

INTERBENCHMARK WORKSHOP ON THE ASSESSMENT OF NORTHEAST ATLANTIC MACKEREL (IBPNEAMAC)

VOLUME 1 | ISSUE 5

ICES SCIENTIFIC REPORTS

RAPPORTS
SCIENTIFIQUES DU CIEM



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

The material in this report may be reused for non-commercial purposes using the recommended citation. ICES may only grant usage rights of information, data, images, graphs, etc. of which it has ownership. For other third-party material cited in this report, you must contact the original copyright holder for permission. For citation of datasets or use of data to be included in other databases, please refer to the latest ICES data policy on ICES website. All extracts must be acknowledged. For other reproduction requests please contact the General Secretary.

This document is the product of an expert group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the view of the Council.

ISSN number: 2618-1371 | © 2019 International Council for the Exploration of the Sea

ICES Scientific Reports

Volume 1 | Issue 5

INTERBENCHMARK WORKSHOP ON THE ASSESSMENT OF NORTHEAST ATLANTIC MACKEREL (IBPNEAMAC)

Recommended format for purpose of citation:

ICES. 2019. Interbenchmark Workshop on the assessment of northeast Atlantic mackerel (IBPNEAMac). ICES Scientific Reports. 1:5. 71 pp. <http://doi.org/10.17895/ices.pub.4985>

Editor

Niels Hintzen

Authors

Anders Nielsen • Andrew Campbell • Aril Slotte • Claus Reedtz Sparrevohn • David Miller • Gary Nelson • Gjert Endre Dingsør • Guðmundur J. Óskarsson • Höskuldur Björnsson • Ian Gatt • Jan Arge Jacobsen • Jens Christian Holst • Jens Ulleweit • Leif Nøttestad • Lisa Ailloud • Lisa Readdy • Magne Aldrin • Martin Pastoors • Neil Campbell • Noel Cadigan • Rose Mari Berge • Samuel Subbey • Sean O'Donoghue • Sindre Vatnehol • Sonia Sanchez • Steve Mackinson • Teunis Jansen • Thomas Brunel



ICES
CIEM

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

Contents

i	Executive summary	ii
ii	Expert group information	iv
1	Background	1
2	IBP Process	2
3	Summary of first meeting (11–13 December, CPH, Denmark)	3
4	Summary of decisions from 1 st meeting	6
5	Summary of final meeting (4–7 March, IJM, The Netherlands)	7
	5.1 Quality assurance of tagging data	10
	5.2 Assessment model results	11
	5.2.1 Using a selected subset of the RFID data	11
	5.2.2 Changes in model configuration relative to young ages	19
	5.2.2.1 Rational for revisiting model configuration	19
	5.2.2.2 Consequences of the changes made	20
	5.3 Reference points results	28
6	Summary of decisions (2 nd meeting)	36
7	Reflection of reviewers	46
Annex 1:	List of participants	49
Annex 2:	Agenda IBP meeting 11–13 th of December	51
Annex 3:	Agenda IBP meeting 4–7 th of March	52
Annex 4:	Recommendations	54
Annex 5:	Future research requirements	56
Annex 6:	Minority Statement and Response to the Minority Statement by the IBP	58
Annex 7:	References	62
Annex 8:	Working Documents	63

i Executive summary

The NEA mackerel inter-benchmark group met at ICES HQ for a data workshop 11-13 December 2018 and a final second meeting 4-7 March 2019 at Wageningen Marine Research, Netherlands. The group consisted of scientists (national labs, universities and industry scientists), stakeholders, managers and re-viewers from in total 15 different countries.

The ToRs of the group show a specific focus on the influence of the tagging data on the assessment, since at the 2018 working group meeting (WGWIDE) sensitivity analyses showed a substantial change in perception of the mackerel stock when tagging data was excluded from the assessment. This indicates a high, and potentially incorrectly specified, weight of this data source. The group furthermore focused on a revised recruitment index and at the quality of the catch sampling.

After thorough review of the data and analyses the group made a number of decisions with regard to the data that should be included in the assessment model and the statistical approach to model these data. The updated recruitment index was considered to be of good quality and was implemented in the assessment.

There were concerns related to the age reading quality, directly applicable to the quality of the catch-at-age data and potentially survey data. Analyses showed that experienced age readers are in agreement in ~41% on the age of older fish (age 7+) and that national catch-at-age frequency distributions originating from similar types of fisheries diverged substantially. The impact of ageing error on the perception of the stock turned out to be minimal and was therefore not considered an issue for this inter-benchmark, although recommendations are made to improve in this field.

Several analyses were undertaken to investigate the quality of the RFID tagging data, including spatial coverage of the recapture stations, detection rate of RFID tags in factories, data exploration to infer exploitation rate from the tagging data itself, changes in perception of cohort size as inferred from the tagging data itself and a variety of alternative assessment model configurations in which the RFID dataset was truncated in time, by age, by number of years tagged fish were allowed to have been out at sea (years-out) and models in which the original tagging data was replaced with alternative methods to aggregate the tagging data into e.g. indices of abundance.

The main topics of discussion, and the decisions made (in between square brackets) related to: #1 the maximum number of years a tagged fish was allowed to have been out at sea before recapture [1 & 2 years-out, to prevent a bias due to tag-loss], #2 the age classes to be included in the assessment [5+, to ensure full mixing of tagged fish in the population], #3 apparent tag loss [could not be verified by data analyses and was therefore not included], #4 years of release of RFID tagging experiments [from 2013 onwards, prior to this year the spatial coverage of detection stations was biased], #5 the treatment of 'survival rate' between the older steel tag data and the newer RFID dataset [as separate processes in the assessment], #6 the data format in which tagging data is used in the assessment (e.g. as single observations, aggregated or as an index of abundance [as single observations]) #7 the statistical treatment of the swept area survey [with a correlation structure imposed, un-changed from 2017 benchmark], #8 the statistical treatment of tagging data [with-out any correlation structure imposed]. #9 & 10: other model parameters [changes in random walk F and observation variance combinations] and #11: New reference points were calculated [MSY $B_{trigger}$ = 2.50Mt (was 2.57), F_{MSY} = 0.23 (was 0.21), B_{lim} = 1.99Mt (was 1.94Mt), B_{pa} = 2.5Mt (was 2.57Mt), F_{lim} = 0.46 (was 0.48) and F_{pa} = 0.37 (was 0.35)].

There was furthermore discussion on the diverging views on estimated stock trends and the absolute estimate of biomass by the assessment model. There were concerns that the recent history overestimated the decline in stock trend since 2011 and that total biomass was underestimated given that fisheries independent indices of abundance (e.g. egg survey, swept-area survey, tagging data), when taken as indicators of absolute biomass, would suggest a higher density of Mackerel. With the final accepted assessment from this inter-benchmark, the previously noted decline since 2010 has shifted to 2014. The egg and pelagic survey groups from ICES are requested to reflect on best methods to use the egg survey and swept-area survey as indicators of absolute biomass.

All analyses and decisions to support this IBP were performed prior or at the final meeting. All decisions were made by consensus. The meeting closed at 15:00 on Thursday the 7th of March.

ii Expert group information

Expert group name	Interbenchmark Workshop on the assessment of northeast Atlantic mackerel (IBPNE-AMac).
Expert group cycle	Annual
Year cycle started	2019
Reporting year in cycle	1/1
Chair(s)	Niels Hintzen, The Netherlands
Meeting venue(s) and dates	11–13 December 2018, ICES Secretariat, Denmark 4–7 March 2019 , Wageningen Marine Research, the Netherlands

1 Background

The NEA Mackerel stock assessment uses many different datasets such as catch, survey indices and tagging data to inform on stock trends and rate of exploitation. Owing to contradictory signal in these data sources, the estimation of stock trends and exploitation rate is uncertain and perception changes from year to year are not uncommon. At the last assessment working group meeting (WGWIDE 2018) the sensitivity to especially the tagging data was highlighted and the group felt ill prepared to explain the sensitivity to the tagging data. For this reason, an inter-benchmark was organised to primarily investigate the quality of the tagging data and its implementation in the assessment. Additionally, other data considerations such as updating recruitment indices and evaluation of inclusion of historic catch data were added to the list of topics to investigate.

2 IBP Process

The IBP has 4 formal meeting events, starting with a physical meeting in Denmark at the 11–13 of December. The initial meeting aims to focus the scope of the IBP and to ensure that the required personal capacity and scientific underpinning are available to the group.

The IBP follows-up on progress at a first WebEx meeting on the 22nd of January where results, particularly aimed at preparing and investigating the accuracy of data being used as input to the assessment, are presented and discussed. Decisions are made that allow for further assessment analyses.

The IBP follows-up on progress at a second WebEx meeting on the 12th of February where results, particularly aimed at assessment model evaluations and sensitivities are presented and discussed. Decisions are made to allow for further assessment and reference point fine-tuning.

The IBP is concluded at a second physical meeting in The Netherlands at the 4–7 of March. At this meeting the final settings of assessment input data, configuration and reference points are decided. A new stock annex describing the exact specification of the newly benchmarked stock is the main output of the IBP.

Participants from many different ICES member countries contributed to the IBP. The IBP is open to anyone with an interest and are expected to contribute actively to analyses and discussions.

3 Summary of first meeting (11–13 December, CPH, Denmark)

The agenda of the first meeting can be found in Annex 2. The summary below reflects the points of discussion and suggested ways forward. In total presentations on 17 WDs (in draft) were presented to the group.

#	Title of WD	Author	Topic it addresses
WD1	NEA Mackerel: Analysis of tagging data	Höskuldur Björnsson	Tagging data
WD2	Some comments on modelling tag data in SAM for the mackerel assessment and a general comment on the “process error” in SAM	Magne Aldrin	Assessment model (tagging data, process error)
WD3	RFID tagging data sensitivity run on the WGWIDE 2018 assessment	Claus R. Sparrevohn	Sensitivity of the assessment to tagging data
WD4	NEA Mackerel: Alternative assessment	Höskuldur Björnsson	Assessment model
WD5	On a possible serious bias in the egg biomass indexes on NE Atlantic mackerel	Jens Christian Holst	Survey indices (Egg survey)
WD6	On possible bias in the IESSNS trawl biomass index on NE Atlantic mackerel	Jens Christian Holst	Survey indices (IESSNS)
WD7	Analyses of the process error in the mackerel assessment	Thomas Brunel	Assessment model (process error)
WD8	Investigating different model configurations to estimate post release survival in the mackerel assessment	Thomas Brunel	Assessment model (tagging data)
WD9	North East Atlantic Mackerel assessment – sensitivity to error structure for the IESSNS	Sonia Sánchez and Thomas Brunel	Assessment model (survey indices)
WD10	North East Atlantic Mackerel assessment – sensitivity to RFID tags	Sonia Sánchez and Thomas Brunel	Assessment model (tagging data)
WD11	Sensitivity of the mackerel assessment to the input data, with particular emphasis on the RFID tagging data	Thomas Brunel	Assessment model (survey/tagging data)
WD12	Developing abundance at age indices from tag-recapture data	Sindre Vatnehol and Aril Slotte	Tagging data
WD13	Truncating the catch time series for Mackerel	Claus R. Sparrevohn	Assessment model (catch data)
WD14	Update of the abundance index of age 0 NEA mackerel based on demersal trawl surveys in October – March	Teunis Jansen	Survey indices (Recruitment index)
WD15	Fishery/Survey Timing	Andrew Campbell	Assessment model

WD16	Issues regarding updated version of RFID-tag data 2019	Aril Slotte	Tagging data
WD17	Investigation of changes in apparent 'survival' in the mackerel tagging time series data	Steve Mackinson et al.	Tagging data

Aril Slotte presented the survey design and his explorations on the collection and data preparation of tagged Mackerel. The similarity in instantaneous mortality between Mackerel tagged with steel or RFID tags and handling between Norway and Iceland stood out. The uncertainty on the precision of the fish scanned in factories in the time of the steel-tagging programmes received substantial attention as well as the potential change in survival of Mackerel through time and the change in their distribution area that may affect estimation of survival when comparing the steel with the more recent RFID tags. The analyses confirmed that efficiency of detecting mackerel in processing factories was appropriate and in line with assumptions by the assessment group, as well as the assumption that the new survey design did not alter the direct survival of mackerel.

Sindre Vatnehol presented a new method to create an index of abundance from the tagging data that could be used in assessment models operationally similar to other age-based indices (WD12). The internal consistency of these indices was shown (and considered high) and forms a visually attractive tool to judge tagging data quality outside of the assessment model.

Höskuldur Björnsson presented explorations of the tagging data and showed the consistency in targeted length-groups over the years (WD1). The discussion focused on the use of a 'tag-loss' parameter that could be added to the assessment model to capture additional mortality of fish carrying a tag for longer periods (spending more time in the sea before recapture).

Thomas Brunel presented a number of assessment analyses showing the effect on estimated survival rates when the tagging data would be subset into a number of year-blocks rather than one continuous time-series (still differentiating RFID from steel tags) (WD8). These analyses showed a transition in estimated survival from ~0.4 to 0.1, even for steel tags in the later period. The interpretation by using subsets did not differ from one longer time-series. The change in survival for fish staying longer out in the sea was brought up again and it was discussed that the potential for correlation in release and/or recapture year may have an impact. For that reason, the SAM model was changed to incorporate estimation of correlation of this effect. His analyses furthermore showed a very large impact of the 2017 data, which has a larger impact than previous years. It was suggested to make use of only fish being less than 3 years out at sea to undo any potential effect of additional mortality for fish staying longer out in the sea. Others suggested that the change in distribution of Mackerel over the past years could result in a different mixture of age classes in different areas and therefore no longer result in a random sample of the recaptured fish.

This last suggestion is related to comments by Magne Aldrin presented later on the need for accounting for correlations within the different data types in order to balance their effect on the model estimated (WD28). He suggested that one could expect to see one or more of the following three correlation structures in the tag data: 1) Correlations between all recaptures of fish with the same release year, because the fish tagged in the release year may not be a random sample of the population, 2) correlations between all recaptures in the same year, because the fish recaptured in a given year may not be a random sample, and 3) correlations between recaptures of the same fish cohort tagged the same year, because the fish at a given age tagged that year may not be a random sample. Later, Anders Nielsen presented the results of a SAM analysis accounting for correlations within recaptures with both equal release year and recapture year.

