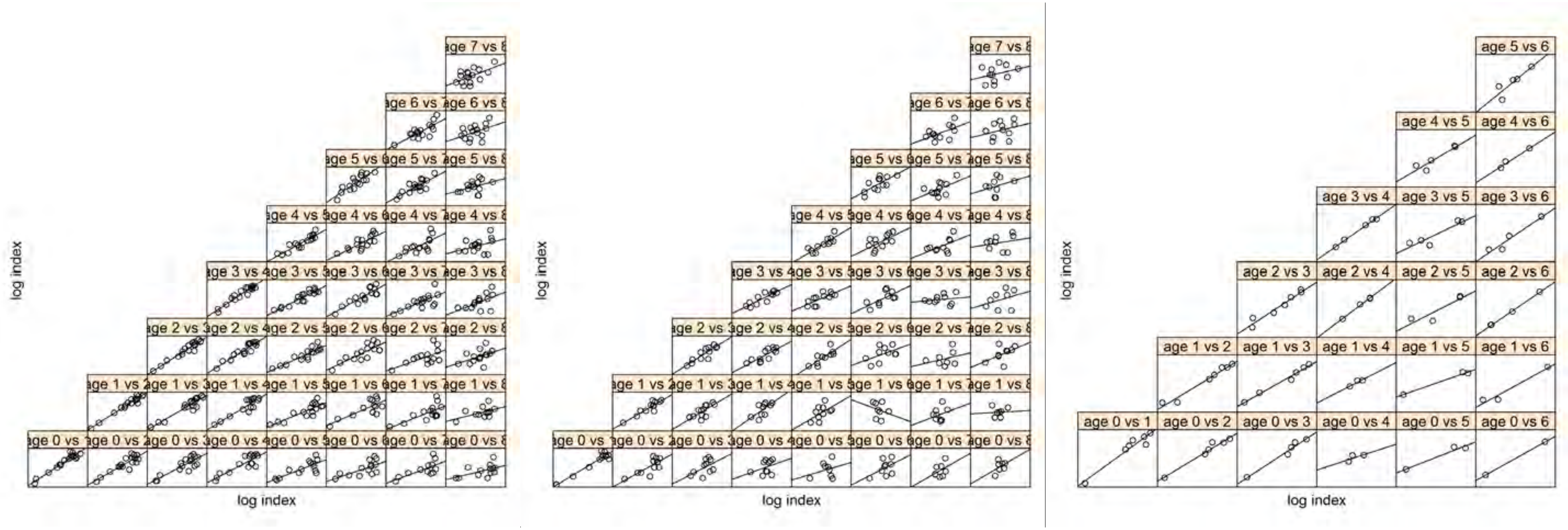


Figure 6. Within-survey correlations comparing index values at different ages for the same year classes, three tuning series, current assessment time-series (1991–2017, left panel), the trimmed assessment time-series (1991–2009, middle panel) and proposed new time-series (2011–2018, right panel). The straight line is a linear regression.

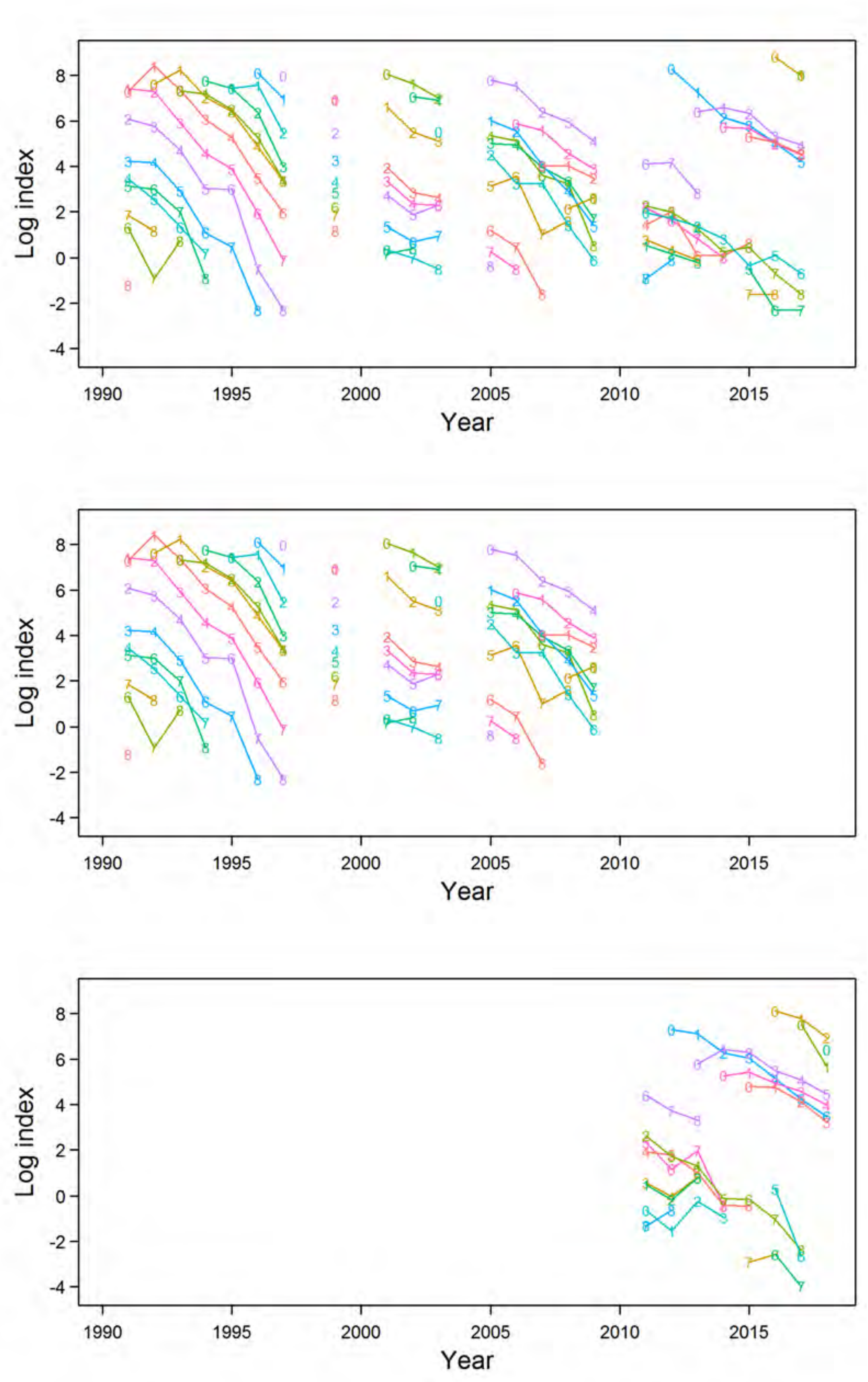


Figure 7. Log catch curves in the three tuning series, the current assessment time-series (1991–2017, upper panel), trimmed assessment time-series (1991–2009, middle panel) and proposed new time-series (2011–2018, lower panel).

WD 2: Prediction of mean catch weight-at-age and estimation of mean stock weight-at-age for haddock in Division 27.6b (Rockall)

Andrzej Jaworski, Marine Scotland – Science

March 2019

Prediction of mean catch weight-at-age

Methods for predicting mean catch weight-at-age (referred to in this section as "weight-at-age" or just as "weight") of the haddock stock in Division 27.6b (Rockall) were analysed. Their performance in short-term forecasts of yield (catch in weight) was tested in a retrospective analysis conducted for a relatively recent 12-year period (from 2006 to 2017). This analysis was a follow-up of the evaluation carried out for four haddock stocks in the Northeast Atlantic, which included Rockall haddock, and which was done for the period 1997–2008 (Jaworski, 2011). Such retrospective analyses of growth in the context of stock assessment can effectively be conducted with stock-specific historical data (Lorenzen, 2016).

In this analysis, data on catch weights were used that were available from annual assessments of the Rockall haddock stock for the period 1991–2017 (ICES, 2018). Throughout this time-series, the growth rate of Rockall haddock typically decreased with increasing weight and age (Figure 1). The growth rate was calculated in relation to the weight of the same year class in the previous year. With six age groups available in the data (ages 1 to 6), this resulted in the growth rate for six age groups (ages 2 to 7). As it can be seen, within certain age groups, the growth rate was dependent on weight, being lower in year classes with a larger mean weight, which was particularly the case for the youngest (ages 2–3) and oldest ages (ages 6–7). This effect was less pronounced or absent at intermediate ages (at age 4–5). Compared to the earlier study conducted for Rockall haddock (Jaworski, 2011), the pattern was less visible here and the points in the plot were far more scattered. This was mainly due to higher weights observed in the recent period (ICES, 2018). For example, the maximum weight-at-age was earlier less than 1 kg, while in the recent period, it has been much higher (up to 2.6 kg).

The high variation in growth rate observed in the current analysis made the yield prediction less accurate, compared to the previous evaluation. Also, most methods in this evaluation tended to underestimate yield rather than overestimate it, as was the case with the earlier evaluation. This can in part be explained by low catches in the recent period, as opposed to higher catches observed in the earlier study. In this evaluation, there were some methods that were biased (by 9–10%), but relatively precise, for example, "mean of five previous weights" (Method A), "mean of three previous weights" (Method B) and "linear models for year classes" (Method E; for details see Jaworski, 2011) (Figure 2). Three other methods: "previous weight" (Method C), "expected growth with weight, age and year" (Method M) and "expected growth with weight and year class" (Method K) were unbiased, but relatively imprecise in weight forecasts. Methods A, B and E also performed well in terms of absolute errors (for total yield and yield-at-age, Figure 3). In addition, examination of plots of the observed vs. predicted values for the total yield (Figure 4) provided more evidence of Methods A, B and E being good candidates for further consideration. Eventually, Method B was chosen as the optimal option. It can further be seen in Figure 5 that earlier in the time-series, this method gave reliable weight predictions, while in the recent period, its performance worsened, alongside the other methods, as a result of higher weights in the recent period.

It has to be noted that running the retrospective analysis for just 12 years has its advantages as it indicates methods that proved to be sufficiently robust and effective for the recent period. As the next step in this evaluation, the analysis was extended to include more years of retrospection (from 1996–2017) to find the overall effectiveness of the different methods in a longer term, with catch levels varying widely in the past. Similar plots were made to those in Figures 2–4, but this time, for a longer retrospection series and are shown in Figures 6–8. The bias for Method B seen with the shorter retrospective was less marked here (3%, Figure 6). These results gave additional support to choosing Method B as the optimal one.

In the short time forecast for Rockall haddock conducted by the working group in 2018, a 10-year mean was used to predict weight-at-age in the catch (ICES, 2018). The rationale behind this was the high variability in catch weight estimates for recent year classes. In earlier assessments, a 3-year (before 2015) and 5-year (in 2015–2017) means were used. A 10-year mean was also attempted in this evaluation, both with a 12-year and a longer retrospective (from 2001–2017). However, those predictions were far less accurate compared to the 3-year or 5-year-mean predictions.

In conclusion, Method B, a 3-year mean, was found optimal and readily applicable for weight forecasts in future assessments of haddock in Division 27.6b.

Estimation of mean stock weight-at-age

In the recent period, high catch weights have been observed, mainly with respect to landing weights (Figure 9; ICES, 2018). In the assessment, the mean weights-at-age in the stock are basically assumed to be the same as the catch weights. However, in the last assessment, the assumed stock weights were slightly reduced for some recent years (Figure 10, upper panel) as a result of low discards and high discard weights. Those assumed stock weights were used to estimate the spawning–stock biomass.

In this section, the utility of stock weights as observed in the survey is explored to potentially be used in the assessment. There is some information on mean stock weights from the surveys prior to 2011, but it is not clear how they were calculated (most likely, they were weighted mean weights for demersal sampling areas). From the year 2011 on, the abundance index has been calculated based on depth strata (see the working document on the index calculation), although this index has not formally been used yet in the assessment. By analogy to the index calculation, all otoliths and individual weights taken within each of the four strata are combined to form a weight–age–length key (WALK). This WALK is applied to all LFs in the stratum individually to produce age frequencies and mean weights for each haul. Finally, for each stratum, the products of age frequencies and weights are summed and divided by the sum of age frequencies. The resulting mean weights are summed and divided by the number of valid hauls to provide mean weights per haul. This procedure can be summarised as:

$$w_{i,a} = \frac{\sum_{h=1}^{H_i} \left[\left(\sum_{l=l_{\min}}^{l=l_{\max}} N_{i,a,l,h} w_{i,a,l,h} \right) / \sum_{l=l_{\min}}^{l=l_{\max}} N_{i,a,l,h} \right]}{H_i}$$

where $N_{i,a,l,h}$ is the number of fish at age a and length l caught during haul h, $w_{i,a,l,h}$ is the mean weight of fish at age a and length l caught during haul h, H_i is the number of valid hauls in stratum i and $CPUE_{i,a}$ is the catch per unit of effort of fish at age a in stratum i.