Neil Campbell presented the change in the Scottish proportion of landings over the year and the shift herein (WD15). He suggested that this may affect the fitting of survey indices as there was a one-directional shift in timing of the fishery. It was suggested to investigate this aspect for the entire fleet by month and make modifications to SAM to cope with this.

Jens Christian Holst investigated different potential biases in data sources such as the estimation of egg densities and the estimate of the swept-area survey (WD5 & 6). The discussion pointed out that the egg survey results presented were not relevant for mackerel as they use a different survey design and analyses method while the potential bias in the swept area survey were countered with other information showing no significant impact of the curved trawls as well as change in temperature in the mackerel habitat (oceanographic variability in general). A discussion followed on some influential hauls in the survey in 2017 but the survey group had examined these already and felt they should be kept in because sufficient evidence was available to validate the observed densities.

Thomas Brunel furthermore presented analyses focussing on process error which seemed to not play a major role in the assessment (WD7). This was followed by a presentation by Magne Aldrin who suggested a different mathematical implementation of the process error in the SAM model (WD2). Anders Nielsen commented that the suggested modified mathematical formulation will give slower convergence and in addition will have no practical consequences for the results. This suggestion was discussed but the group felt that the existing formulation was both correct and appropriate. It was suggested though to have a sensitivity run showing the differences between the two implementations.

Teunis Jansen updated the group on pending updates of the recruitment index series which is currently lacking data for 2016–2017 (WD14). He informed the group on additional quality checks and work that had to be executed to make these new numbers available.

The group continued to discuss and investigate different assessment configurations and sensitivity runs with e.g. different input data such as an age-based index derived from the tag data. These discussions led to agreeing on a number of follow-up steps and an agreed time-line when to deliver these. These can be found below.

[illegible]

4 Summary of decisions from 1st meeting

The proportion of fishing that takes place prior to the timing of the surveys has changed through the years. The impact of changing the assumption that these proportions change through time was tested in the assessment model. Given that there was no distinguishable difference, the group decided to maintain the fixed time-series.

A revised and updated recruitment index was presented to the group and reviewed the analyses and results. On that basis, the group decided that the updated recruitment index should be used in the benchmark assessment process.

Furthermore, it was decided to not change the way the swept area index and the egg-survey were used in the assessment.

5 Summary of final meeting (4–7 March, IJM, The Netherlands)

The agenda of the second meeting can be found in Annex 3. The summary below reflects the points of discussion and suggested ways forward. In total presentations on 11 WDs (in draft) were presented to the group.

#	Title of WD	Author	Topic it addresses
WD18	Note SAM configuration for Mackerel	Anders Nielsen	Assessment model
WD19	NEA Mackerel Catch Sampling	Andrew Campbell	Catch data
WD20	Assessment sensitivity to different criteria for selecting subsets of the RFID dataset	Thomas Brunel	Assessment model (tagging data)
WD21	NEA Mackerel Analysis of assessment based on tagging data	Höskuldur Björnsson	Assessment model (tagging data)
WD22	Changes in catch at age distributions of mackerel 2005-2016	Thassya C. dos Santos Schmidt and Aril Slotte	Catch data
WD23	Pelagic AC Input Inter benchmark NEA mackerel	Pelagic Advisory Council (PelAC)	Stakeholder information
WD24	Exploring trends in the age distributions within the tag-recapture data	Sindre Vatnehol and Aril Slotte	Tagging data
WD25	Review of procedures and data in the Norwegian mackerel tag-recapture material	Dankert Skagen	Tagging data
WD26	Spatio-temporal distribution of RFID tag recaptures	Konstantina Dimitrakopoulou and Aril Slotte	Tagging data
WD27	Deriving Auxiliary Stock Information from the mackerel RFID Tags	S. Subbey, B. Sandercock, K. Dimitrakopoulou, N. Jourdain, D. Skagen, A. Slotte, S. Vatnehol	Tagging data
WD28	Some results for various way to handling tag and catch observations	Magne Aldrin	Assessment model (tagging data)

Aril presented WD 16. The discussion focussed on the differences predicted by the assessment model in survival rate that were unlikely to be caused by a difference in tagging practice (manual vs automatic jigging for the steel vs RFID tags). It was discussed that tag loss could also explain the differences in survival rate which may arise from a different method of placing the tag (in muscle for steel tags vs abdomen for RFID tags [red: for food safety]). It was hypothesized that mackerel could get rid of tags through body movement, and as such there could be a difference between sexes.

As in the December meeting the scanning efficiency of the factories was discussed, but with a 94% tested efficiency it was considered sufficiently dealt with. It was discussed that the number

of factories participating in the recapture analyses has expanded from 2013 onwards and included factories also outside of Norway to more appropriately reflect the expansion of the Mackerel stock in recent years.

The ageing procedure for the tagged fish remained a point of discussion throughout the meeting. Since tagged fish cannot be selected out when recaptured (red. For practical reasons in the factory), age-length keys are used. However, there is substantial disagreement between experienced age-readers on the age of older fish.

Höskuldur presented an alternative assessment model (WDs 4 and 21) for robustness-test purposes and evaluated the impact of incorporating tag-loss. The tag-loss estimate (~ 0.17) indicated that tag-loss cannot be ruled out in recent years and the discussion focussed on reasons how tag-loss could be explained. It was mentioned that the mathematical implementation may not necessarily match reality in which tag-loss can only occur once while the assessment implementation indicates a continuous rate.

Thomas presented WD 20 in which he showed how the perception of cohort size changed when tagging data of the same cohort, but of fish that had spent more years out at sea, would be added. When integrating only 1 or 2 years-out in the assessment, it was pointed out that this results in down weighting the tagging data (the over dispersion parameter increases). Discussions focussed again on whether tag-loss could be introduced to account for this effect and thereby reduce the need for a subjective way of selecting data. Reviewers pointed out that using only 1 or 2 years-out may conflict with the assumption that fish are well mixed in the population. Thomas furthermore showed that sample size of the tags or e.g. the swept-area survey has a marked effect in the assessment. Subsampling the tagging data results in a markedly different perception of the stock. Here it was pointed out though that with conflicting survey estimates in the assessment, removing part of the data will always result in changes in perception.

Magne presented WD 28 in which he demonstrated the usability of correlation structures in especially the tagging data. He demonstrated it based on three ways of using the tagging data: 1: using the raw tagging data, 2: using the tagging data but aggregated by recapture year and 3: converting the tagging data to an index of abundance. All showed similar, but slightly different patterns. It was noted that correlation structures were to be investigated later on in the week when model performance was on the agenda.

Andy presented WD 19 in which he highlighted the differences in catch-at-age, catch-at-length distribution between Irish and Scottish samples (Norwegian data left out because it was similar to Irish data). The differences could not be resolved at the meeting, but it was noted that the quality of age reading was poor for older ages (41% for ages 7 and older). A sensitivity assessment was evaluated in which Scottish age readings were replaced with Irish which showed hardly any difference in perception while on the data input side there were substantial differences in catches by cohort. The discussion focussed on alternative assessments, e.g. length based assessments or separating the fit to ages from the fit to total catch (currently combined in SAM), but these seemed not to be possible in the short term. Alternatively, using an age-misspecification matrix could be an option for the future. It was decided to proceed with the data the working group has been using (including Scottish age sampling) given the minute impact as shown in Figure 5.1 below.

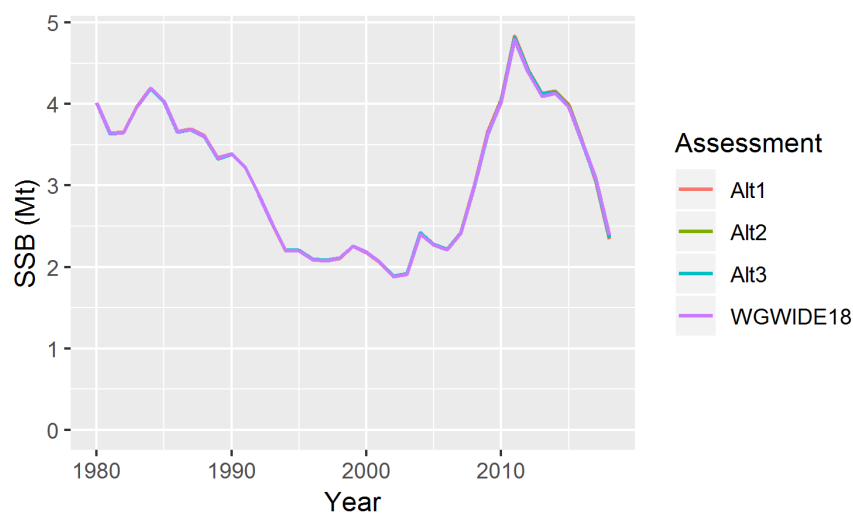


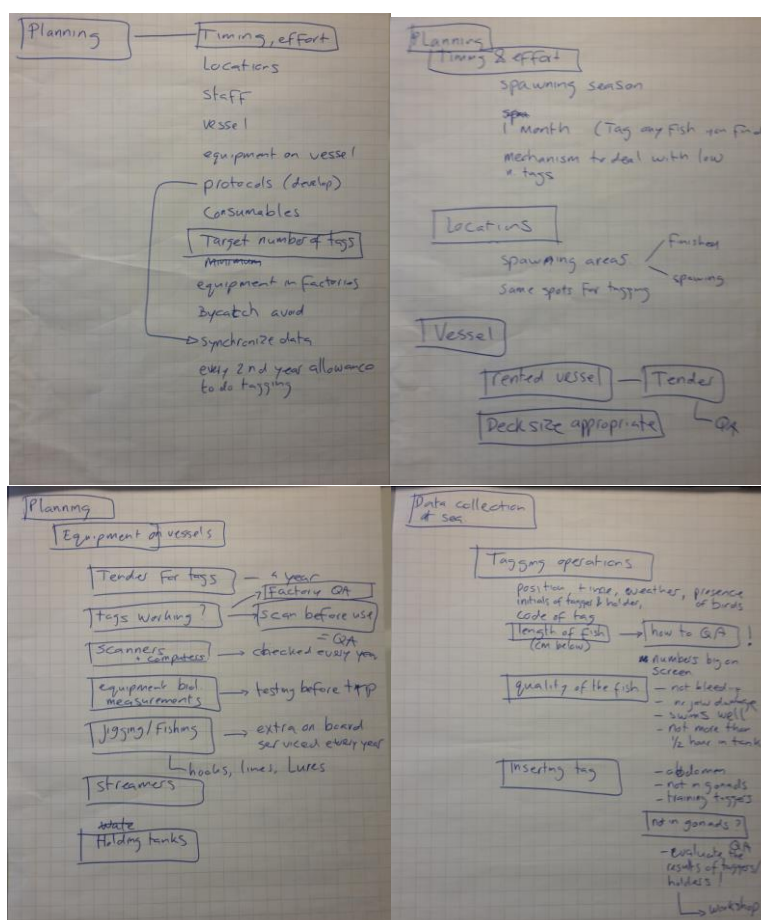
Figure 5.1. Comparison of assessment trends using alternative assumptions on age-composition of the Scottish part of the catch.

5.1 Quality assurance of tagging data

Several working documents were presented that analysed different parts of the tagging data and the role of that data in the context of the SAM assessment model.

IBPNEAMac also looked at more generic quality assurance protocols that could be developed around the tagging data or more broadly to the different inputs and outputs of the stock assessment model (WD 23). Quality assurance (QA) is the set of procedures and protocols that are aimed at avoiding mistakes and defects in manufactured products. We discussed that in principle, all the data that is used within the context of a stock assessment process, should have a form of quality assurance that is well documented and that could be tested on whether procedures have been followed. In practice, such quality control procedures are not fully developed yet for every data source that is used in the mackerel assessment.

Specifically for the tagging data, we discussed the potential elements of a quality assurance programme that captures the planning process, data collection at sea, data collection on recaptures, data on landings and biological sampling, data storage and transfer, data analysis and input into the assessment. Each of these steps would then be worked out in more detailed steps, like in the examples below.



It was agreed that there is an overall need within ICES to develop robust quality assurance procedures for the different assessments and that the work on the tagging data could be used as a first example on how this could be done. The illustrations will be digitized and further worked out in follow-up workshops.

5.2 Assessment model results

The detail of the decisions taken during the benchmark is given in section 6. The present section focusses on the consequences of these decisions for the assessment, which are illustrated both in the estimated stock trajectories and in model diagnostics.

Those decisions lead to 2 main changes in the assessment:

- 1) A selection of the RFID data is now used in the assessment
- 2) Model parameters referring to the fit to the catches and the variability of the fishing mortality are now distinct for the young age-classes – very marginal catches for ages 0 and 1 – and for the fully recruited fish.

The consequences of these two main changes is described below.

5.2.1 Using a selected subset of the RFID data

The reasons for selecting this specific subset of the data are exposed in section 6 (#1,2 and 5) and are not developed here. Using this subset of the RFID dataset has two main consequences:

- First, the information on cohort abundance provided by the RFID dataset is different in the subset chosen and in the entire dataset. This comes from the fact that cohort abundances for the first years of RFID tagging (2011 and 2012) were informed both by early recoveries (e.g. in 2013) and late recoveries (e.g. 2017). As late recoveries seemed to consistently give a higher perception of abundance-at-age (see WD 20 and decision #5 in section 6), there is automatically a positive bias for abundances at age in old release years, which is not present in recent release years (which are only informed by recent recaptures). This results in a spurious declining trend in the abundance information provided by the RFID data (especially for ages 2-5), which is not –or less- visible when a selection of tag with one year at liberty is operated (figure 5.2.1). The information in the subset of the RFID data now indicates less decrease (or a decrease starting later) for the younger ages than in the whole dataset, and an increase for the older age groups. Combined as an index of the overall biomass of the stock, the subset of the RFID data now gives the perception of a stock increasing in between 2013 and 2014 and being rather stable since then (WD 12).
- Secondly, the number of data points included in the selected subset is substantially lower than in the full RFID dataset (49 data points compared to 208 previously). In the last accepted assessment (WGWIDE2018), the RFID data had much more influence on the recent stock trajectory than other data sources, especially since the incorporation of the 2017 recaptures (WD 10). Analyses presented at IBPNEAMAC2019 showed that the weight of each dataset is not only linked to the quality of the information contained, but for a large part is related to the number of data points (WD 20). Given that the information provided by the whole RFID dataset is of limited quality (moderate to poor internal consistency), the high weight of this data source in the last accepted assessment seemed inappropriate. By reducing the number of data points, the influence of the RFID series is reduced and the other data sources get more weight. The comparison of the “leave one out” runs for the WGWIDE 2018 update assessment and the assessment run on the subset of the RFID data illustrates well this change in weight of the different data sources (figure 5.2.2). In the WGWIDE 2018 assessment removing any of the surveys from the assessment had much less effect than removing the RFID data. In the assessment run with the subset of the RFID data, removing the egg survey has nearly the same effect than removing the RFID data, and removing the IESENS causes a revision of the same

magnitude in the opposite direction. This indicates a more balanced weight of the different data sources when the subset of the RFID data is used.

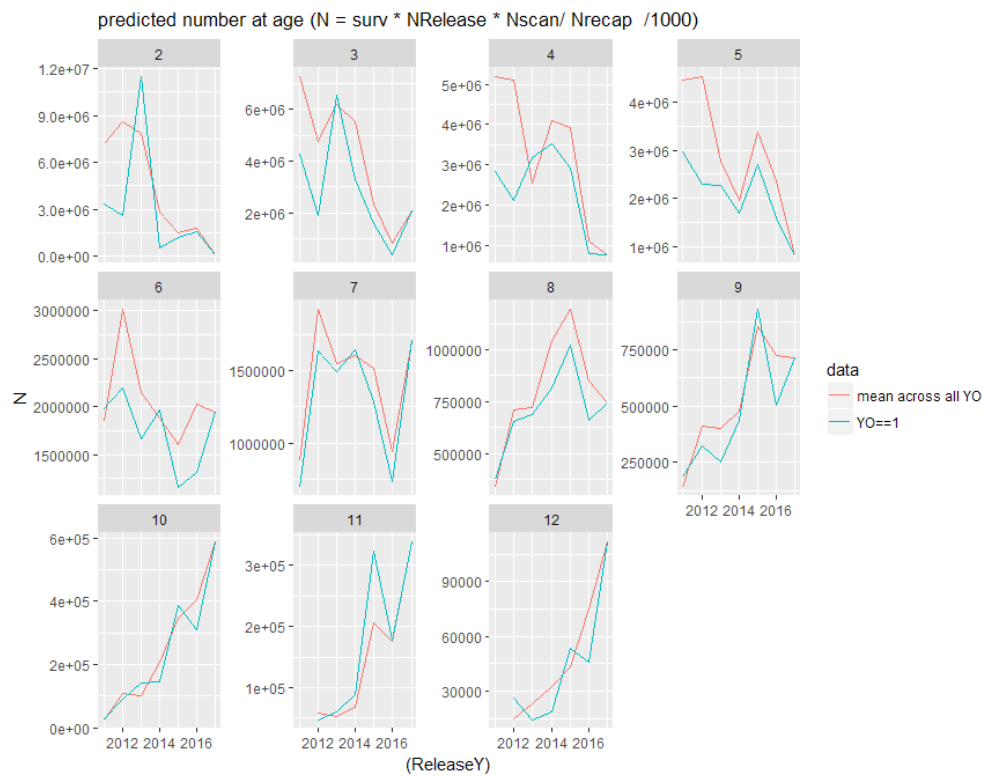
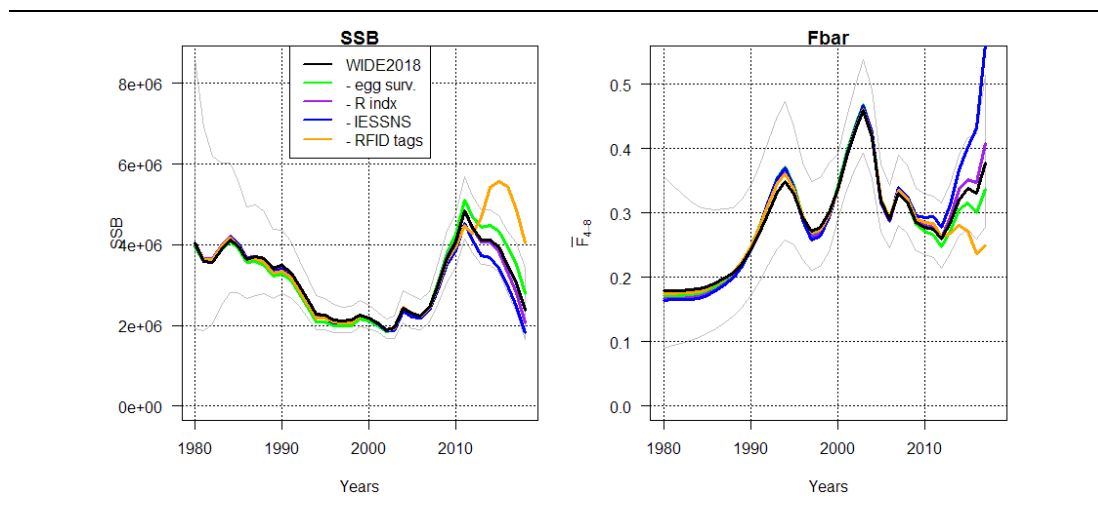


Figure. 5.2.1 : illustration of the information from the RFID data (recaptures transformed into abundance at age) for the recaptures after one year at liberty compared to the whole RFID dataset (for which the average of the successive recaptures is shown).

Update assessment (WGWIDE 2018)



Assessment using the subset of the RFID data

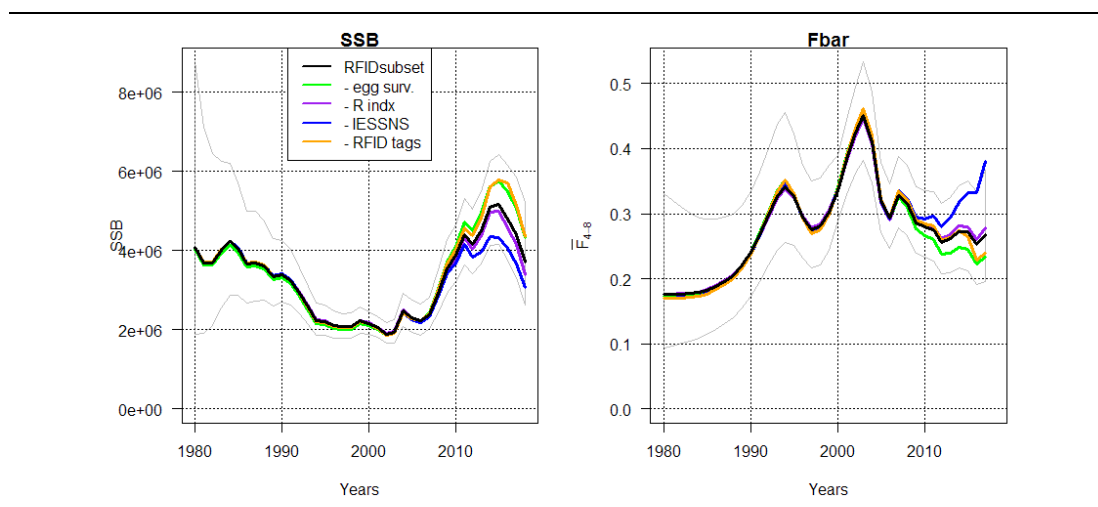


Figure 5.2.2 : comparison of the leave one out runs (assessments run with one of the data sources removed) of the 2018 update assessment (WGIWDE) and the assessment run with the selected subset of the RFID data.

The consequences for the assessment of using this subset were the following:

- A different model fit to the data (figure 5.2.3), with an improvement of the fit to the IESSNS index (lower standard deviation of observation errors), and a poorer fit the other data sources (although the difference is small for the catches). The fit to the RFID and steel tag is unchanged (note that the assessment with the subset was run with a common overdispersion for both tagging datasets, which had little impact on the assessment, see WD 18). Process variances were not markedly affected. The catchability of the IESSNS

and of the egg survey were revised downwards, reflecting the change in the estimated stock size in the recent years.

- There was a pattern present in the residuals of the RFID tag recaptures for the update assessment from WGWIDE 2018 (figure 5.2.4) : for age at release 2 to 5 and for the oldest releases (mainly years 2011 to 2013), the residuals for the recaptures from the first years after release (1 or 2 years out) were mainly positive, while negative residuals were observed for the tags having remained at liberty more than 2 years. This pattern suggests a decrease in the tag concentration over the years (consistent with tag loss or malfunction), which is not explicitly modelled in the current assessment.
- After operating the selection of the RFID data decided at IBPMAC 2019, only recent recaptures are used, residuals for the remaining data points do not show any pattern (figure 5.2.5).
- The retrospective analyses indicate that the assessment using the subset shows more instability than the update assessment (figure 5.2.6). There are indeed larger revisions for each retrospective peel for the assessment using the subset, probably due to the fact that the RFID data is shorten by 2 years. However, the revisions can occur in both directions (at least for the most recent peels) which indicates that there is no systematic bias in the assessment.
- There is a large change in the perception of the stock trajectory for the period after 2010 (figure 5.2.7). The lower influence of the RFID data (formerly pulling the update assessment down), combined with the fact that the remaining RFID data no longer indicate a decline in the stock, lead to a SSB increasing between 2012 and 2015 and declining afterward to just under 4 mt. The update assessment from WGWIDE 2018 gave a SSB declining since 2011, going down to just above 2.0mt in 2018. The trend in fishing mortality in the WGWIDE 2018 update assessment is an increase between 2011 to 2017, while in the assessment using the subset of the RFID, fishing mortality is rather stable over this recent period.

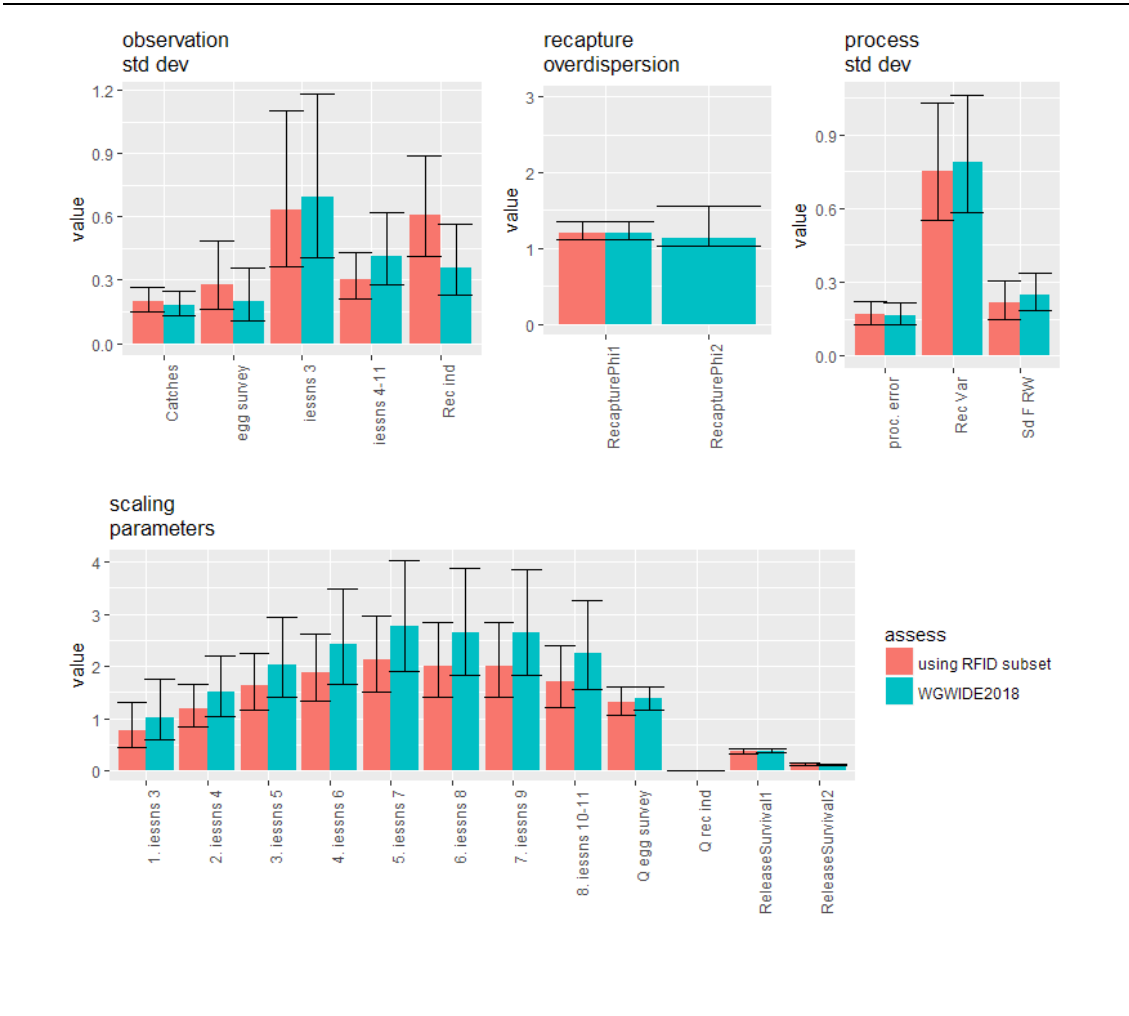


Figure. 5.2.3 : comparison of estimated model parameters for the assessment run at WGWIDE 2018 and the assessment using the selected subset of the RFID tagging data