For each age, the mean weight (by analogy to the age frequency) for each stratum is raised by the stratum area. These raised weights are then summed and the result divided by the total area in the assessment region. The final weight for each age is given by:

$$w_a = \frac{\sum_{i=1}^{S} w_{i,a} A_i}{\sum_{i=1}^{S} A_i}$$

where A_i = area (m²) of stratum i and S = number of strata.

The estimated mean stock weights from the survey (for 2011–2018) can be seen in Figures 10 and 11 (lower panels). The estimates are quite reliable for ages 0–6 as the produced growth curves for year classes are smooth and they seem to well approximate von Bertalanffy curves. The visible problem is that occasionally, some stock weights were missing in the survey, but missing values can be approximated by von Bertalanffy predictions.

The mean stock weights from the assessment were relatively smooth in the past (Figure 11, upper panels), but in the recent period, they have become more erratic and rather uncertain.

Assuming that stock weights from the survey represent adequately the real weights in the stock, they could potentially be used for catch weight predictions in the short time forecast. The utility of such predictions was evaluated earlier for Icelandic haddock (ICES, 2009; Jaworski, 2011). However, the available time-series of stock weights from the survey is too short (eight years) to run similar retrospective evaluations.

In conclusion, the available information on stock weights from the survey can potentially be used effectively in estimations of spawning–stock biomass and catch predictions.

References

ICES. 2009. Report of the North Western Working Group (NWWG), ICES Headquarters, Copenhagen, 29 April–5 May 2009, ICES CM 2009\ACOM:04.

ICES. 2018. Report of the Working Group on Celtic Seas Ecoregion (WGCSE), 9–18 May 2018, Copenhagen, Denmark. ICES CM 2018/ACOM:13. 1251 pp.

Jaworski, A. 2011. Evaluation of methods for predicting mean weight-at-age: an application in forecasting yield of four haddock (*Melanogrammus aeglefinus*) stocks in the Northeast Atlantic, Fisheries Research, 109, 61–73.

Lorenzen, K. 2016. Toward a new paradigm for growth modeling in fisheries stock assessments: embracing plasticity and its consequences, Fisheries Research, 180, 4–22.

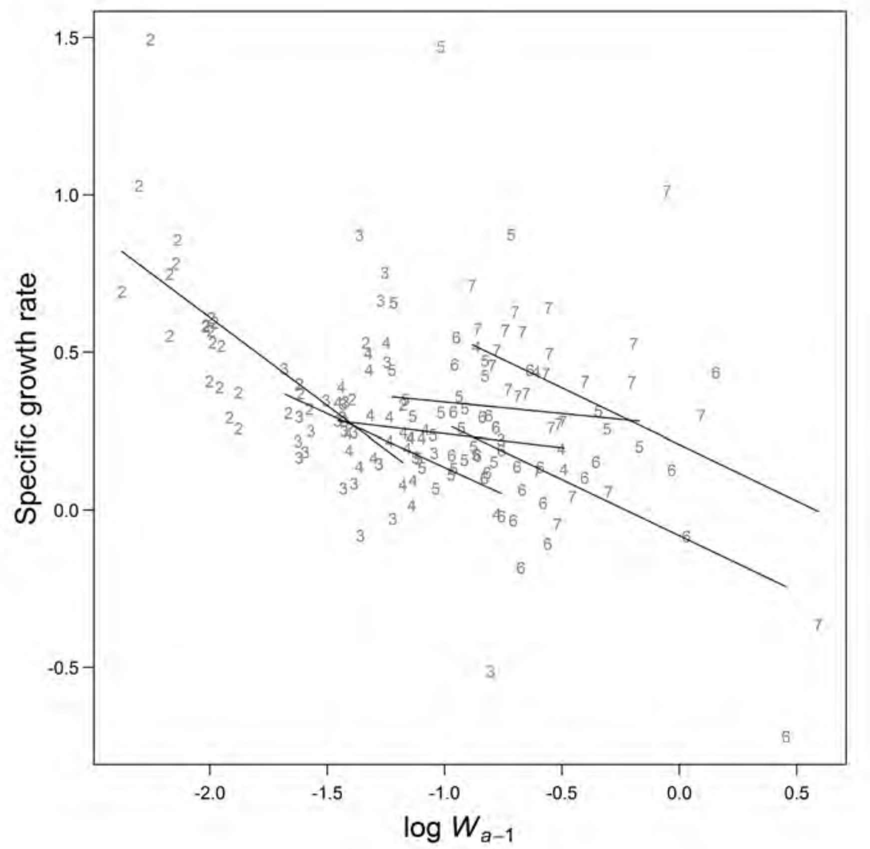


Figure 1. Haddock in Division 27.6b. Specific growth rate as a function of log weight and age in 1991–2017. The lines show fitted values for the model with different slopes and intercepts for ages 2–7 (Jaworski, 2011). Each single line represents the fit for an age group.

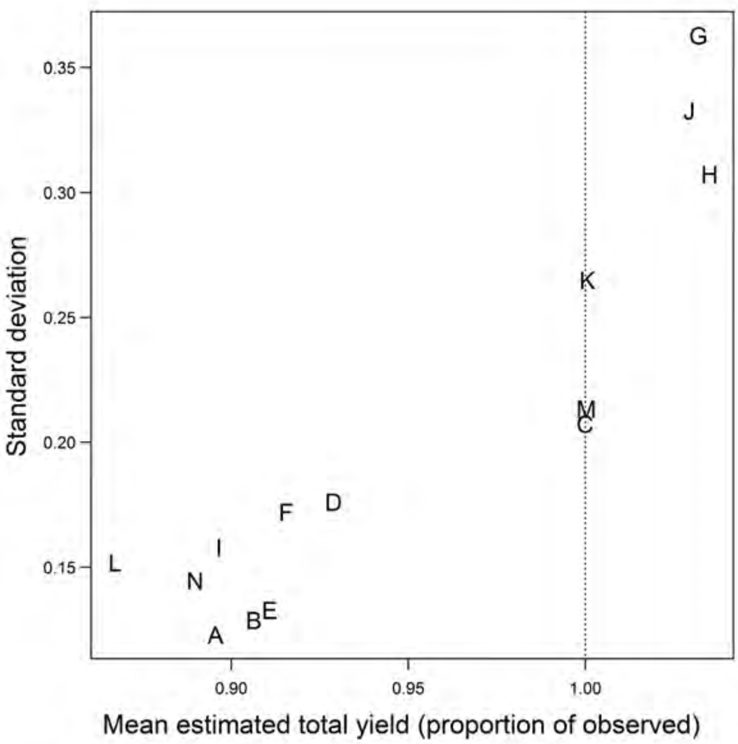


Figure 2. Haddock in Division 27.6b. Mean and standard deviation of the total yield forecasts for the retrospective series 2006–2017, obtained by different methods (see Jaworski, 2011, for details). The forecasts are expressed as the proportion of the observed total yield in a given year. The dotted vertical line shows unbiased forecasts while the 0-value on the *y*-axis (not shown) corresponds to the highest possible (perfect) precision. Method B is currently used in forecasts.

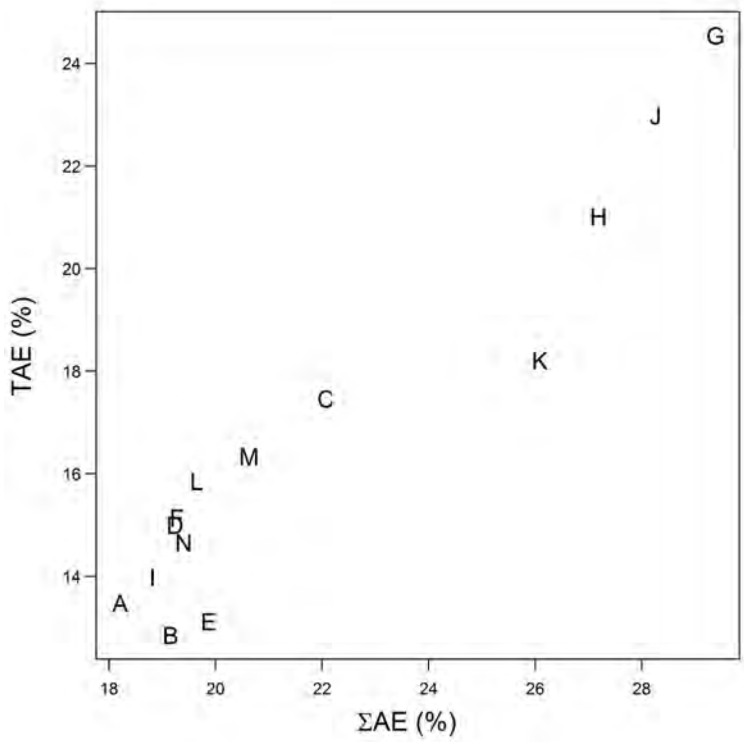


Figure 3. Haddock in Division 27.6b. Sum absolute error (ΣAE) for yield-at-age forecasts and absolute error (TAE) for total yield forecasts for the retrospective series 2006–2017, obtained by different methods. Both errors are expressed as % of the observed total yield in a given year. Mean errors for this retrospective series are shown here. Method B is currently used in forecasts.

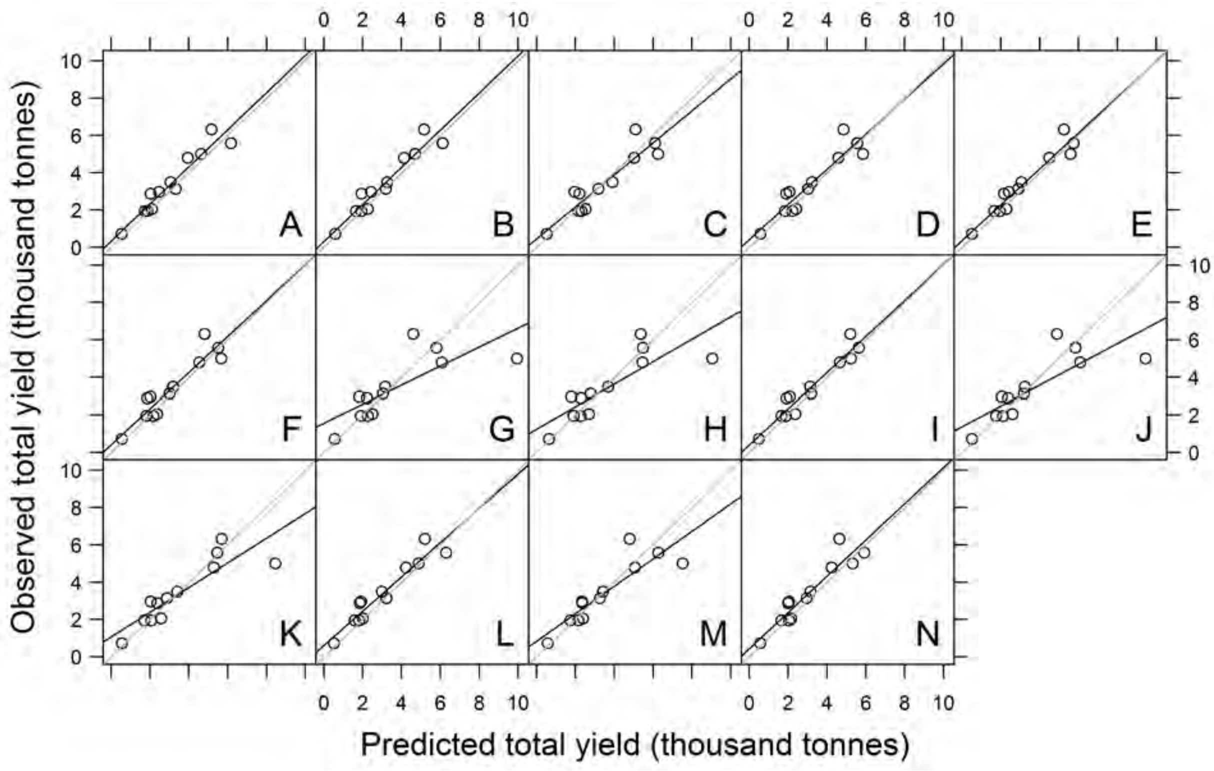


Figure 4. Haddock in Division 27.6b. Predicted vs. observed plots of the total yield for the retrospective series 2006–2017 by using different methods. The fitted linear regression line is shown (black line) with the line $y = \hat{y}$ of the "perfect fit" (grey line) added for reference. Method B is currently used in forecasts.

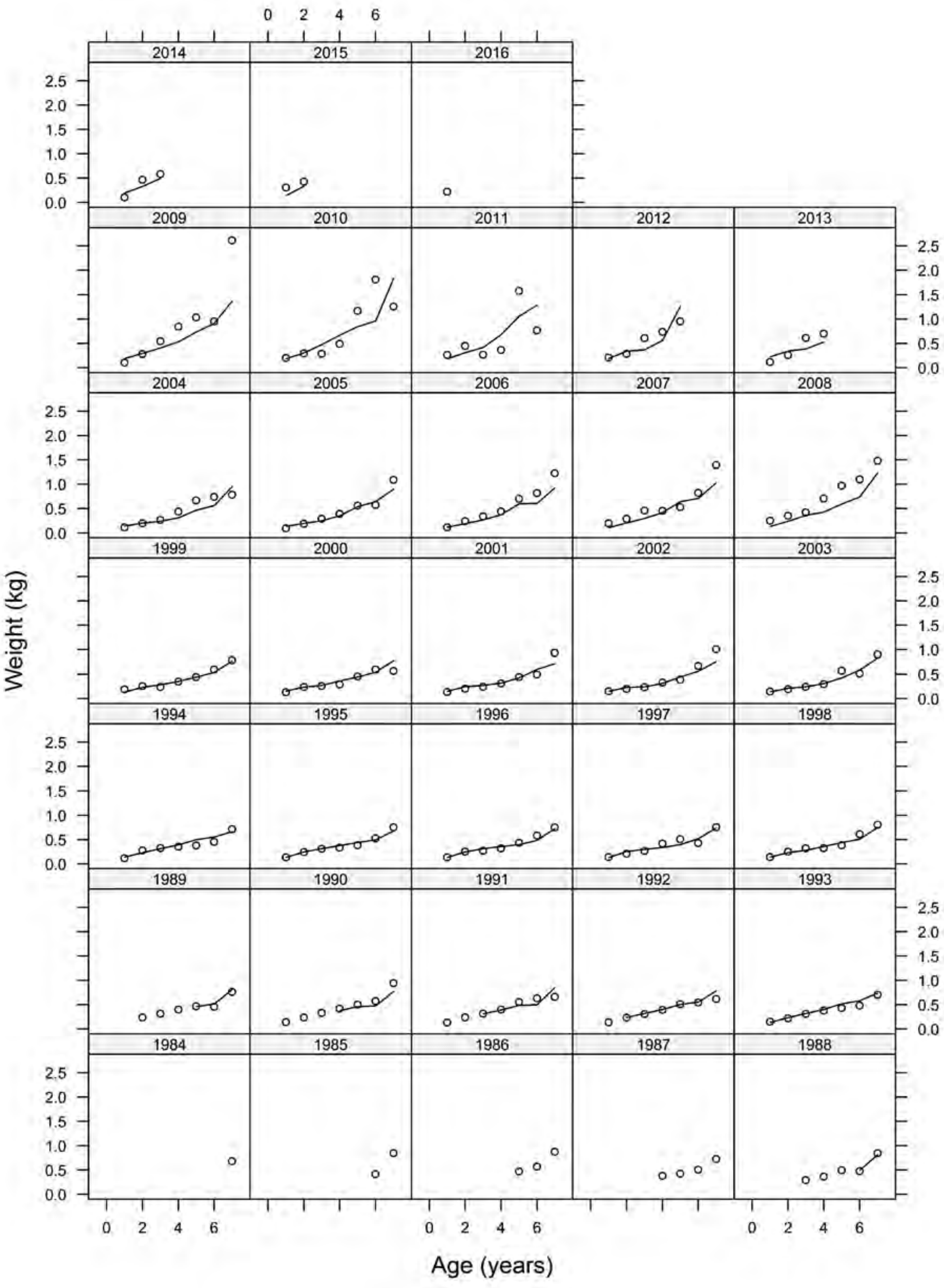


Figure 5. Haddock in Division 27.6b. Observed (circles) and estimated (by Method B, solid line) mean weights-at-age in the catch tracked by year class.

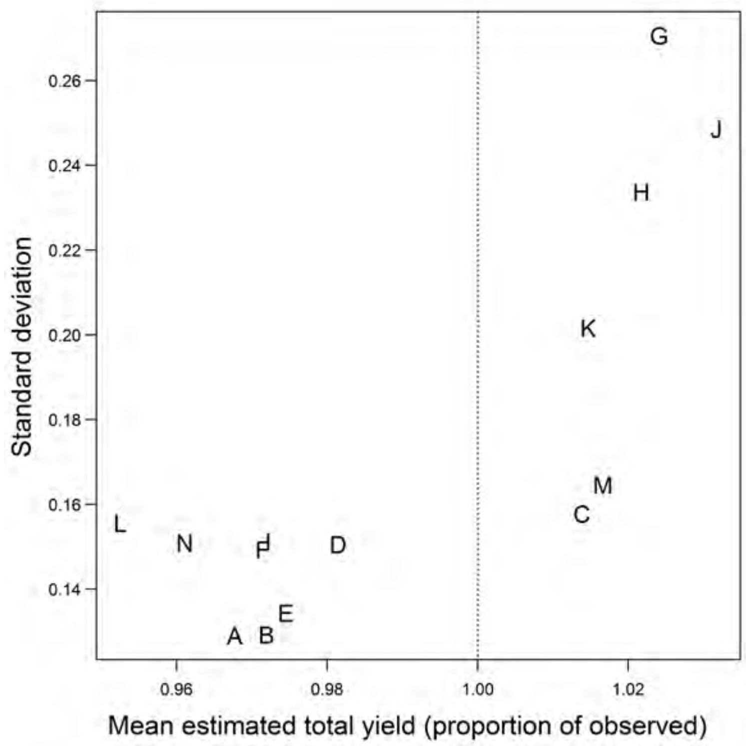


Figure 6. Haddock in Division 27.6b. Mean and standard deviation of the total yield forecasts for the retrospective series 1996–2017 (see Figure 2).

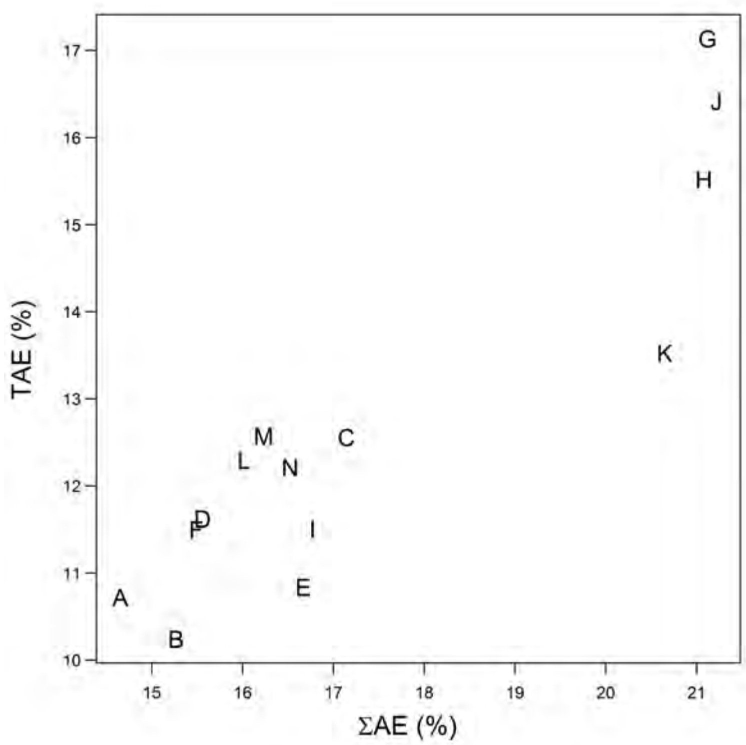


Figure 7. Haddock in Division 27.6b. Sum absolute error (ΣAE) for yield-at-age forecasts and absolute error (TAE) for total yield forecasts for the retrospective series 1996–2017 (see Figure 3).

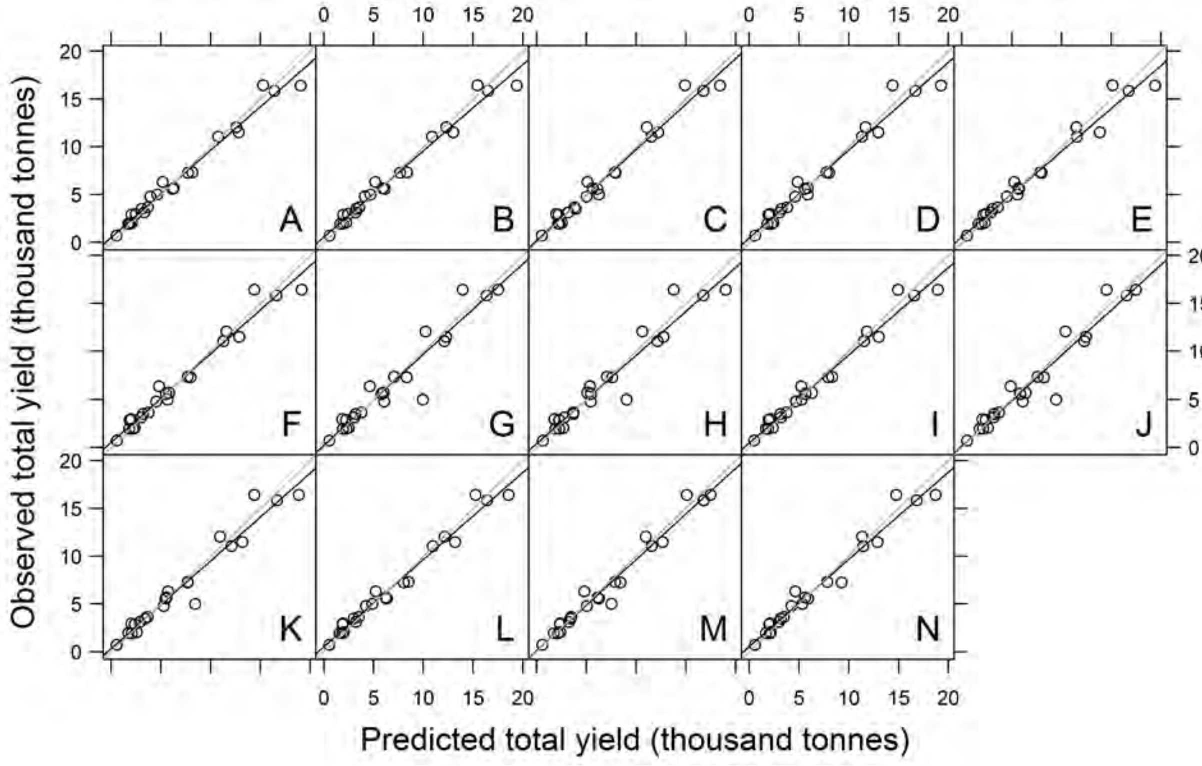


Figure 8. Haddock in Division 27.6b. Predicted vs. observed plots of the total yield for the retrospective series 1996–2017 by using different methods (see Figure 4).