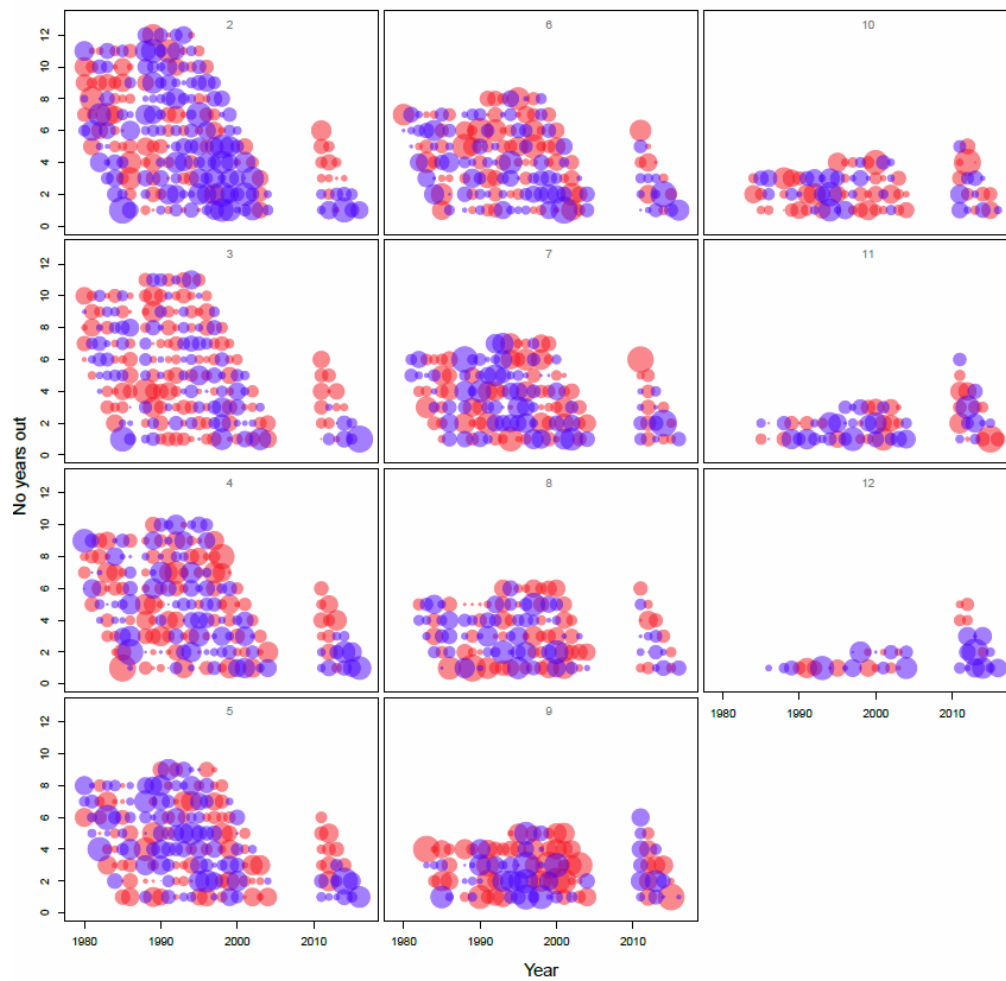


Figure 5.2.4 : model residuals for the tagging data (red : negative, blue : positive) for the assessment using the entire RFID data. For each panel, representing an age at release, the X-axis correspond to the year of release, and the Y-axis corresponds to the number of years at liberty before recapture.

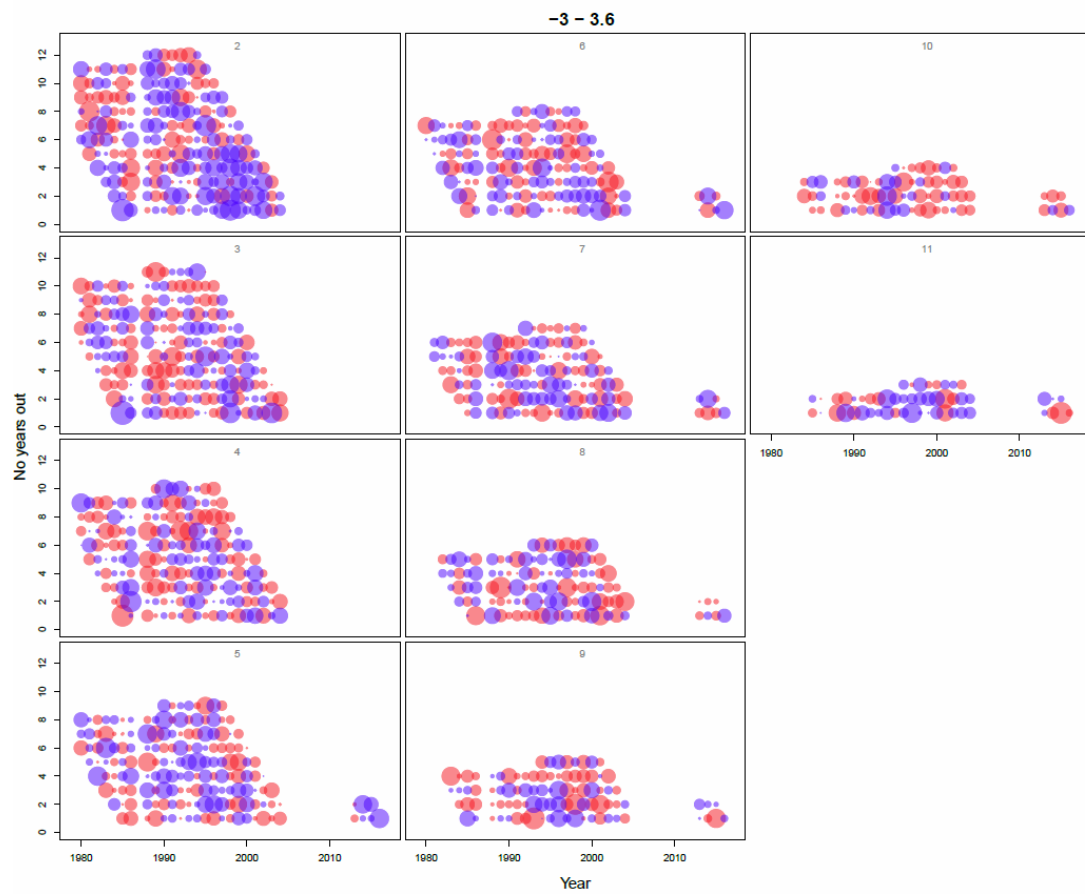


Figure 5.2.5 : model residuals for the tagging data (red : negative, blue : positive) for the assessment using the subset of the RFID data. For each panel, representing an age at release, the X-axis correspond to the year of release, and the Y-axis corresponds to the number of years at liberty before recapture.

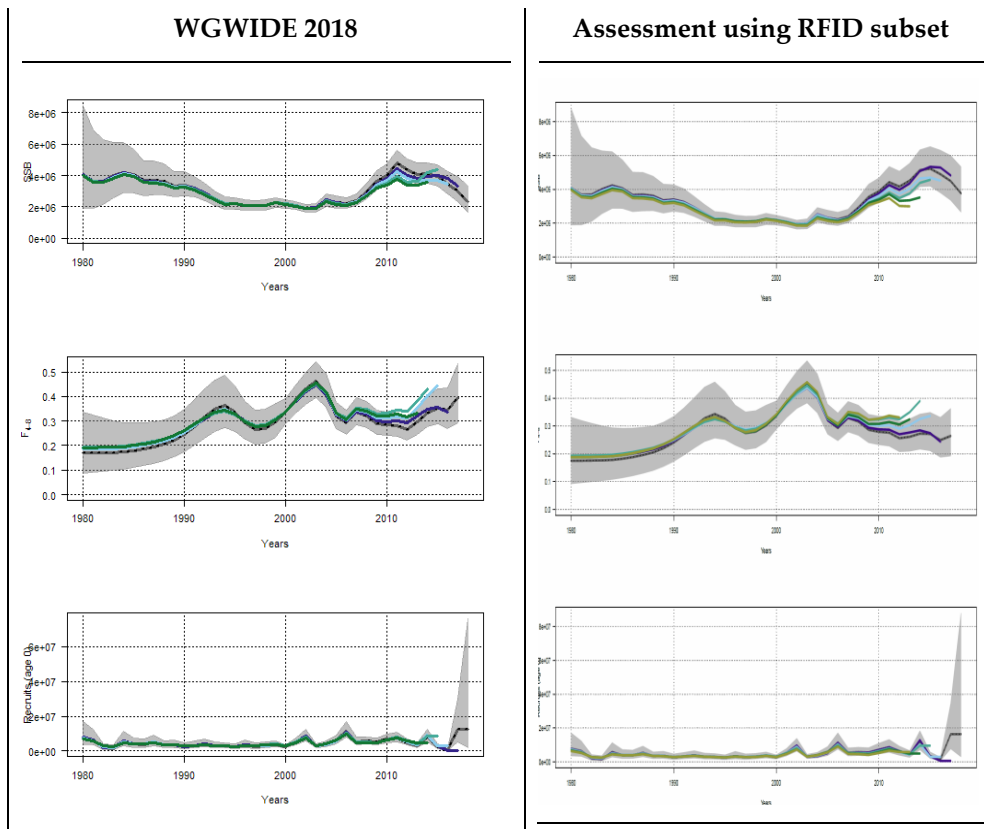


Figure.5.2.6 retrospective analyses for the WGWIDE 2018 assessment and the assessment run on the subset of the RFID data.

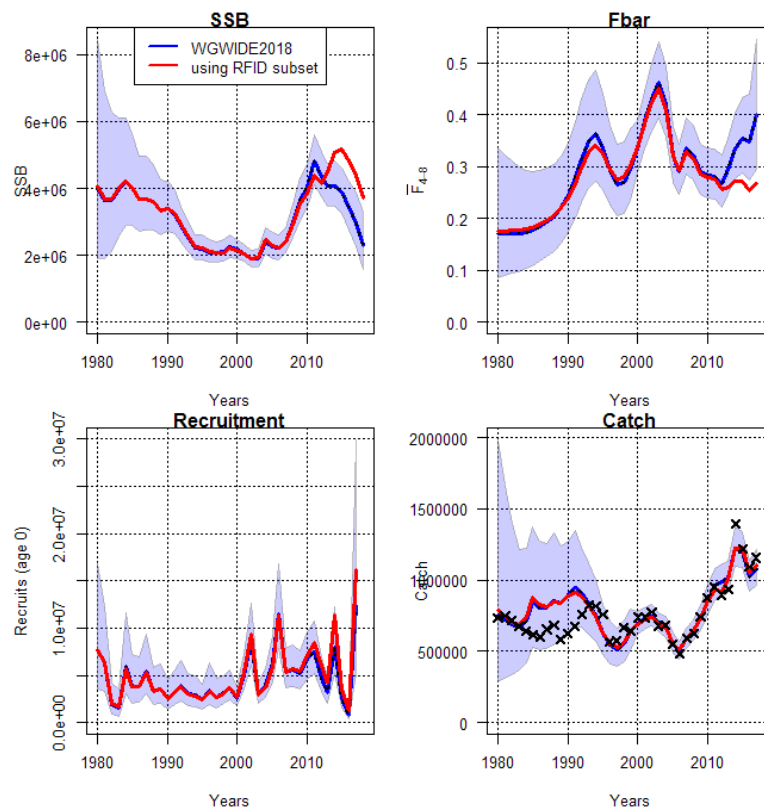


Figure. Figure 5.2.7 : summary of the WGWIDE 2018 assessment and of the assessment run on the subset of the RFID data.

5.2.2 Changes in model configuration relative to young ages

5.2.2.1 Rational for revisiting model configuration

Some of the settings in the mackerel assessment were adopted for technical reasons in the 2014 benchmark, when the SAM model including tagging data was used for the first time. In order to get the model with tagging data to converge with the optimizer used at that time (ADMB), some simplifications had to be made for some of the parameters. In particular, the observation variance for the catches and F random walk variance were not age-specific. During the 2017 benchmark, SAM had moved to TMB, which is more efficient for parameter estimation, but for lack of time the configuration for catch observation and F random walk variances was not revisited. There was more time to do so at IBPMAC 2019, and options for these configurations were carefully scrutinized.

Having a single catch observation variance for all age groups means that the model is expected to fit equally well to the different ages. Inspection of the residuals for the catch at age in recent update assessments shows that residuals for young ages (0 and 1) are consistently larger (in absolute value) than for the rest of the catch-at-age matrix. This suggests that the assumption of equal variance is not appropriate. Catches of age 0 and 1 are mainly coming from discard data, which are less precisely estimated than landings, and for which data are submitted for only part

of the countries. Catch data for 0 and 1 year olds are therefore expected to be noisier than older ages, which are well sampled.

In this context, a likely consequence of using a single non age-specific parameter is that the model fits the catch data for the young ages too well, which means that stock abundance for these young ages probably follows too closely this noisy catch data. On the opposite, the observation variance is probably too high for the older ages (as the estimate is influenced by the noisier data for ages 0 and 1), which means that a tighter fit to the catches should be expected.

As for the catch observation variance, the use of a single F random walk variance is likely to be inappropriate. Since ages 0 and 1 are not targeted by the fisheries, but rather appear as incidental catches, it should be expected that they are subject to a more variable fishing mortality than the targeted age-groups. Here again, having a single non age-specific parameter is a model misspecification, likely resulting a wrong representation of the dynamics of the young age groups, and more importantly, to some extent of the exploited age-groups as well.

Different exploratory runs were conducted, first by decoupling these two parameters between age-groups and then by regrouping ages with similar values. The final model configuration (explained in the section 6), treats separately age 0, age 1 and ages 2 and older both for the catch observation variances, and for the F random walk variances.

5.2.2.2 Consequences of the changes made

When parameters are decoupled, catches observation variance for age 0 is estimated to be very large (close to 1, figure 5.2.8), and variance of age 1 is also much higher than in the assessment with a single parameter. Variance for older ages (2–2) is markedly lower than when all age-groups are coupled. This results in a much looser fit to the catch data for ages 0 and 1 with the new configuration compared to the previous one, but also in a tighter fit for ages 2 and older (figure 5.2.9).

The model with the new configuration has a much lower observation variance (better fit) for the recruitment index. As the model is given the freedom not to trust the catch data for the age 0, it relies much more on the only other source of information available. The low observation variance for the recruitment index does not necessarily mean that the index is very good predictor for recruitment, it just illustrate the lack of any competing information to influence the model. Model fit to the other surveys is not influenced by the change in configuration.

The variances for the processes are also strongly modified by the change in configuration. As the recruitment is mainly driven by the recruitment index, which is less variable than the catch-at-age 0 (previously the main driver of the estimates of age 0), the random walk variance for the recruitment is estimated much lower with the new configuration. Fishing mortality random walk variance for the fully recruited ages (2–12) is lower than in the model with a single non age-specific parameter. This mean that the new model has less variable fishing mortality for the exploited ages (and hence F_{bar} 4–8). Random walk variances increase for age 1 and for age 0 to a level similar to the single non age-specific parameter, but the uncertainty on these 2 parameters appears to be large.

The assessment with the new configuration appears to be more stable than the previous configuration (figure 5.2.10 to be compared to figure 5.2.6). The first 3 peels of the retrospective plot are very consistent for both SSB and F_{bar} . Larger departures are observed for retrospective year 4 and 5 but it is to be expected as they no longer contain RFID data.