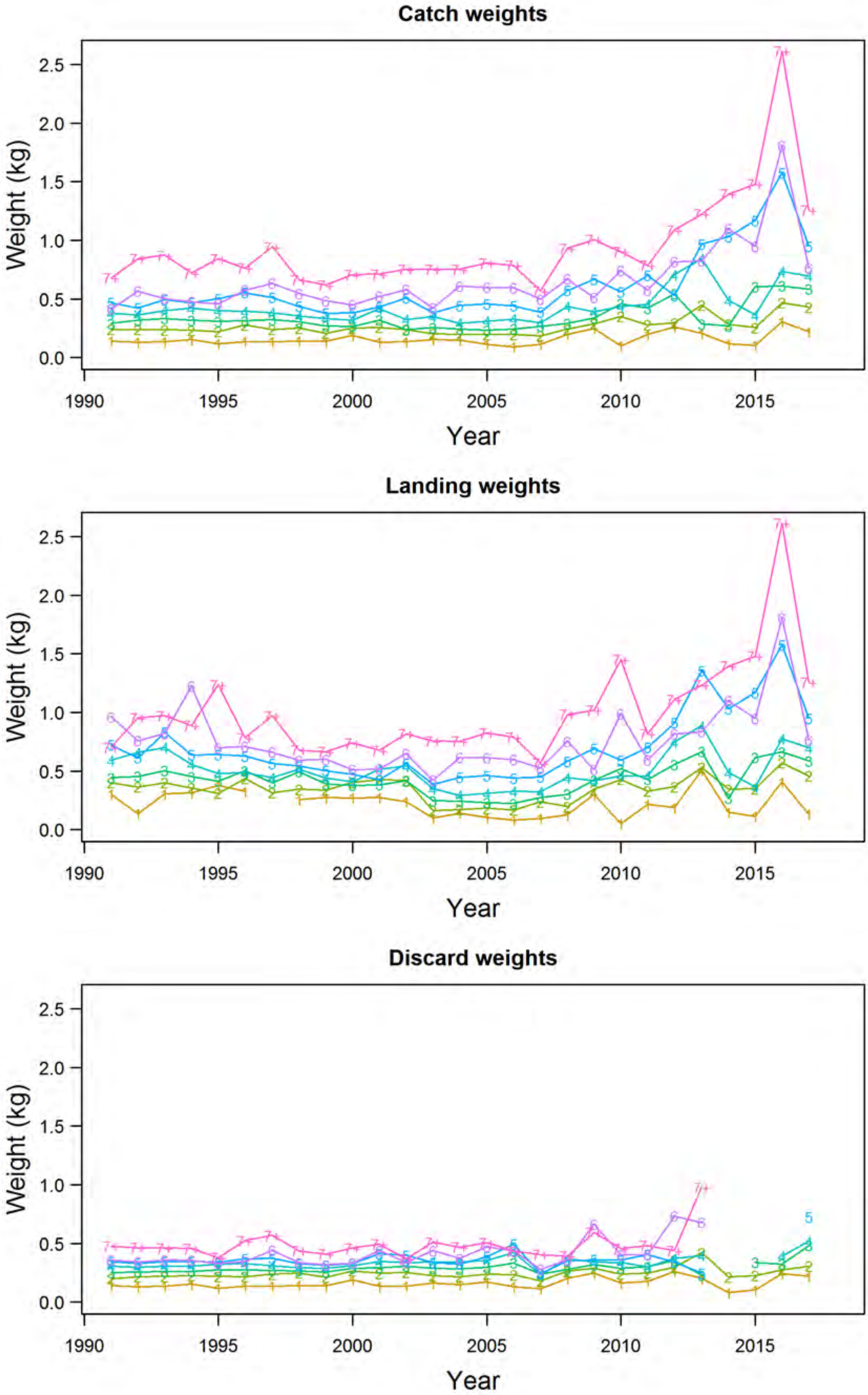


Figure 9. Haddock in Division 27.6b. Catch, landing and discard weights (ICES, 2018).

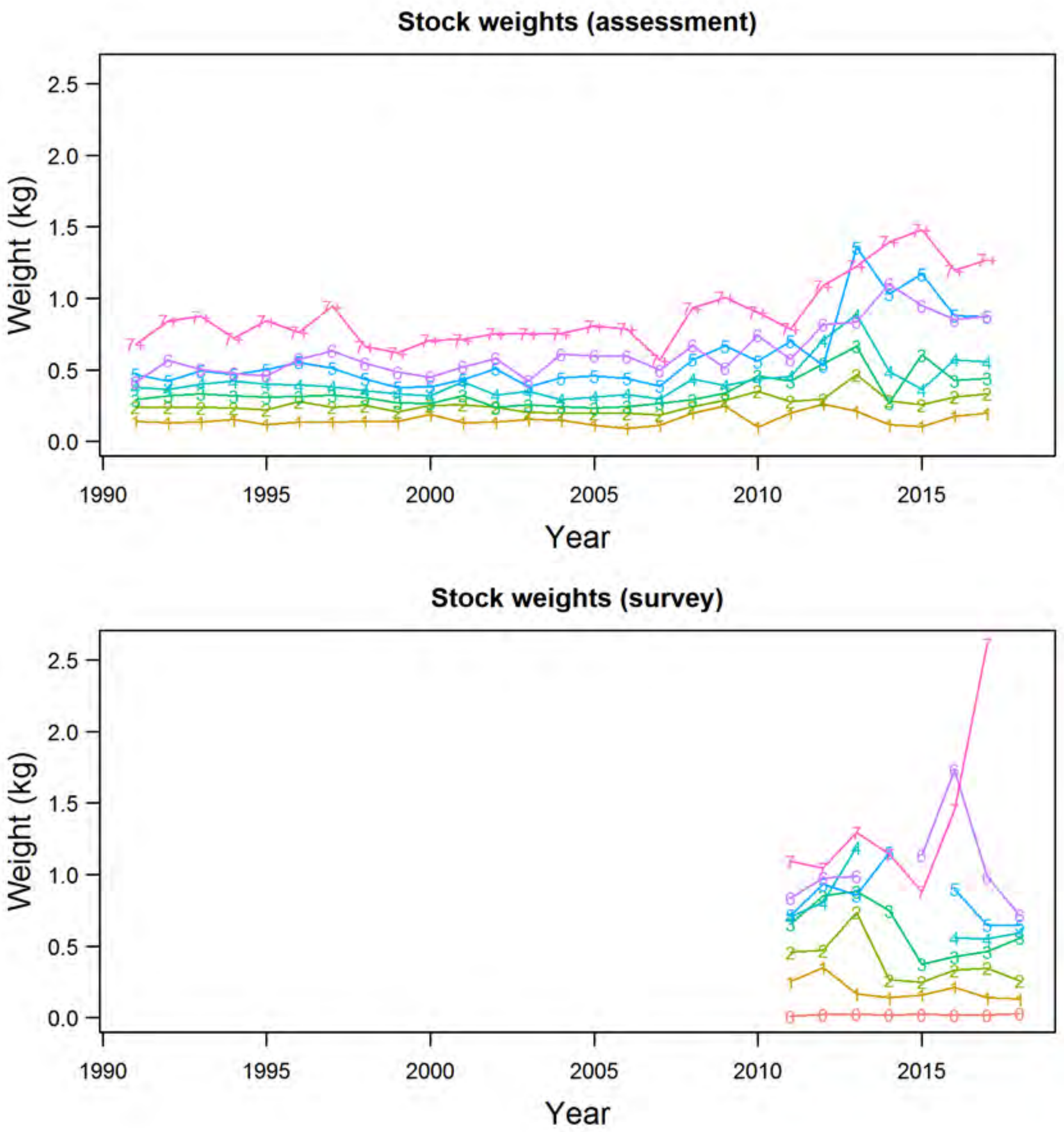


Figure 10. Haddock in Division 27.6b. Stock weights by year as presented in the assessment (upper panel, ICES, 2018) and estimates from the survey for the years 2011–2018 (lower panel).

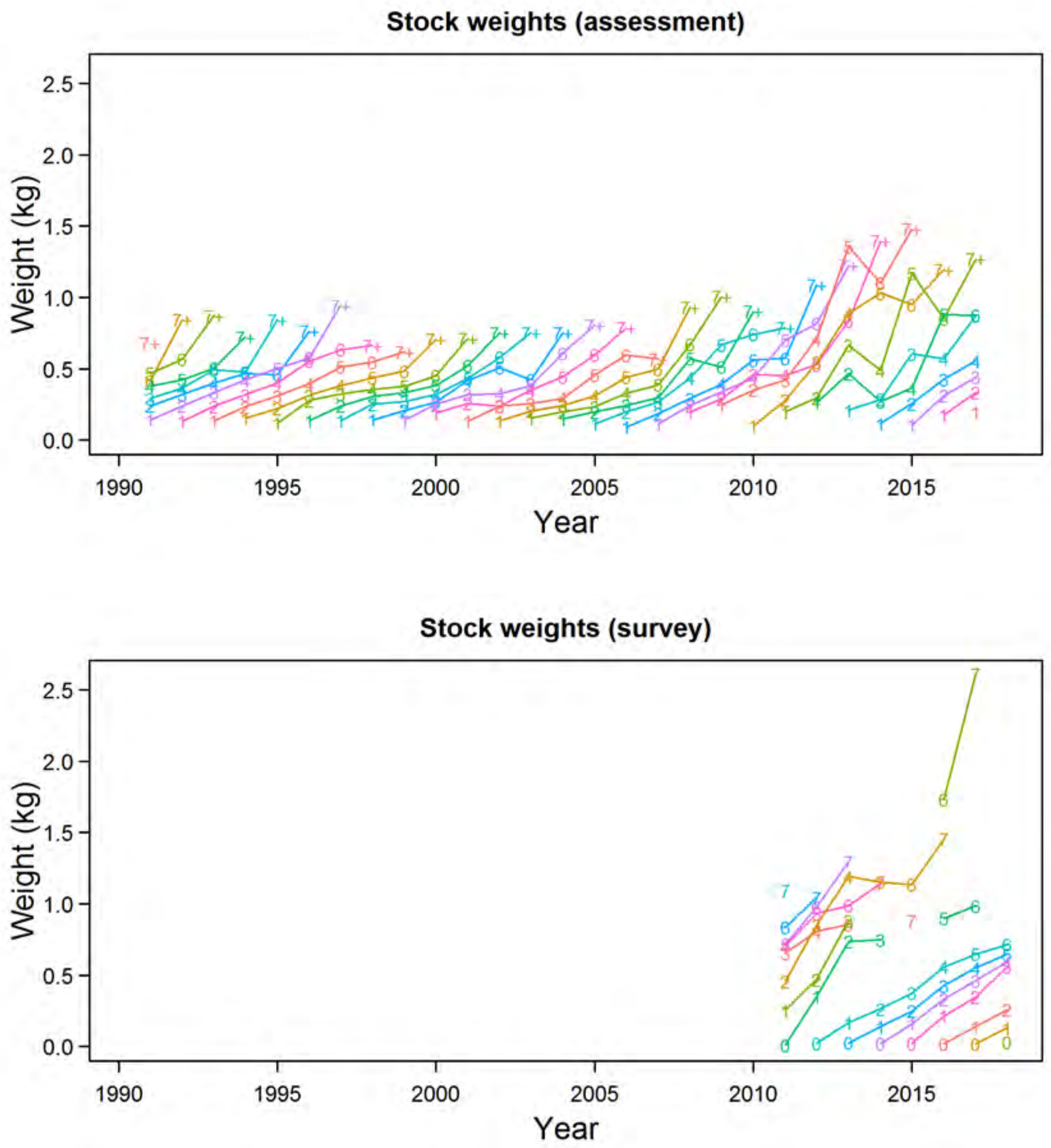
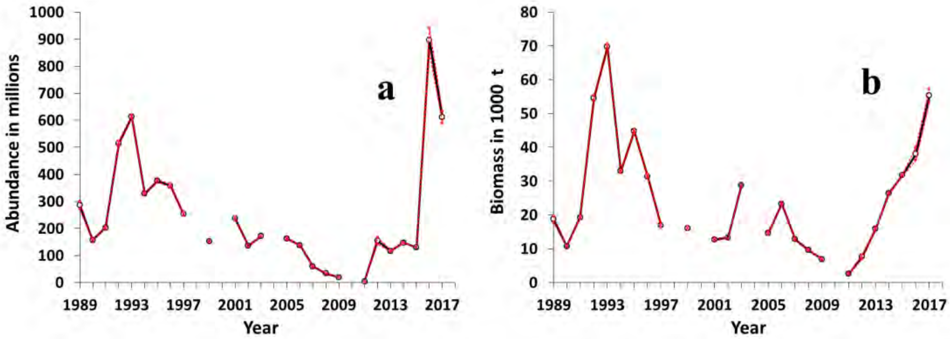


Figure 11. Haddock in Division 27.6b. Stock weights by year class as in the assessment (upper panel, ICES, 2018) and estimates from the survey for the years 2011–2018 (lower panel).