The new configuration has modified substantially the observations variances, which results in a weighting of the different data sources which is quite different from the previous assessment

(figure 5.2.11). Removing any of the surveys, or the RFID tags, has much less effect on the assessment (figure 5.2.11 to be compared to 5.2.2). This implies that the assessment is mainly driven by the catch data. Although they still have a noticeable influence on the assessment, the other data sources have a much lower weight compared to the previous assessment. While it may seem regrettable that the assessment is dominated by the catch data, it probably also reflects the many limitations that affects the reliability of the different survey indices available for the mackerel and the tagging data.

The overall perception of the SSB trajectory is not much modified by these changes in model configuration (figure 5.2.12). The estimated F_{bar} is slightly higher in the recent years than with the previous configuration, and notably less variable. The recruitment estimate is also much less variable, and gives a quite different perception of the strength of some year classes than previous configuration (no strong 2014 year-class, no extremely weak 2016 year class).

Caution note on recruitment at age 0 estimated from the new assessment :

At the age at effective recruitment to the fishery (age 2 or 3), the model is informed by better quality catches, survey and tag information. Abundance-at-age 2 and 3 are therefore more reliable than at age 0. Since hardly any fishing mortality occurs before age 2 or 3, abundance estimates at these ages should be very good indicators of cohort strength. However, there are quite some discrepancies between cohort strength as seen at age 0 and at age 2-3 (which are both similar, figure 5.2.13). For example, although the 2002 and 2011 year classes are exceptionally large, they are estimated to be only average or moderately large at age 0. On the opposite, small 2003 or 2015 year classes are estimated average at age 0.

For lack of informative catch data on age 0, the estimated recruitment tightly follows the recruitment index. Since the model has a process error, the estimated cohort strength can be progressively (and substantially) modified between age 0 and age 2 or 3 to match with the perception of cohort strength at these ages.

Practically, this means that any inference about the size of recent cohorts should be made with caution. This may be problematic when it comes to carrying out a short term forecast. Since mackerel is sexually mature early, abundance estimates at age 0 and 1 in the terminal assessment year will contribute for a substantial part to the forecasted SSB in the advice year (then being 2 and 3 years old). This should also be kept in mind when looking at stock-recruitment models for this stock (see section 5.3).

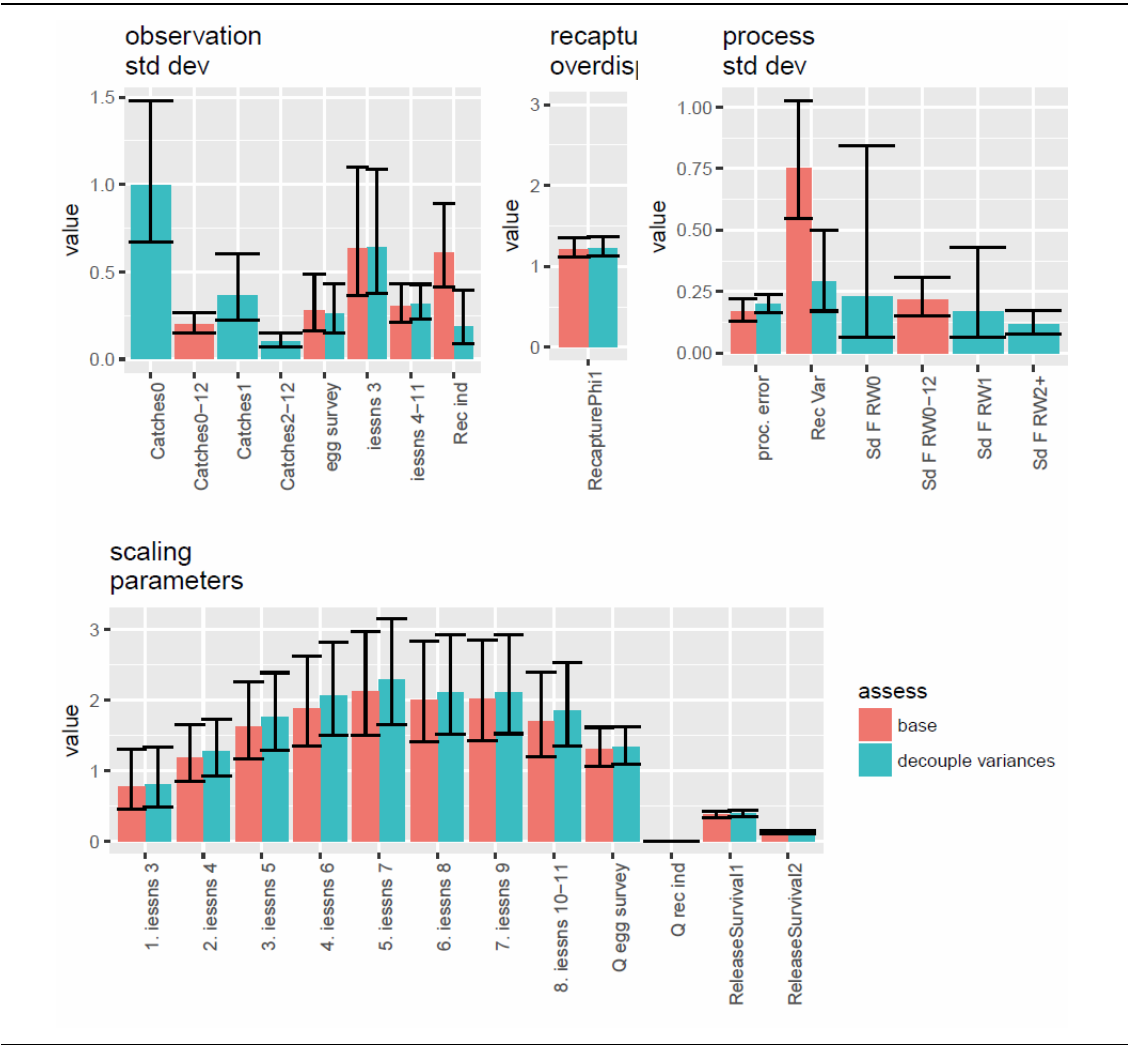


Figure. 5.2.8: comparison of estimated model parameters between the assessment run with the previous configuration and the assessment run with catch observation variances and F random walk variances decoupled.

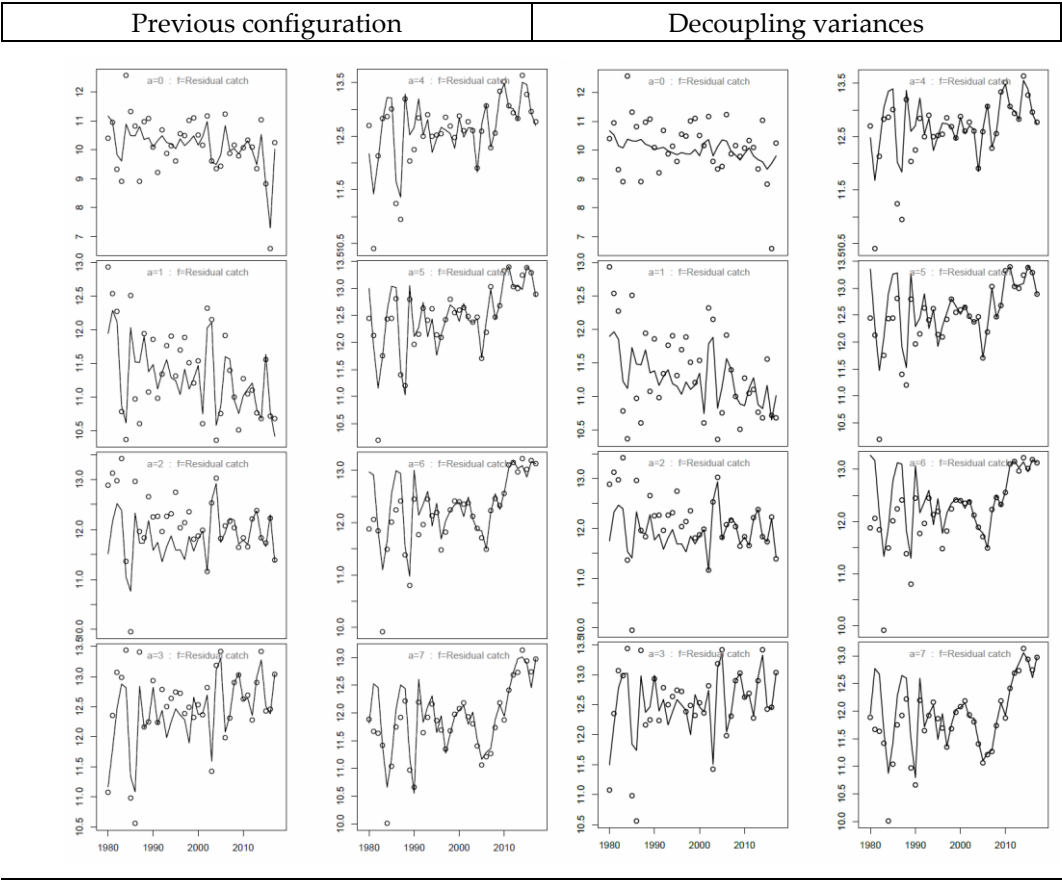


Figure. 5.2.9 : observed vs. predicted catch-at-age data (for ages 0 to 7) for the assessment run with the previous configuration and the assessment run with catch observation variances and F random walk variances decoupled

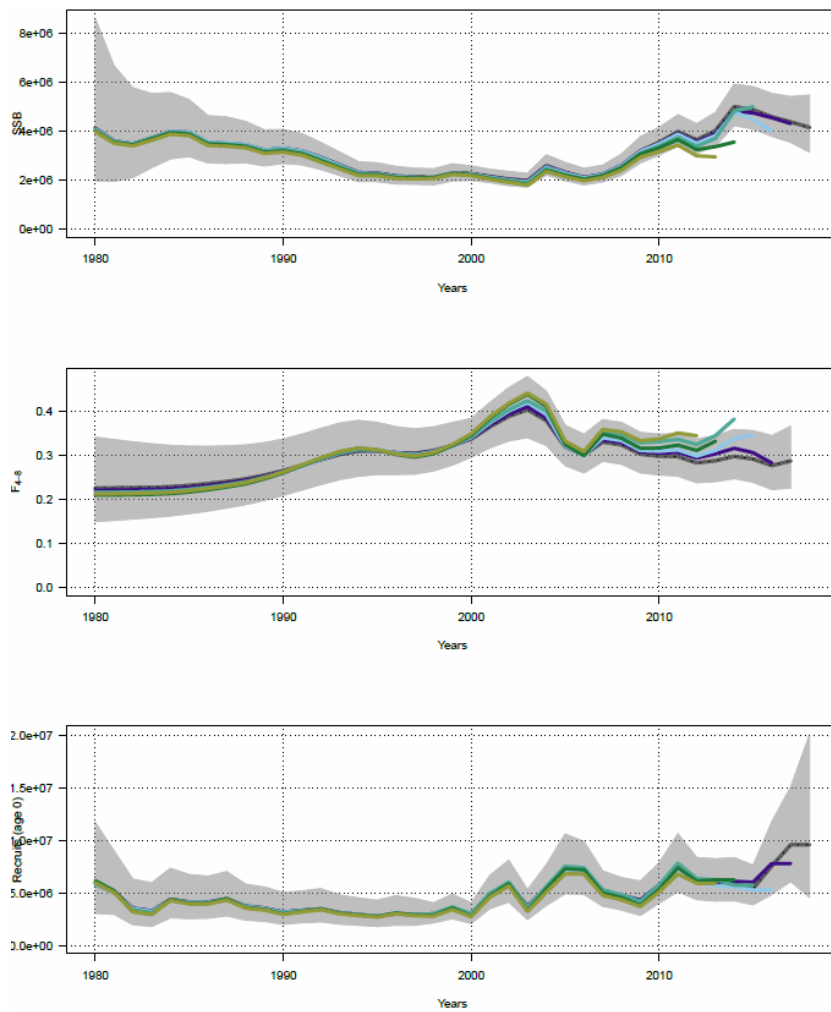


Figure. 5.2.10 : retrospective analysis for the assessment with variances decoupled

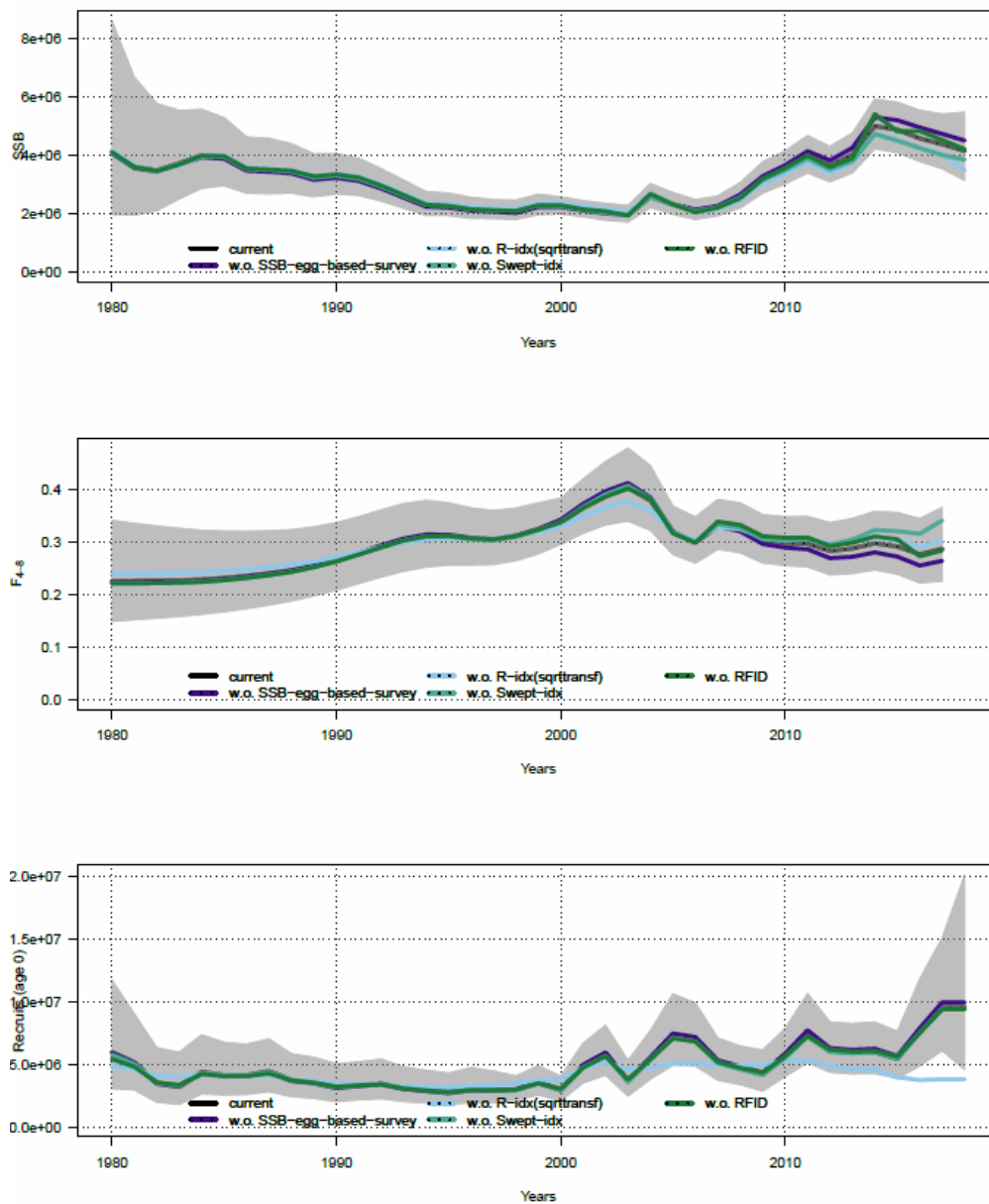


Figure. 5.2.11 : leave one out runs for the assessment with variances decoupled

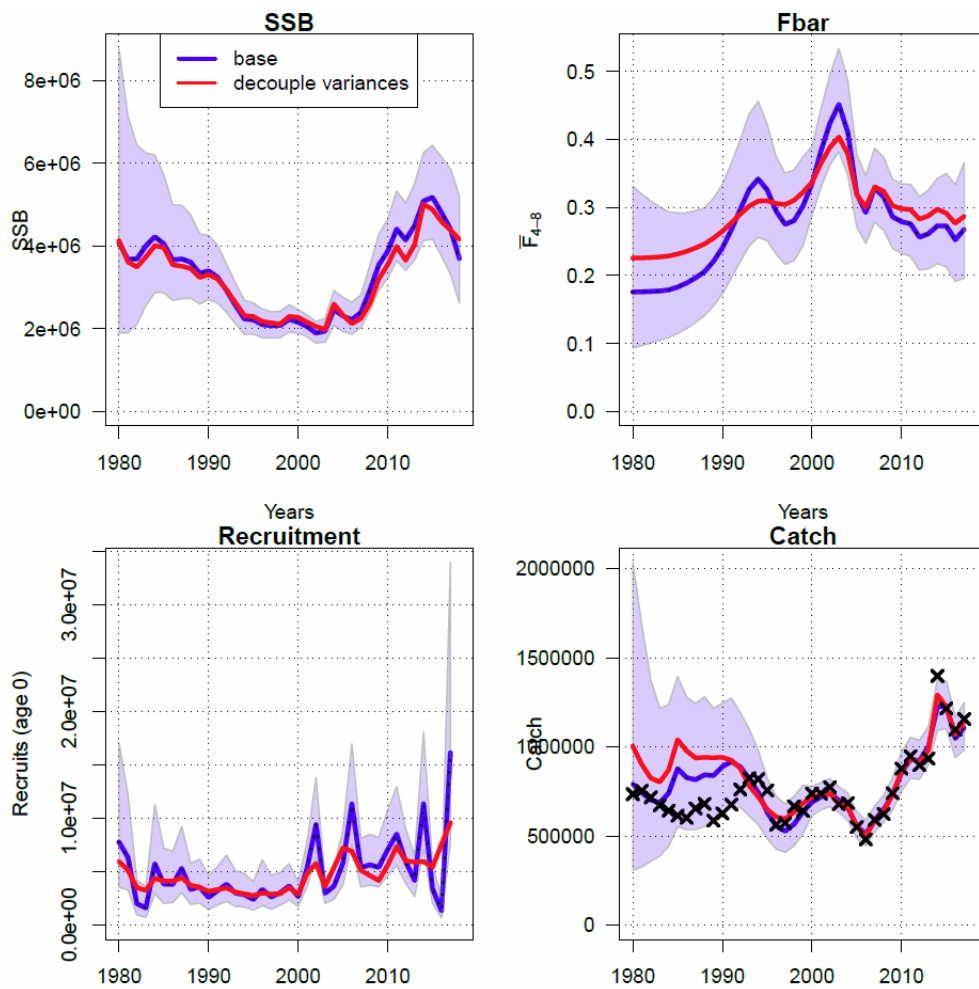


Figure. 5.2.12 : comparison of estimated stock trajectories between the assessment run with the previous configuration and the assessment run with catch observation variances and F random walk variances decoupled

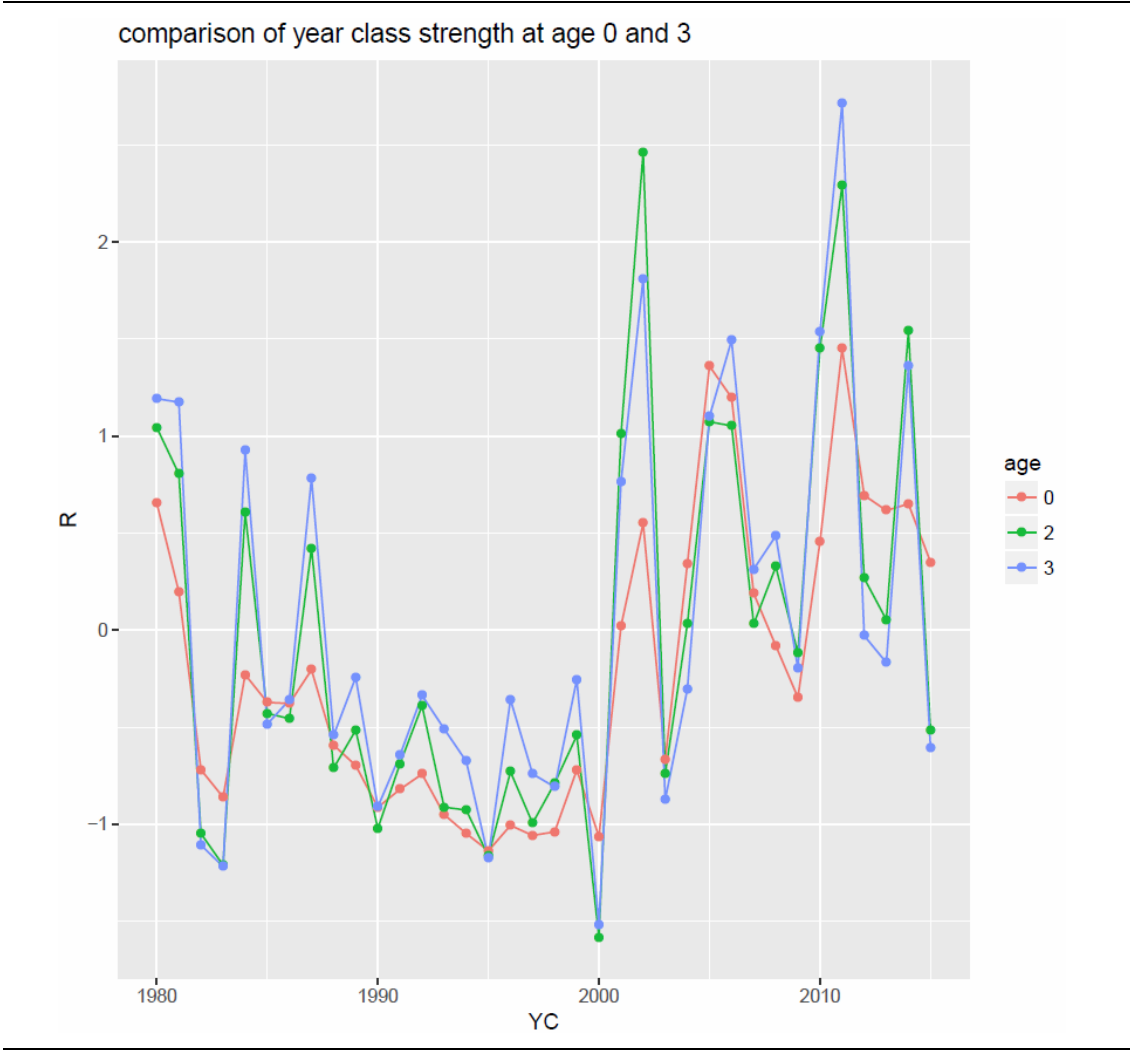


Figure. 5.2.13 : comparison of year-class strength perceived at age 0, age 2 and age 3 (estimated abundance at age standardised, and plotted with respect to the year of birth of the cohort

5.3 Reference points results

The most recent revision of the NEA Mackerel reference points was carried out in 2017 following the benchmark of the assessment (ICES, 2017a) and the management plan evaluation (ICES, 2017b) with the analysis leading to the following reference point estimates:

Framework	Reference Point	Value	Technical Basis
MSY Approach	MSY Btrigger	2.57Mt	B_{pa}
	F_{MSY}	0.21	Stochastic simulation
Precautionary Approach	B_{lim}	1.94Mt	B_{loss} from 2017 benchmark assessment (2002)
	B_{pa}	2.57Mt	$B_{lim} \times \exp(1.645 \times \sigma_{SSB})$, $\sigma_{SSB} = 0.17$
	F_{lim}	0.48	The fishing mortality that, on average, leads to B_{lim}
	F_{pa}	0.35	$F_{lim} \times \exp(-1.645 \times \sigma_f)$, $\sigma_f = 0.20$

Under IBPNEAMac term of reference c, the reference points were re-examined, using the agreed inter-benchmark assessment data and configuration and in accordance with ICES guidelines (ICES, 2017c).

B_{lim}

A plot of recruitments (at age zero) versus spawning stock biomass from the updated assessment is shown in figure x.1. As with the previous assessment, there is no clear relation between stock and recruitment and no evidence of impaired recruitment levels at the lowest observed SSB (a stock of type 5 as described in the ICES guidelines). While there is a cluster of low recruitment observations corresponding to a period of relatively low SSB (2–2.5Mt, 1994–2000), a number of the highest recruitments have also been observed at this level of SSB. The basis for the selection of B_{lim} therefore remains as previously, namely B_{loss} . For the updated assessment B_{loss} occurred in 2003 (2002 previously) although it is at a similar level (1.99Mt).

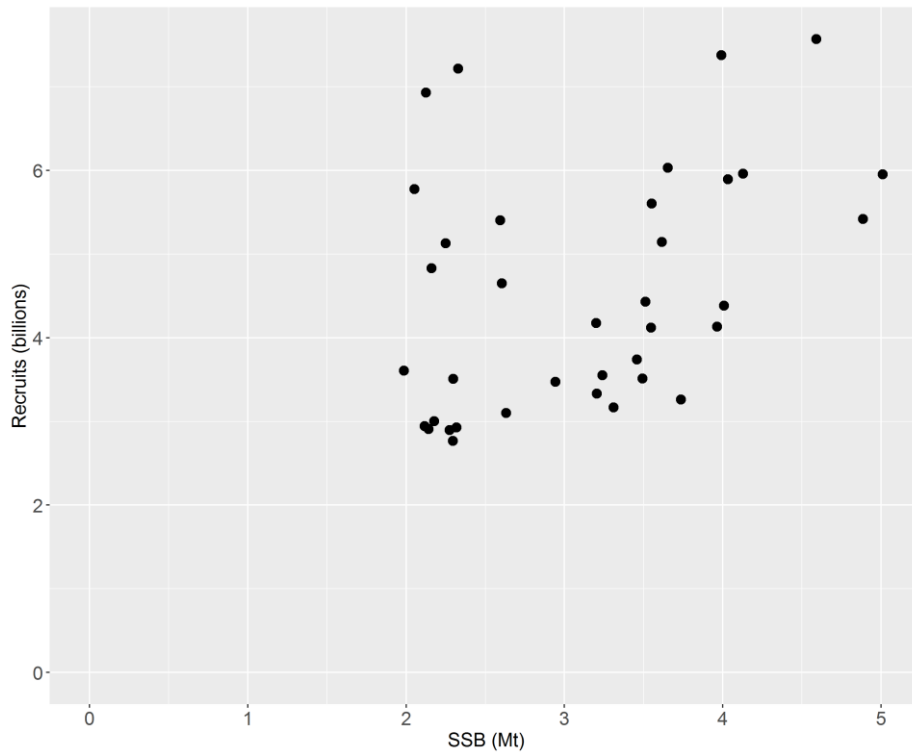


Figure 5.3.1: NEA Mackerel SSB-Recruit pairs from final IBPNEAMac 2019 stock assessment.

B_{pa}

B_{pa} is a biomass reference point above which the stock is considered to have full reproductive capacity, having accounted for estimation uncertainty. It is calculated from B_{lim} and takes the uncertainty in the assessment estimate of SSB into consideration such that when at B_{pa} , the true SSB has a less than 5% probability of being below B_{lim} .

$$B_{pa} = B_{lim} * \exp(1.645\sigma_{SSB})$$

Where σ_{SSB} is the standard deviation of $\ln(SSB)$ at the start of the year following the terminal year of the assessment. For the updated assessment, the estimate available for σ_{SSB} is given for SSB at spawning time in the assessment terminal year and has a value of 0.14, leading to a B_{pa} value of 2.50Mt.

F_{lim}

The precautionary F reference point F_{lim} is defined at the exploitation rate which, on average, leads to an SSB of B_{lim} . The appropriate value for F_{lim} was calculated by stochastic simulation using the ICES *EqSim* software (part of the ICES MSY R package). The simulation was conducted taking into consideration variability in recruitment, biology and selection, each of which is described below:

Recruitment

Due to a lack of reliable catch information before 2000, the catch data is heavily down-weighted in the assessment and model estimates have high uncertainty during this period. Initially, stock-recruit data pairs from 1990 onwards (excluding the most recent year, 2017) were used when estimating the stock-recruitment model for the stochastic simulations, using the non-parametric bootstrap procedure in the ICES MSY R package. However, this resulted in a poor model fit with a significant residual pattern. The dataset was further trimmed to include estimates from 1998-2016 as this was considered more representative of the contemporary situation and resulted in an improved model fit with a 26% probability for the Ricker model, 25% for Segmented Regression and 48% for Beverton and Holt and is shown in figure x.2. Note that the final data point (corresponding to the highest estimated recruitment) is the 2017 estimate. This is not considered by the fitting procedure as it is considered highly uncertain.