Annex 3: Supplementary Figures

Exploration of Survey Data



FigureA.1. Abundance (a) and biomass (b) of haddock, estimated using the swept-area method with geographical stratification based on rectangles of 15′ latitude and 15′ longitude by RV 'Scotia' survey. Red dashed line indicates the confidence interval with 0.95 reliability level (Index S2).

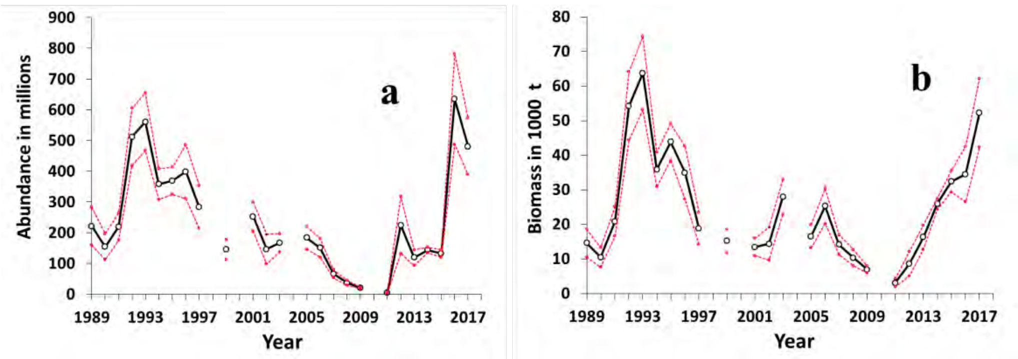


Figure A.2. Abundance (a) and biomass (b) of haddock, estimated using the swept-area method with geographical stratification based on bathymetry by RV 'Scotia' survey. Red dashed line indicates the confidence interval with 0.95 reliability level (Index S3).

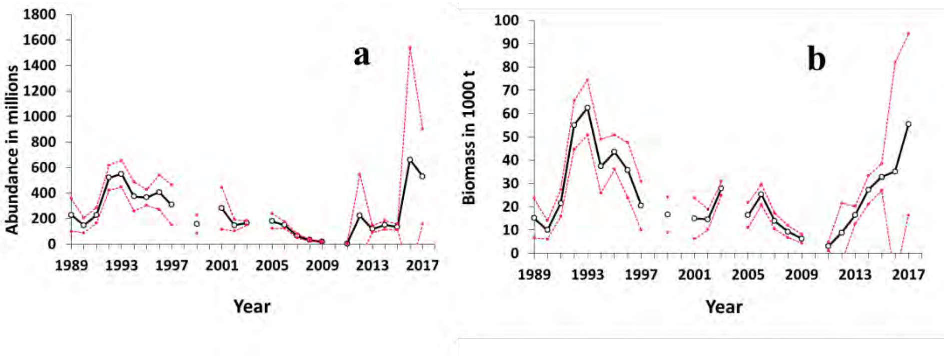


Figure A.3. Abundance (a) and biomass (b) of haddock, assessed using the swept-area method without geographical stratification by RV 'Scotia' survey. Red dashed line indicates the confidence interval with 0.95 reliability level (Index S4).

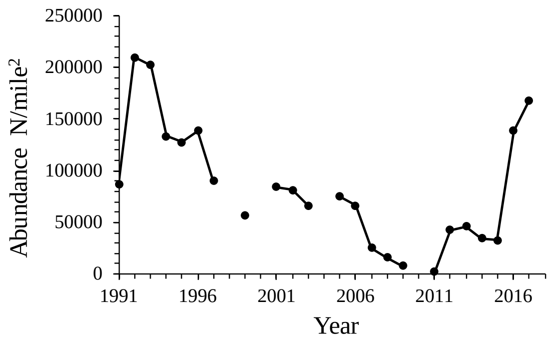


Figure A.4. Abundance Index (N/mile²), estimated using the swept-area method with geographical stratification based on rectangles of 15′ latitude and 15′ longitude by RV 'Scotia' survey. (Index S2).

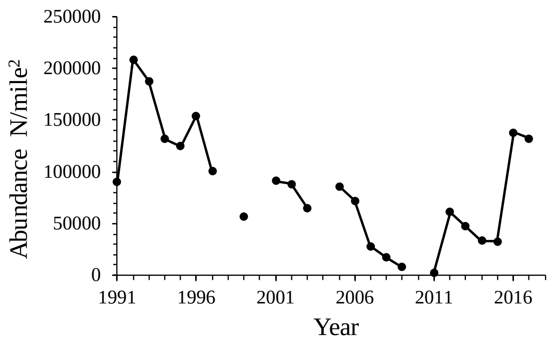


Figure A.5. Abundance Index (N/mile²), estimated using the swept-area method with stratification based on bathymetry by RV 'Scotia' survey. (Index S3).

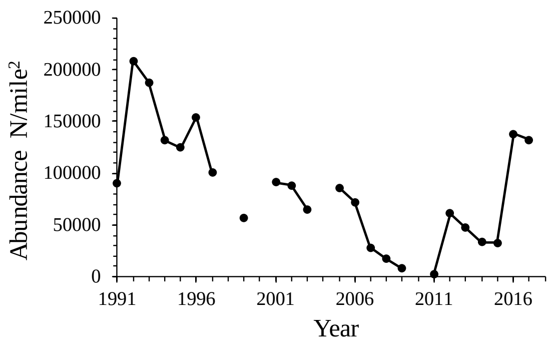


Figure A.6. Abundance Index (N/mile²), estimated using the swept-area method without geographical stratification by RV 'Scotia' survey. (Index S4).

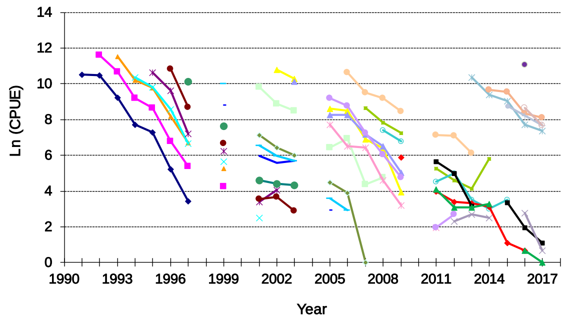


Figure A.7. Log survey cpue by year class. Survey Index **(N/mile^2), estimated using the swept-area method with geographical stratification based on rectangles of 15′ latitude and 15′ longitude by RV 'Scotia' survey. (Index S2).**

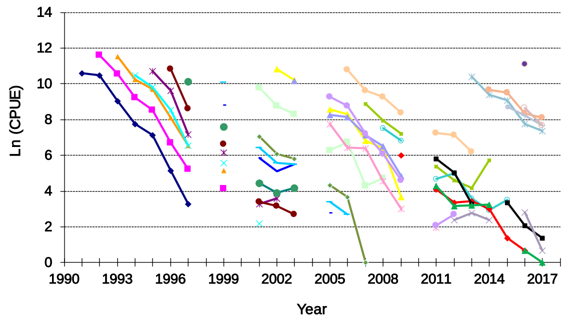


Figure A.8. Log survey cpue by year class. Survey Index **(N/mile^2), estimated using the swept-area method with stratification based on bathymetry by RV 'Scotia' survey. (Index S3).**

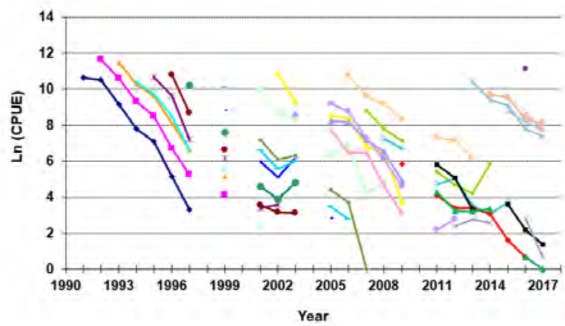


Figure A.9. Haddock in Division 6.b. Log survey cpue by year class. Survey Index **(N/mile^2), estimated using the swept-area method without geographical stratification by RV 'Scotia' survey. (Index S4).**

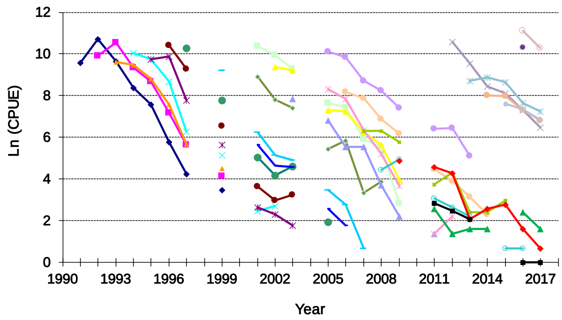


Figure A.10. Haddock in Division 6.b. Log survey cpue by year class. WGCSE 2018 Survey Index (N/10 hours), assessed by RV 'Scotia' survey. Correct ALK for ages 4–5 in 2015. (Index S1).

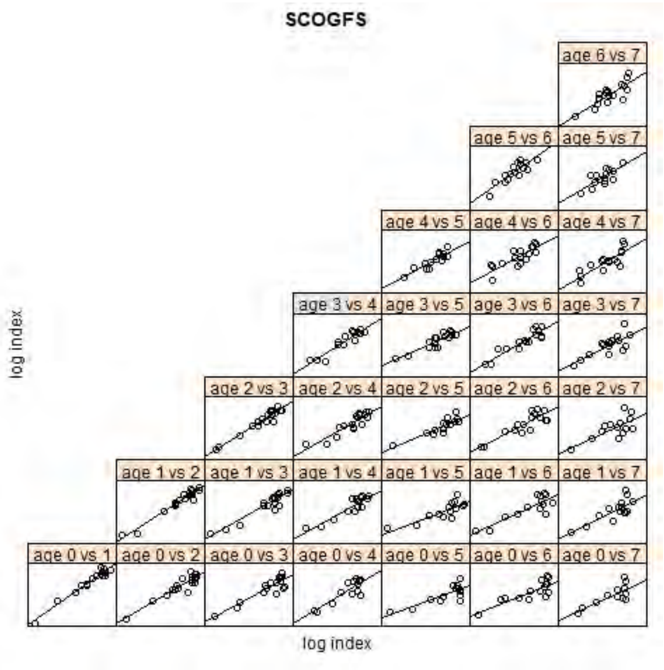


Figure A.11. Haddock 6b. Within survey correlations between consecutive ages of the same cohort. Survey Index (N/mile²), estimated using the swept-area method with geographical stratification based on rectangles of 15′ latitude and 15′ longitude. (Index S2).