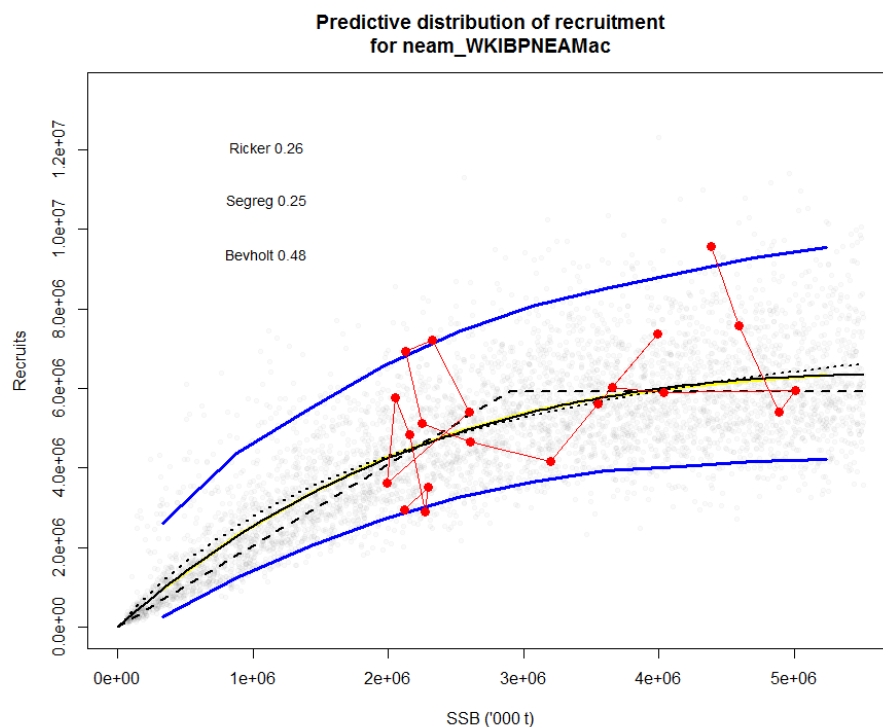


Figure 5.3.2: NEA Mackerel SSB-Recruit pairs from final IBPNEAMac 2019 stock assessment.

The (red) line depicts the temporal recruitment estimates in figure x.2 and indicates that autocorrelation may be a feature of the time series with clusters of recruitments both above and below the median model fit. R's *acf()* function output (figure x.3) indicates a significant correlation at lag 1. The *EqSim* option of including autocorrelation in recruitment was therefore included within the simulations.

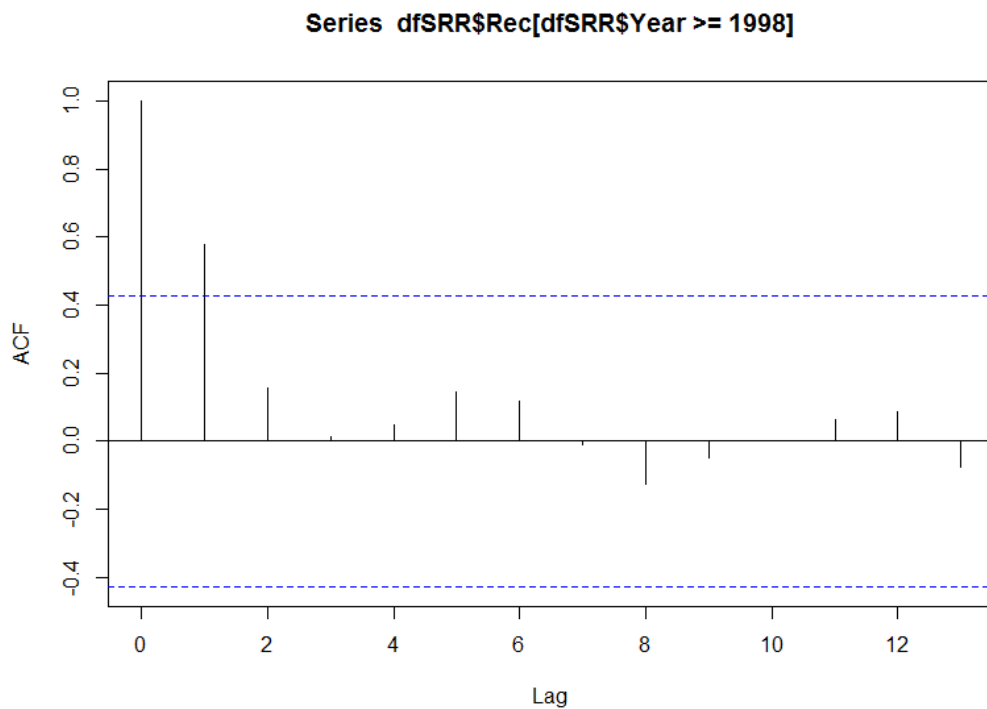


Figure. 5.3.3 – Estimate of the autocorrelation within the recruitment time series (1998-2016) from the R *acf()* function.

Biology (weight at age, proportion mature) & Selection

The default *EqSim* setting for both biological vectors and fishery selection is the most recent 10 years with options to either use the mean over this period or to resample individual years at random for future years. However, there is a clear trend over the past 10 years in both the biological inputs and the assessment estimates of selection and, since the simulations should reflect the current productivity of the stock, this period was reduced to the most recent 5 years with vectors selected at random from the period 2013-2017.

Running *EqSim* for a range of F values with the settings as described above but excluding assessment error and any reduction in F regardless of SSB (*i.e.* without implementing the ICES advice rule), the results indicate an F_{lim} value of 0.46.

F_{pa}

F_{pa} is the estimate of fishing mortality which is designed to ensure that the true F is above F_{lim} with a 95% probability. Its value is calculated based on F_{lim} , whilst taking the assessment uncertainty in F into consideration:

$$F_{pa} = F_{lim} * \exp(-1.645\sigma_f)$$

With an estimate of 0.14 for σ_f , this leads to an estimated value of 0.37 for F_{pa} .

MSYB_{trigger}

Within the ICES MSY framework, MSYB_{trigger} is a biomass reference point that triggers a management response to avoid stock depletion when fishing at F_{MSY} . It is defined as the 5th percentile on the distribution of SSB when fishing at F_{MSY} . For stocks which have been fished at levels greater than F_{MSY} in the recent past, B_{pa} is an appropriate proxy for MSYB_{trigger}. Fishing mortality on NEA Mackerel has been significantly greater than the F_{MSY} estimate for a number of years, and particularly in the most recent period. Thus, the B_{pa} value of 2.50Mt was selected as the appropriate value for MSYB_{trigger}.

F_{MSY}

The estimation of F_{MSY} involves a 4 step process

1. An initial estimate of F_{MSY} from an *EqSim* simulation including stochastic recruitment, fishery and assessment/advice. This gives an estimate of F leading to maximum yield without a constraint on biomass levels (*i.e.* with MSYB_{trigger} = 0).
2. Selection of an appropriate value of MSYB_{trigger}, the biomass which triggers a management response to reduce target fishing mortality (see above).
3. A second *EqSim* simulation as in (1) but also including the MSYB_{trigger} as determined in step 2.
4. Final selection of F_{MSY} is the lower of the F_{MSY} value from (1) or $F_{P.05}$ from (3) where $F_{P.05}$ is an upper limit on F that is considered precautionary.

The estimate of assessment and advice error follows the procedure described in the ICES WKMSYREF3 report (ICES, 2015). This is incorporated in *EqSim* as a two parameter error function applied directly on the target F incorporating the conditional standard deviation in the log domain (F_{cv}) and autocorrelation as an AR(1) process (F_{phi}). The most recent estimates of fishing mortality (from the updated assessment) are compared to those in the annual short term forecasts with realised catches. This leads to estimates of 0.28 and 0.26 for the *EqSim* F_{cv} and F_{phi} parameters.

The *EqSim* output from the initial (step 1) run are shown in figures 5.3.4 and 5.3.5.

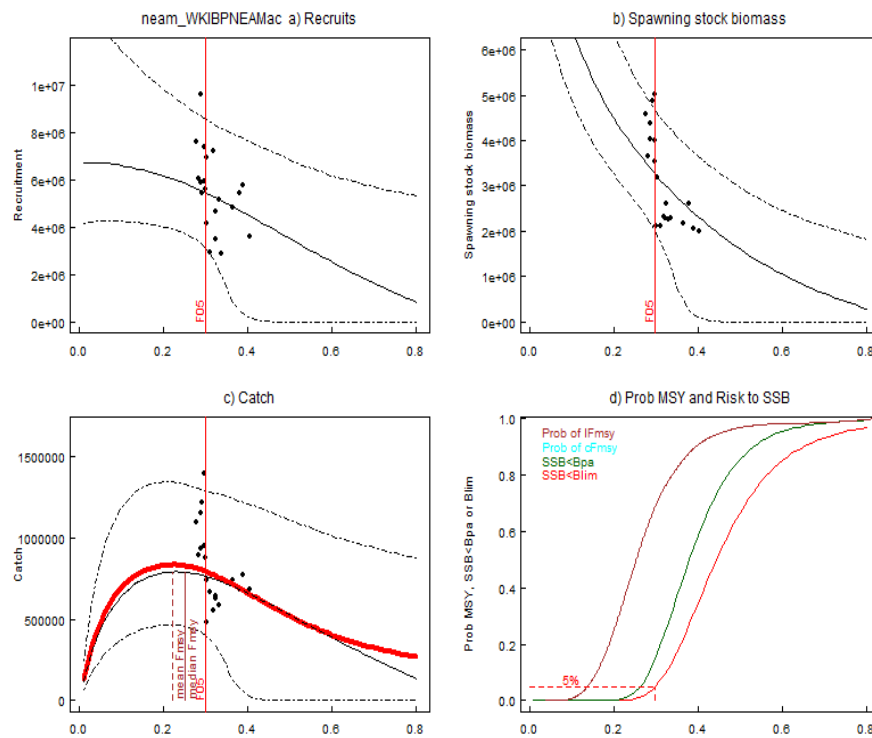


Figure 5.3.4: EqsSim simulated recruitment (a), SSB (b), Catch (c) and Probability of SSB falling below Blim, Bpa (d) in long term equilibrium. Dashed lines depict 5th and 95th percentiles, solid lines the median of 1000 iterations. Assessment estimates of recruitment and SSB and ICES catch shown as black dots on relevant plots. No advice rule implemented (MSYBtrigger = 0).

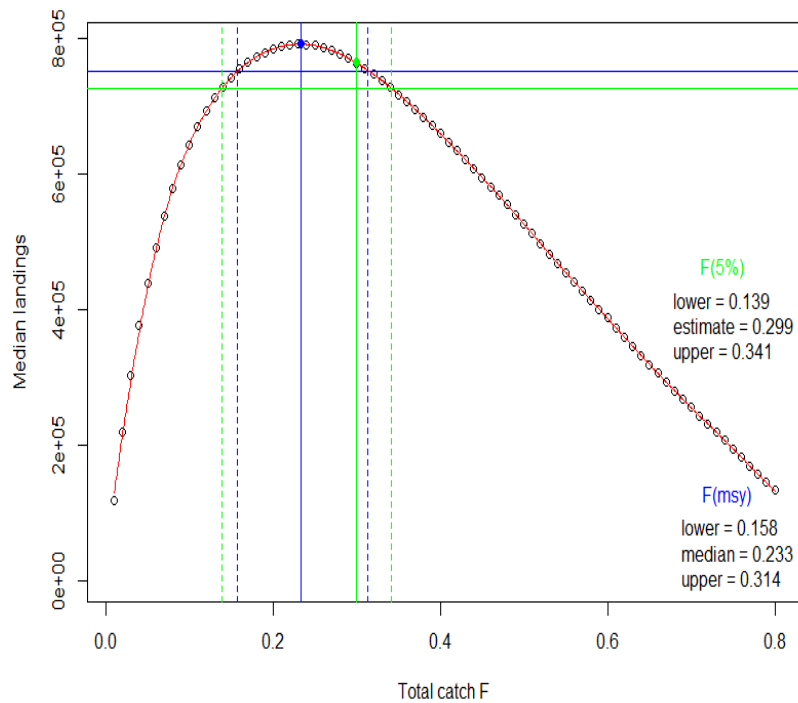


Figure. 5.3.5: EqsSim output. Median yield at long term equilibrium as a function of fishing mortality. FMSY median and range limits shown in blue. Fp.05 in green. No advice rule implemented (MSYBtrigger = 0).

The candidate F_{MSY} , leading to maximum yields of the order of 800kt it 0.23 (range 0.16 to 0.31). Re-running *EqsSim* incorporating the ICES advice rule with an $MSYB_{trigger}$ value of 2.5Mt leads to reductions in target F when the SSB falls below the trigger value. The output from this run is shown in figure 5.3.6.

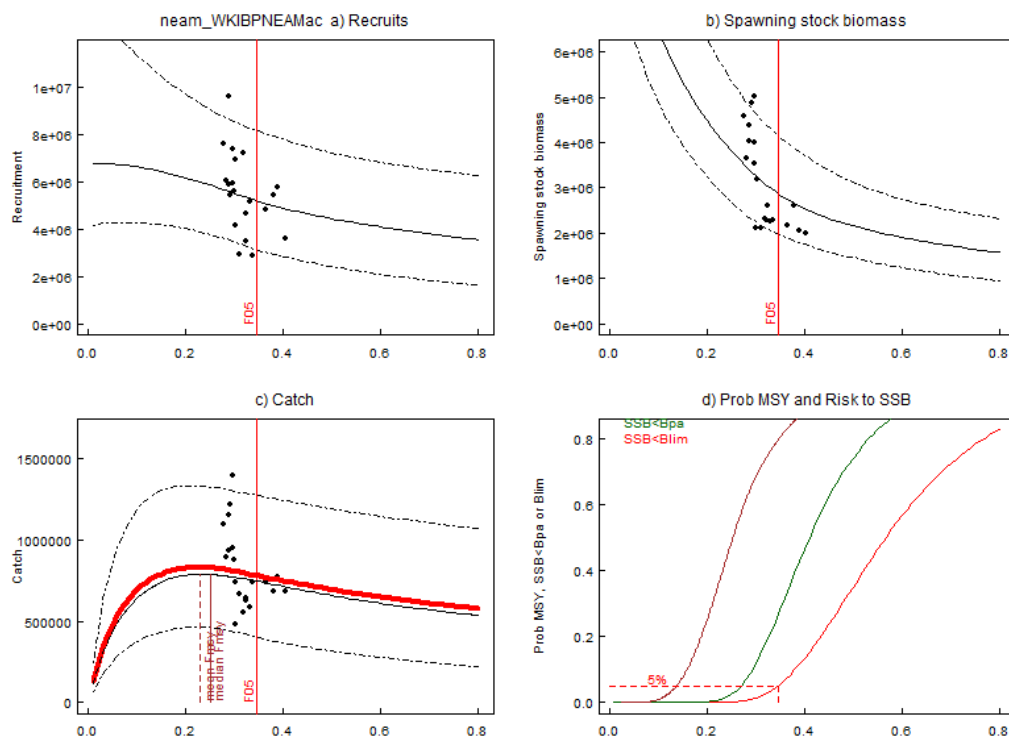


Figure 5.3.6: EqsSim simulated recruitment (a), SSB (b), Catch (c) and Probability of SSB falling below Blim, Bpa (d) in long term equilibrium. Dashed lines depict 5th and 95th percentiles, solid lines the median of 1000 iterations. Assessment estimates of recruitment and SSB and ICES catch shown as black dots on relevant plots. Advice rule implemented with $MSY_{Btrigger} = 2.5Mt$.