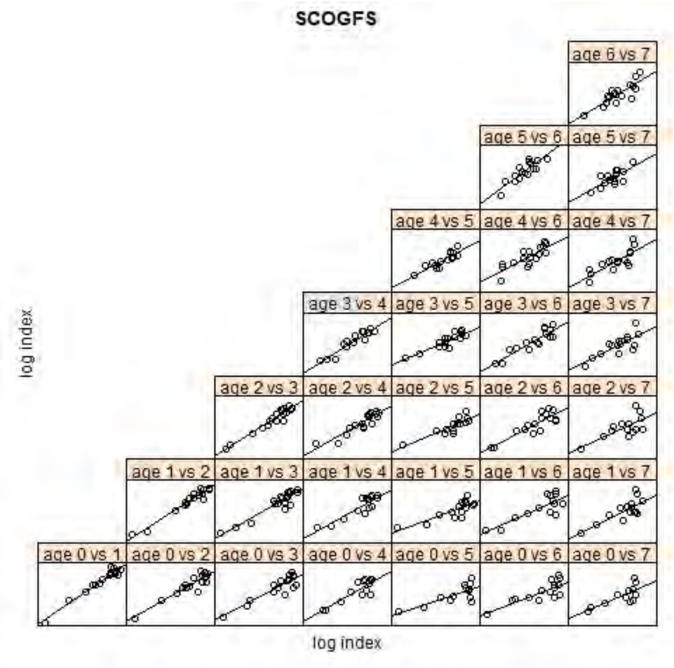


Figure A.12. Haddock in Division 6.b. Within survey correlations between consecutive ages of the same cohort. Survey Index **(N/mile²), estimated using the swept-area method with stratification based on bathymetry. (Index S3).**

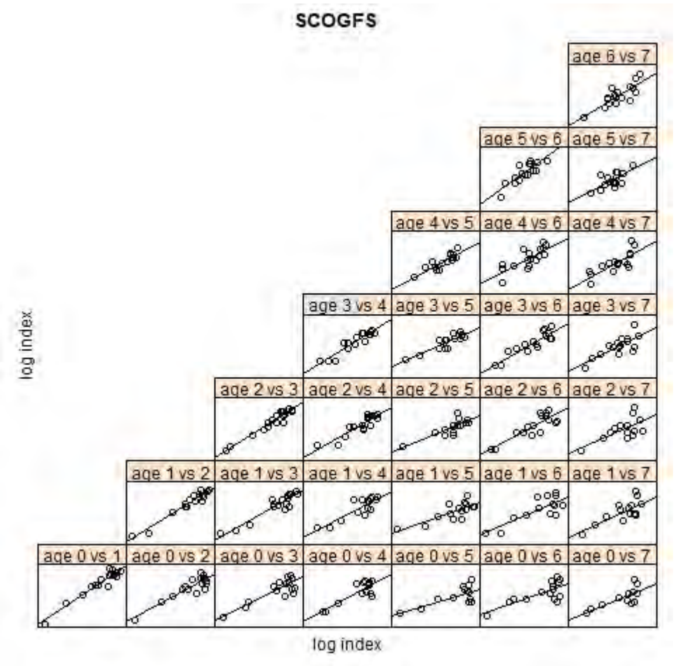


Figure A.13. Haddock in Division 6.b. Within survey correlations between consecutive ages of the same cohort. Survey Index (N/mile²), estimated using the swept-area method without geographical stratification. (Index S4).

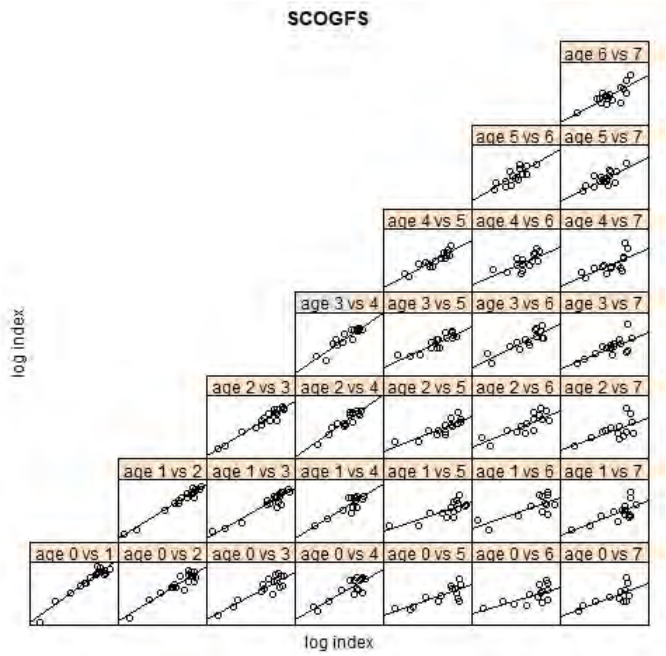


Figure A.14. Haddock in Division 6.b. Within survey correlations between consecutive ages of the same cohort. WGCSE 2018 Survey Index (N/10 hours), assessed by RV 'Scotia' survey. Correct ALK for ages 4–5 in 2015. (Index S1).

Baseline assessment

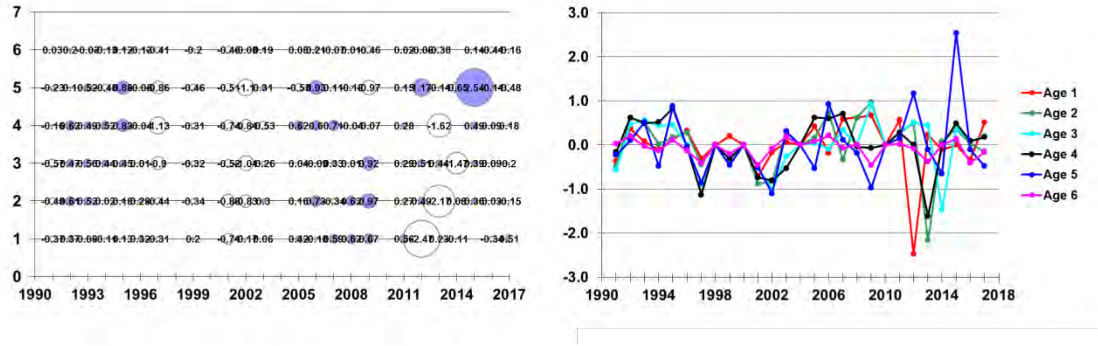


Figure 15. Haddock in Division 6.b. Baseline run. Log catchability residual plots. Final XSA settings and survey data from 2018.

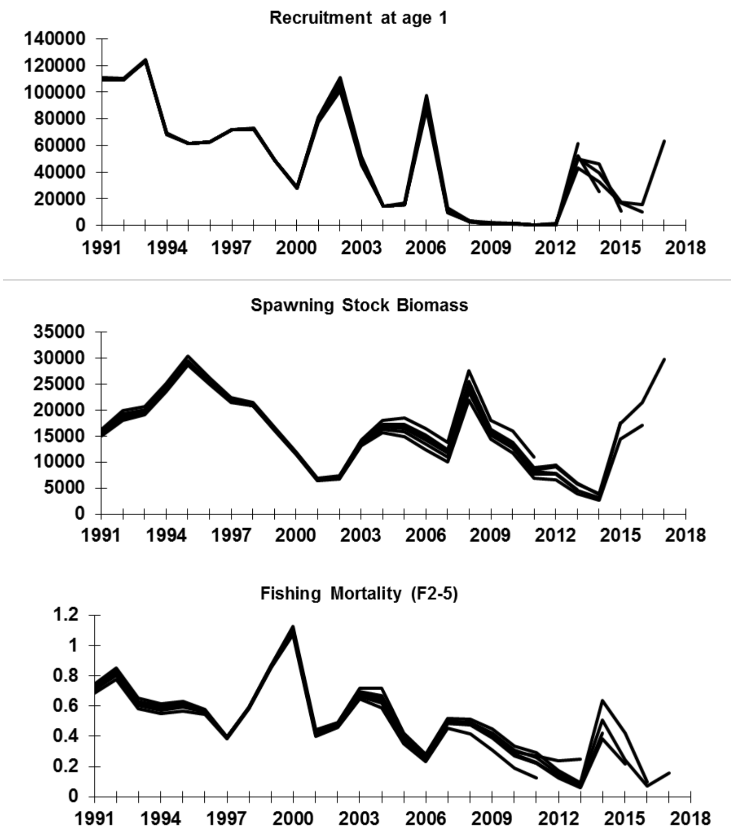


Figure 17. Haddock in Division 6.b. Baseline run. Retrospective analyses. Final XSA settings and survey data from 2018.

Assessment Run S2

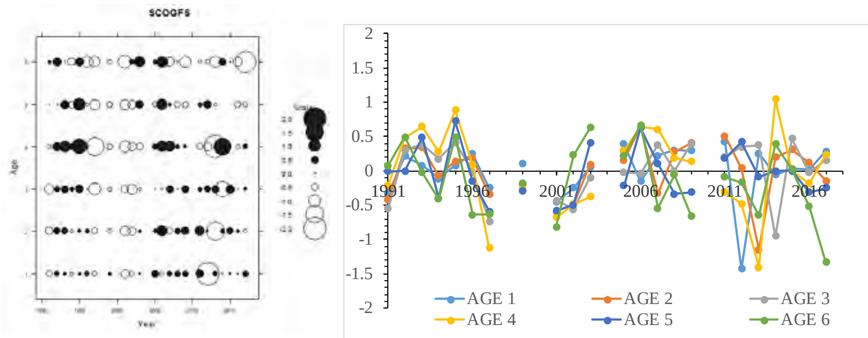


Figure A.15. Haddock in Division 6.b. Run S2. Log catchability residual plots. Age plus group 7+. Survey Index (N/mile²), estimated using the swept-area method with geographical stratification based on rectangles of 15′ latitude and 15′ longitude. (Index S2).

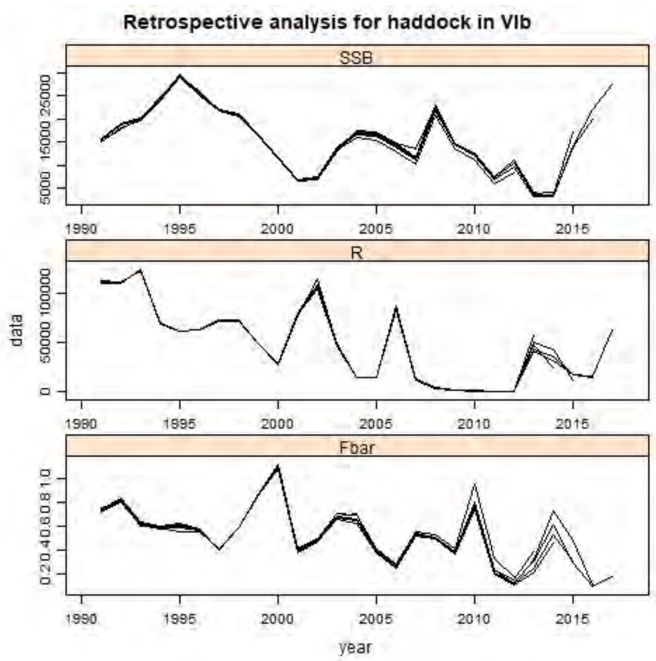


Figure A.16. Haddock in Division 6.b. Run S2. Retrospective analyses. Age plus group 7+. Survey Index (N/mile²), estimated using the swept-area method with geographical stratification based on rectangles of 15′ latitude and 15′ longitude. (Index S2).

Assessment Run S3

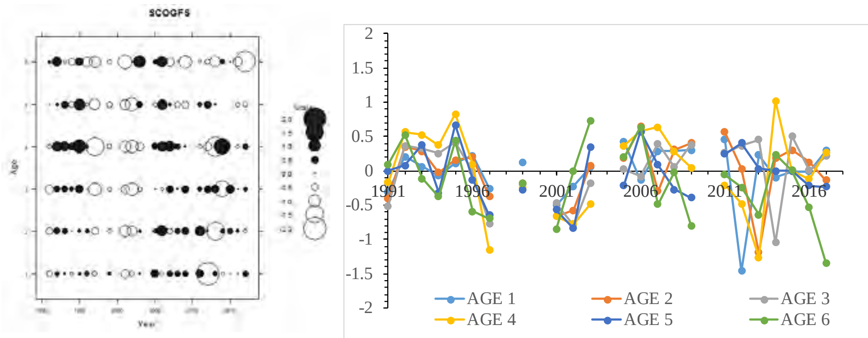


Figure A.17. Haddock in Division 6.b. Run S3. Log catchability residual plots (shrinkage 1.0, catchability dependent on stock size-at-ages <4). Age plus group 7+. Survey Index (N/mile²), estimated using the swept-area method with stratification based on bathymetry by RV 'Scotia' survey. (Index S3).

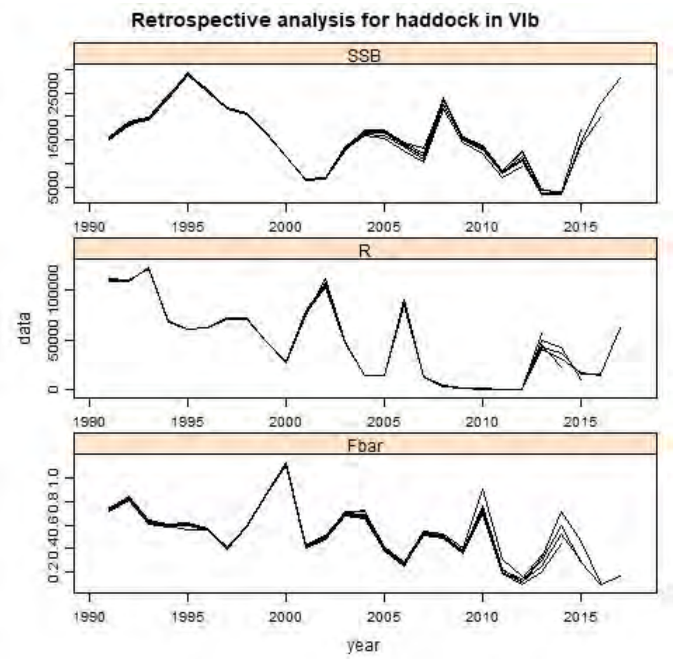


Figure A.18. Haddock in Division 6.b. Run S3. Retrospective analyses. Age plus group 7+. Survey Index **(N/mile²), estimated using the swept-area method with stratification based on bathymetry by RV 'Scotia' survey. (Index S3).**

Assessment Run S4

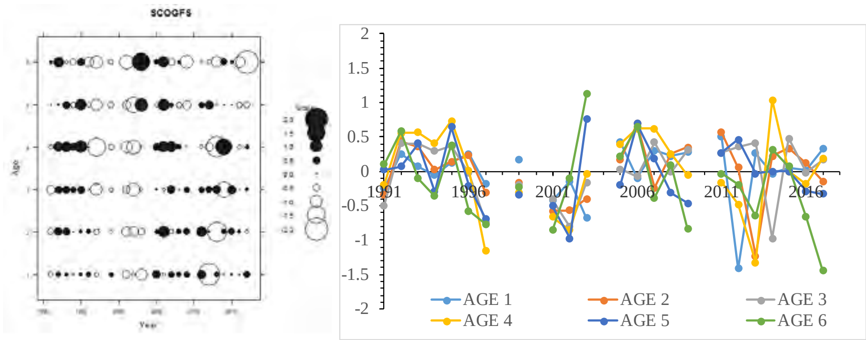


Figure A.19. Haddock in Division 6.b. Run S4. Log catchability residual plots. Age plus group 7+. Survey Index (N/mile²), estimated using the swept-area method without geographical stratification. (Index S4).

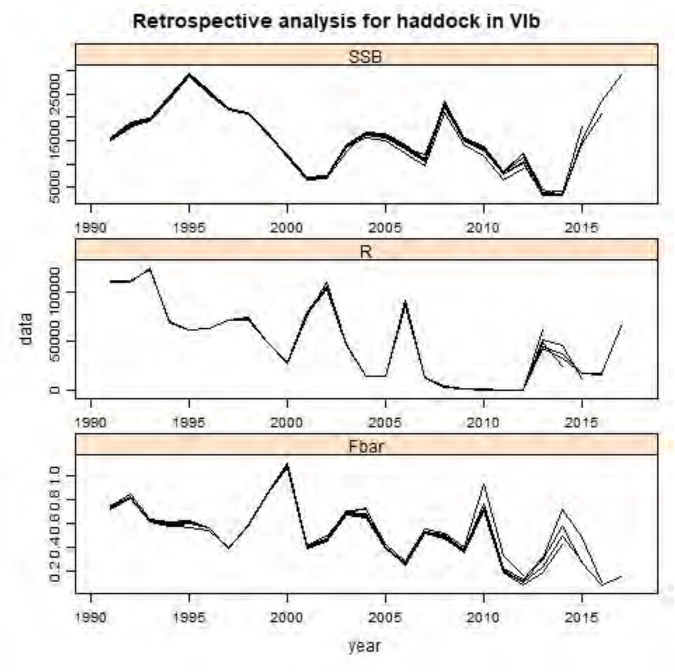


Figure A.20. Haddock in Division 6.b. Run S4. Retrospective analyses. Age plus group 7+. Survey Index **(N/mile²), estimated using the swept-area method without geographical stratification. (Index S4).**

Assessment Run P2

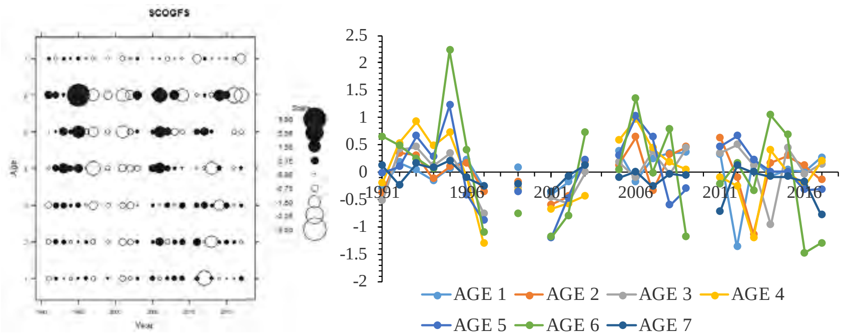


Figure A.21. Haddock in Division 6.b. Run P2. Log catchability residual plots. Age plus group 8+. Survey Index (N/mile²), estimated using the swept-area method with geographical stratification based on rectangles of 15′ latitude and 15′ longitude. (Index S2).

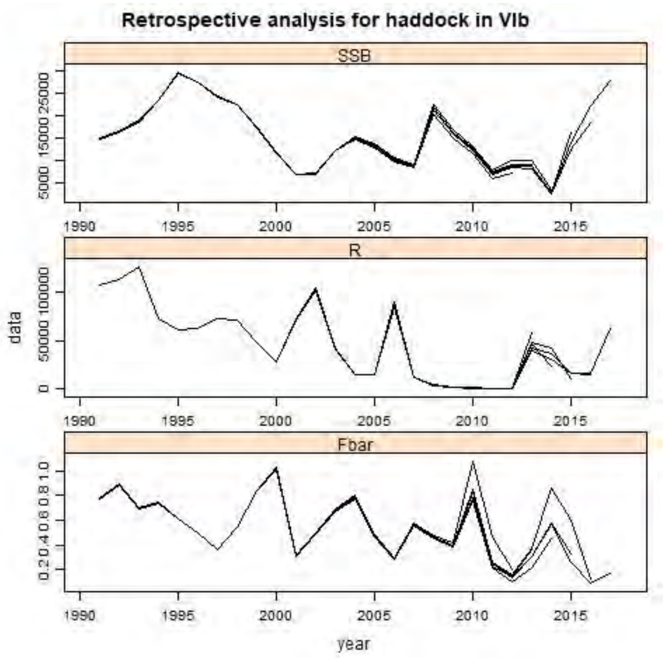


Figure A.22. Haddock in Division 6.b. Run P2. Retrospective analyses. Age plus group 8+. Survey Index (N/mile²), estimated using the swept-area method with geographical stratification based on rectangles of 15′ latitude and 15′ longitude. (Index S2).

Assessment Run P3

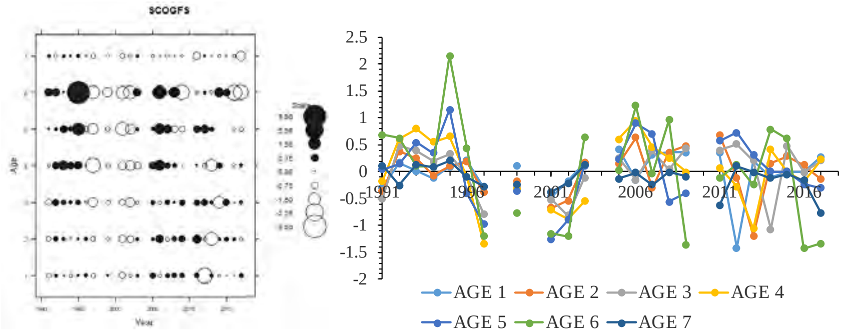


Figure A.23. Haddock in Division 6.b. Run P3. Log catchability residual plots. Age plus group 8+. Survey Index (N/mile²), estimated using the swept-area method with stratification based on bathymetry by RV 'Scotia' survey. (Index S3).

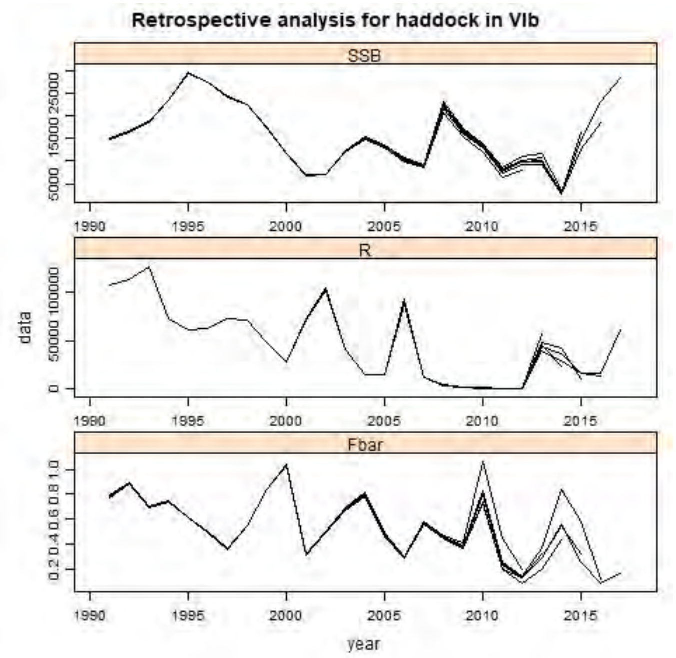


Figure A.24. Haddock in Division 6.b. Run P3. Retrospective analyses. Age plus group 8+. Survey Index (N/mile²), estimated using the swept-area method with stratification based on bathymetry by RV 'Scotia' survey. (Index S3).

Assessment Run P4

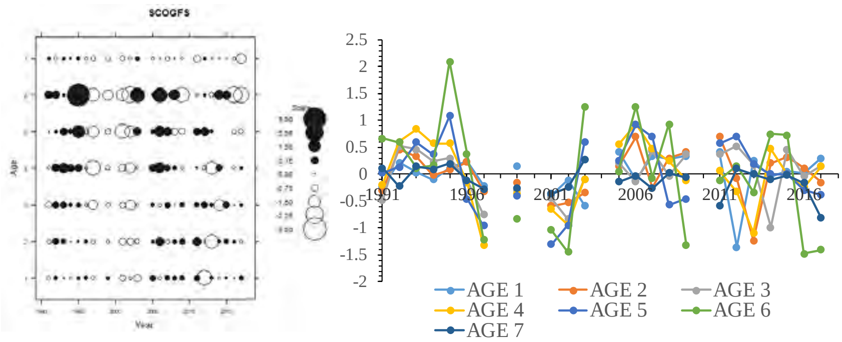


Figure A.25. Haddock in Division 6.b. Run P4. Log catchability residual plots. Age plus group 8+. Survey Index (N/mile²), estimated using the swept-area method without geographical stratification. (Index S4).

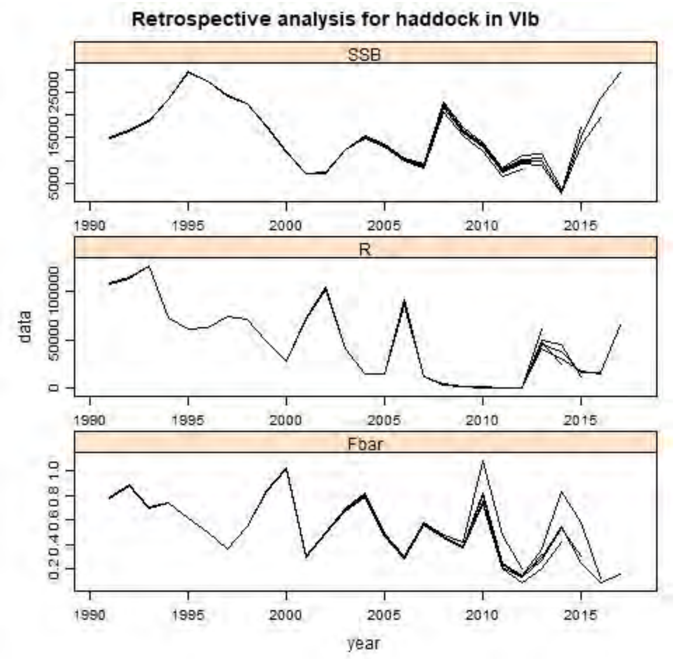


Figure A.26. Haddock in Division 6.b. Run P4. Retrospective analyses. Age plus group 8+. Survey Index (N/mile²), estimated using the swept-area method without geographical stratification. (Index S4).

Assessment Run P5

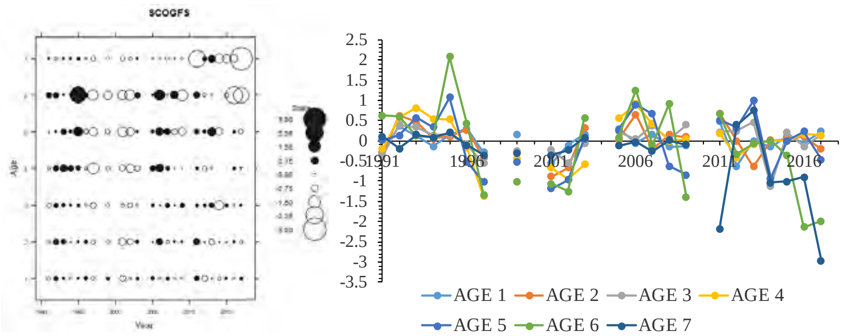


Figure A.27. Haddock in Division 6.b. Log catchability residual plots (shrinkage 1.0, catchability dependent on stock size at-ages <4). Age plus group 8+. Two survey indices (N/10 hours): i) 'standard' index 1991–2009, and ii) new index 2011 onwards.

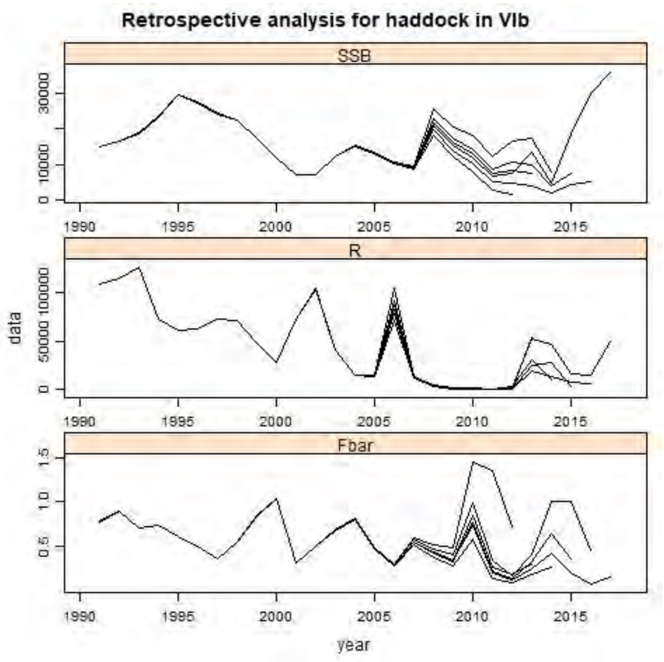


Figure A.28. Haddock in Division 6.b. Retrospective analyses (F shrinkage 1.0). Age plus group 8+. Two survey indices (N/10 hours): i) 'standard' index 1991–2009, and ii) new index 2011 onwards.

Annex 4: External Reviewer Report

External reviewer report

WKROCK turned out to be a challenging benchmark workshop due to insufficient preparation of data inputs, model runs and thorough documentation of the assessment process before the workshop. While a working document on survey design and indices and another one on different ways to predict mean weight-at-age for the short-term projections/forecast were available before the workshop, catch-at-age raising in InterCatch had to be done during the benchmark workshop after doubts appeared regarding the raising procedure used so far. This stresses the importance of carrying out a complete data compilation workshop before the benchmark workshop itself like is usually done for other ICES benchmarks. No test runs with alternative input had been prepared. Because of the limited preparation status, only little could be achieved and work focused more on a review of the current assessment rather than on advancing the assessment and subjecting it to a complete review of data inputs, modelling assumptions and an examination of the advice consequences of alternate model runs.

Following topics were discussed during the benchmark workshop

Catch-at-age

For recent years it was unclear what method has been used to raise wanted and unwanted catch. Data were uploaded to InterCatch but the extraction of data in a format suitable to be input into the model was not finalized. Therefore, data from 2012 onwards were reprocessed in InterCatch to ensure a proper raising of wanted and unwanted catch. This is a step forward to a more transparent process of creating input data to the assessment, but it also means that a significant amount of time was spent on this task and not on other important aspects of the review.

Mean weights in the catch

Mean weight-at-age is especially uncertain for older age groups of small cohorts. An outlier is obvious for 2016 where the mean weight for ages 5+ show a serious jump upwards. Only very few fish of ages 5+ were sampled during this year. This outlier has important consequences on the assessment and it is important to ensure that it does not get a high weight in short-term forecasts for the next years. It is unclear whether it is a sampling artefact or the consequence of very low year classes that experienced fast growth through density-dependent mechanisms. In general, there is a trend towards higher mean weight-at-age and higher interannual variability in recent years. Because of this and because samples collected during the survey do not show such pronounced variations in weight-at-age between years, it was concluded to use smoothed stock weights-at-age (five years moving average?) instead of the raw catch-at-age values.

Different methods to estimate mean weight-at-age for the short-term projections were tested in a retrospective analysis (WD 2). This analysis shows that, in recent years, a three or five years average shows some bias but less variation around the "truth". Taking only the last available year gives an unbiased estimate but much more variation around the "truth". When analysing the full time-series from 1991 onwards, the bias for the three and five-year average is much lower. The working group considered that a three year average would be the best option provided that the 2016 outlier be either excluded or down-weighted in the short-term forecast.

Survey index

The survey index for the Scottish trawl survey has been reanalysed. The survey design changed from a fixed station design to a random stratified survey in 2011. The depth range covered also increased after 2011 (from 250 to 350 m). So far, only the stations inside the original grid and depth range are used for the years after 2011 and only one time-series from 1991 to 2017 is provided. During the benchmark an error in the 2015 ALK used for the survey has been corrected.

Seven versions of tuning indices were tested and the diagnostics of the model runs were analysed:

Runs 1 and 2: Time-series used so far from 1991 to 2017 (baseline) with and without the corrected ALK for 2015;

Runs 3 to 5: Three versions of swept-area indices with different stratifications (five strata based on bathymetry, geographic strata of 15' latitude wide and 15' longitude long and no stratification);

Run 6: Split the survey index time-series in 2011 and use all stations conducted after 2011 including the stations from deeper areas;

Run 7: Only use the new time-series after 2011.

The working group concluded that an assessment with the currently used survey time-series and the corrected ALK is the best option for now. Results were similar for all versions tested, but the retrospective pattern was considerably worse in the runs using the new time-series from 2011 onwards (Runs 6 and 7). The likely reason is that the time-series is still not long enough to produce robust assessment results. The three swept-area indices could have been also an option. Due to the limited preparatory work carried out before the workshop, there was however, not enough time left during the workshop for a proper analysis and a review of the estimation process behind the indices calculations.

Plus group

Age 7 has been treated as plus group thus far. There is a potential to increase the plus group to higher ages. This would help to follow strong and weak cohorts over more years. However, the sampling level deteriorates at higher age groups and age-reading problems become more apparent. Therefore, only an assessment with age 8 as plus group has been considered as a plausible alternative.

A test assessment was thus conducted with age 8+. XSA results were similar for both options (plus group set at age 7 or at age 8) and the retrospective pattern was slightly better for the 7+ run. The sampling level deteriorates above age 7. In addition, apart from two years (where only stronger cohorts are left in the plus group and all other cohorts are very weak) the proportion of catch numbers in the 7+ group is rather small (<10%). Therefore, the working group concluded that the assessment with a plus group set at age 7 is still appropriate. It is worth noting that the residuals were smaller for the highest age used in the survey regardless whether a 7+ or 8+ group was used. This indicates that the model closely follows the survey index of the oldest age in the tuning data (age 6 or 7 dependent on the plus group). It was not possible for the WG, due to time constraints, to analyse/examine this feature but further investigations could be carried out until the next benchmark or during an inter-benchmark (see also recommendations).

FLXSA

The assessment was conducted using the XSA software from Lowestoft. The benchmark group felt that, while the software robustness and appropriateness were not in doubt, relying on legacy software not being developed anymore was a hindrance to further development of the assessment and to achieving a higher level of transparency and reproducibility. As such, effort was made during the benchmark to replicate the XSA assessment run with Lowestoft software in FLXSA. Although some more details have to be adapted in the code, the results of the Lowestoft XSA and FLXSA were found to be reasonably similar to each other. The FLXSA assessment has been included in the ICES TAF framework, therefore enabling a more transparent and collaborative assessment in future years. Furthermore, the FLXSA implementation will facilitate the planned MSE simulations.

Final conclusions

The external reviewers agree with the conclusions of the benchmark and consider that the final benchmark assessment can be used as basis for advice. It considers however, that more work is still needed in the next years to improve the input data and the assessment methodology. Reference points need to be determined based on the final assessment according to ICES reference point guidelines.

Recommendations

Near future until the next advice

It is strongly recommended to check and update the stock annex including a description of the raising of catch data done in InterCatch.

Longer-term

A data compilation workshop is needed prior to the next benchmark. Otherwise a successful benchmark workshop cannot be guaranteed. The level of preparation needs to be checked more carefully before the final benchmark workshop. Meetings by correspondence well before the benchmark workshop may help to detect problems well in advance.

As it stands, the compilation of the Rockall haddock stock assessment input files from InterCatch is based on numerous manual data processing. It is thus error-prone and not readily reproducible. Further automatization of the input data compilations is highly recommended to avoid unnecessary sources of error and to improve transparency. It is worth noting that this is a general issue for all stocks currently using InterCatch. A new system based on the regional databases and creating input data files using R scripts would be beneficial.

XSA is a deterministic model and therefore results are directly impacted by uncertainties in age readings and the amount of catch reported. Moving to a statistical model like SAM or SS3 would permit to better account for such uncertainties in the model fits. These models also provide confidence regions on the parameter estimates and stock dynamics indicators outputs (F, SSB, R) associated with those uncertainties.

The survey index is derived by applying mean density estimates for different depth strata. However, other factors may also influence the spatial distribution of haddock potentially biasing the survey index. In addition, a large part of the area on Rockall is currently protected and fishing is not allowed. It is assumed for the index calculation that densities inside and outside these zones are similar. It may be tested with historical information whether this assumption is likely correct or not.

The survey from which the index is derived changed its sampling design in 2011. It is recommended to further explore whether the currently used survey index could be split into two separate indices once more years with the new design (and better coverage of depth strata) are available.

The expert that presented the analyses conducted to generate the survey index also mentioned the presence of noticeable diurnal trends in the catches (i.e. haddock seem to be more catchable during the day than at night). The effects that this diurnal variation in trawl catch efficiency can have on the survey index should be further investigated.

When forecasting mean weight-at-age, cohort strength effects seem to play a role. It is recommended to analyse whether methods including cohort effects for mean weight-at-age could improve the short-term forecasts of yield and biomass.

Currently a knife-edge maturity ogive at age 3 is used. A smoother maturity ogive may be derived from further analyses of available data.

Natural mortality is currently kept constant over ages and years. Alternatives may be tested for the next benchmark. At least age-dependent natural mortalities could be derived based on life-history analyses.

The stock coordinator pointed out that discrepancies exist in the ageing of Rockall haddock, noting that different ages were estimated by different laboratories using the same otoliths. The accuracy and precision of ages obtained from haddock otoliths should be further investigated so as to establish whether ageing biases are present and to determine whether/how they will affect the generation of catch-at-age matrices derived from age–length keys.

The residuals in the final benchmark assessment are smallest for the oldest age in the survey time-series. This indicates that the assessment model follows the index for the oldest age more closely than the indices for the younger ages. Runs with the proposed swept-area indices in combination with age 7 as plus group do not show this feature. Further investigations on the estimation procedure behind the proposed swept-area indices are needed next to analyses whether it is problematic if the assessment follows the indices for the oldest age in the survey to a larger extent.

Annex 5: Stock Annex for Haddock in Division 6b

Stock ID	Stock name	Last updated	Link
Had.27.6b	Haddock (*Melanogrammus aeglefinus*) in Division 6.b (Rockall)	May 2020	Rockall haddock