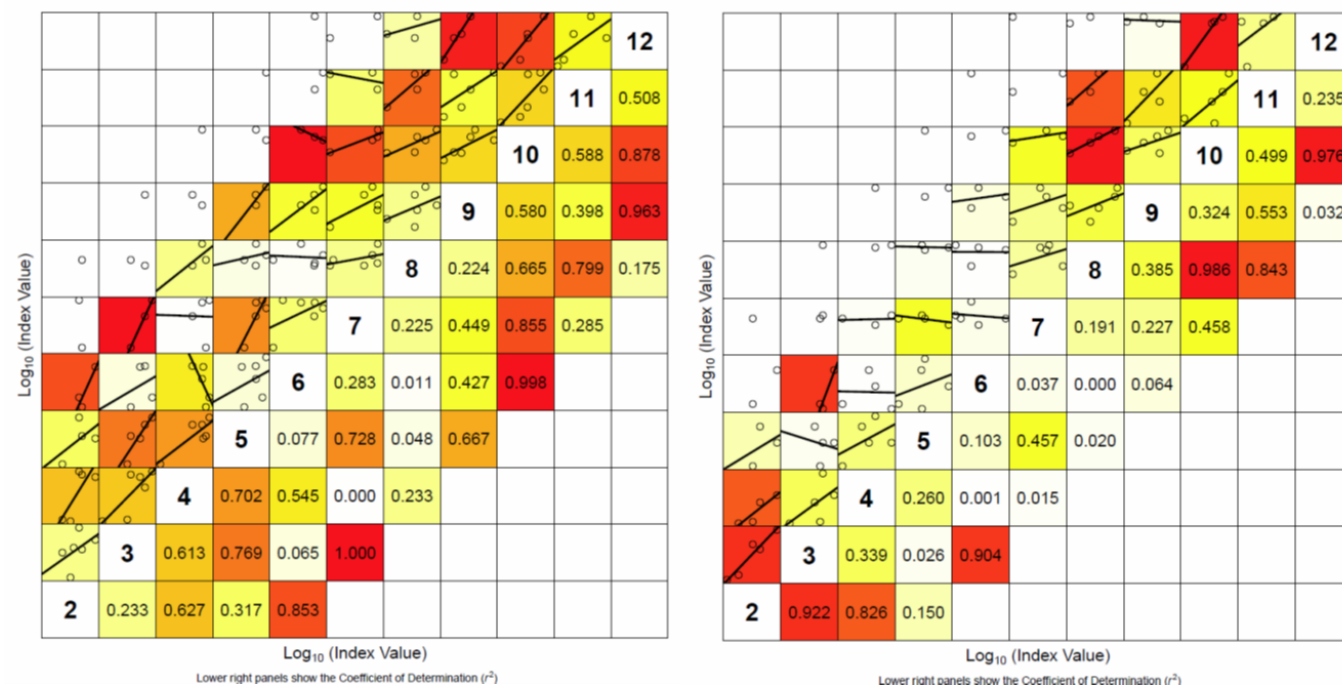
The estimate of $F_{P.05}$ is 0.33 and thus the initial estimate of F_{MSY} of 0.23 remains acceptable as it will not lead to the probability of SSB falling below B_{lim} exceeding 5%.

A summary of the updated reference points is given below:

Framework	Reference Point	Value	Technical Basis
MSY Approach	MSY Btrigger	2.50Mt	B_{pa}
	F_{MSY}	0.23	Stochastic simulation
Precautionary Approach	B_{lim}	1.99Mt	B_{loss} from 2019 interbenchmark assessment (2003)
	B_{pa}	2.50Mt	$B_{lim} \times \exp(1.645 \times \sigma_{SSB})$, $\sigma_{SSB}=0.14$
	F_{lim}	0.46	The fishing mortality that, on average, leads to B_{lim}
	F_{pa}	0.37	$F_{lim} \times \exp(-1.645 \times \sigma_f)$, $\sigma_f=0.14$

6 Summary of decisions (2nd meeting)

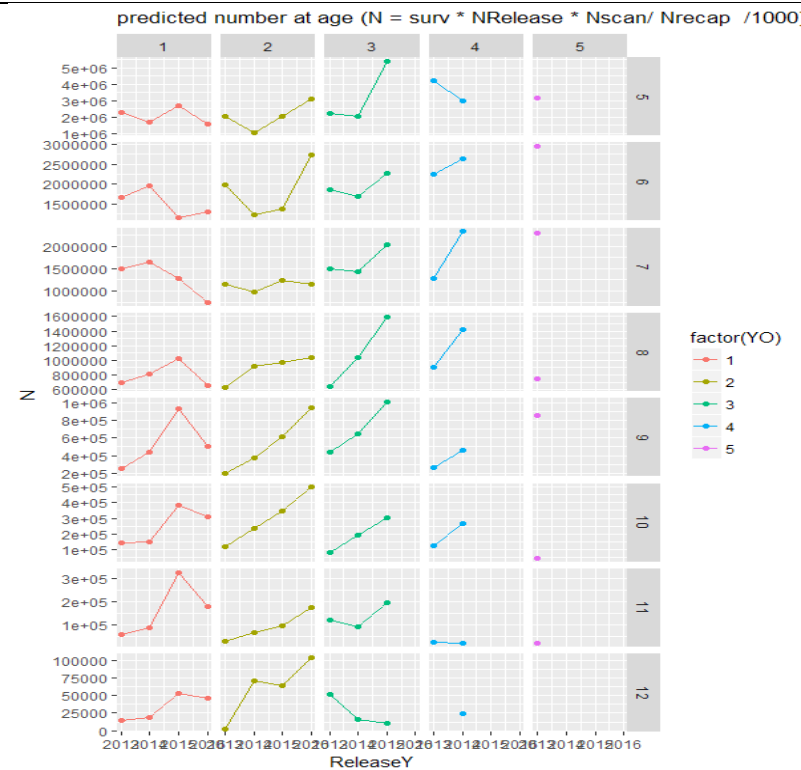
Data decisions	Topic of the decision	Decision made	Scientific support for the decision
#1	Maximum number of years tagged fish spent at sea before recapture, data to be included in the assessment	Using only recapture after 1 and 2 years at liberty	<p>This was decided as a pragmatic way to 1) avoid possible bias due to tag loss, and 2) reduce the size of the data set (and hence its influence on the assessment), 3) add a constant number of observations at each update assessment.</p> <p>Abundance index data estimated from the RFID tags aggregated over different age groups demonstrated the clear tendency of increased estimated abundance in release years with numbers of years between release and recapture, and that this tendency was most evident in release years 2011-2012 and for the estimates including the youngest fish (WD 12). This was also demonstrated at the un-aggregated level where predicted abundance by age in a release year from the SAM model tended to increase over time as more recapture years were added, also more evident trends in release years 2011-2012 and at the youngest ages (WD 20). To test if this bias could be related to mixing issues, the predicted abundance by age in a release year from the SAM model was tested using the RFID time series split by the 4 recapture areas (WD8). The results suggested that within recapture areas the trend with increasing abundance predicted with number of years between release and recapture was less evident, supporting that this bias could be related to mixing issues rather than tag loss or long-term higher mortality of tagged versus non-tagged fish.</p> <p>The main reason for excluding recapture years rather than trying to accommodate for the potential bias in the assessment model, was the uncertainties related to reasons behind the observed bias trends, and how this may develop in near future. Main conclusion was that we need more knowledge and more time before including all recapture years and adjusting within the model.</p> <p>The decision to exclude recaptures being out more than two years were also supported from analyses of internal consistency in the data. Analyses of the RFID data (recaptures expressed as cohort abundance at release time) conducted on the whole data set showed the data from the tag recaptured after 1 year was relatively informative (internal consistency analysis showed that it was to some extent possible to follow cohort strength in successive age-groups, WD 20). Data after 2 years at liberty was less informative, and for subsequent numbers of years at liberty, completely uninformative.</p>



It was therefore concluded that the data from the tags having spent 1 year at liberty was the most appropriate to be used in the assessment (as providing reasonably good information, and as subsequent numbers of years at liberty were of lesser quality and correlated poorly with this best dataset).

However, after applying decisions 1 and 2, and conducting again the consistency analysis on the resulting subset of the data, there was no longer any clear preference for 1 year compared to 2 years at liberty (but the analysis was limited by the low number of year remaining in the data). It was further observed that for age 5 to 7 (rows on figure below), data from 1 and 2 years at liberty (columns on figure below), had a similar level (i.e. mean across years), but that this level increased for 3, 4 and 5 years at liberty. This was interpreted as the influence of tag loss, or any other source of bias potentially resulting in a decrease in recapture rate with the number of year at liberty.

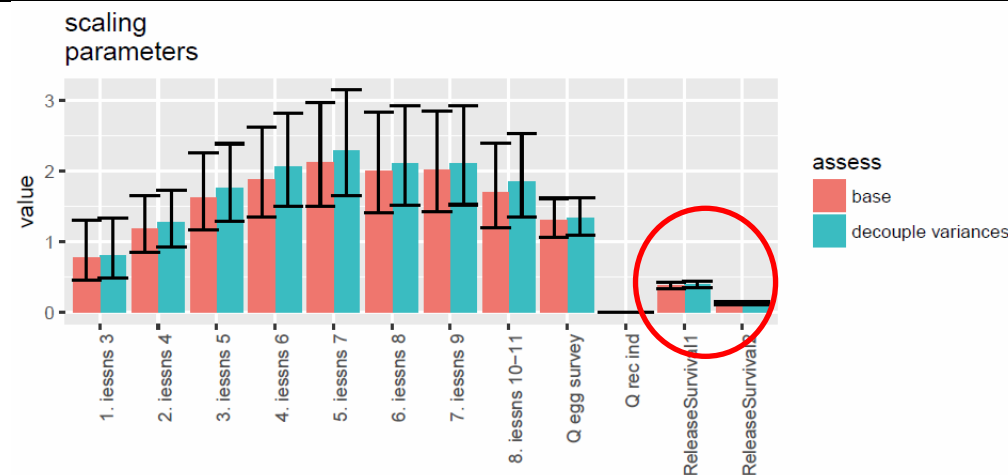
In order to avoid introducing this potential bias in the assessment, the final decision was to use only 1 and 2 years at liberty.



#2	Age classes in the tagging data to include in the assessment	Use only age at release ≥ 5	<p>Few age 2-4 fish are tagged in the RFID program.</p> <p>This may possibly 1) result in more uncertain recapture numbers (since tag concentration is low), and 2) strongly violate the assumption of complete mixing as we might be tagging only a very specific component of the first time spawning fish.</p> <p>In the steel tag time series age structure of tagged relative to scanned fish was rather similar, both young and old fish were tagged and scanned (WD 24). However, in the RFID time series very low numbers of fish at ages 2-4 has been tagged compared to the large numbers scanned of these ages. This means that the concentration of tags in these age groups of strong incoming year classes are very low, and therefore noisy. With the expansion of mackerel happening in recent years it seems that strong year classes such as the 2010 and 2011 have had a northerly distribution looking at the catch at age by area, not really migrating down Ireland to spawn, but staying more in the North Sea, Norwegian Sea, Icelandic area until ages 5-6 years (WD 22). Recapture maps of young mackerel tagged at body length<33 cm (typically ages 2-4) off Ireland, suggest that they are recaptured closer to the main tagging (WD 26). This suggest that the tagged fish of these age groups 2-4 are not representative, not mixing fully in areas where these age groups have been fished and also surveyed with the IESSNS survey (WGWISE report 2018). Support for the decision to exclude ages 2-4 the RFID tag data series is also found in a work looking at total mortality by age using the Browney tag</p>
----	--	----------------------------------	--

			recovery model (WD 27) where these young ages appear to have lower survival than the older fish, which likely is a result of lower recapture rates due to incomplete mixing.
#3	Incorporation of tag-loss in the mathematical formulation of the assessment model	No tag loss in the model formulation	<p>Exploratory runs were conducted with tag loss (or malfunction) implemented in the model formulation for expected recaptures (as an exponential decay function). The model gave acceptable diagnostics. The estimated tag loss was small for the steel tags (0.03 [0.01-0.08]) but higher for the RFID tags (0.13 [0.09-0.19]).</p> <p>However, it was decided that it was premature to incorporate tag loss (or malfunction) in the assessment because :</p> <p>There is no firm proof of tag loss or tag malfunctioning yet. A number of experiments have been suggested to test whether this indeed happens for the RFID tags in mackerel (see annex 4 recommendation)</p> <p>Literature on RFID tag loss on salmonids indicates that it is mostly a problem for females at the first reproductive event after tagging, as the tags (placed in the abdominal cavity) can be expelled with the eggs. If this is the case for mackerel, most of the tag loss would occur between release and the first recapture year (although some tags may be recovered before the spawning season). Tag loss would then be confounded with post release survival, and should not influence recapture rates in the subsequent years</p> <p>The main indication for tag loss in mackerel is the decrease in recapture rates for a given tagging event after many years at liberty (see WD 20, WD 4). However, inspection of the recapture data from 2018, made available during the IBPNEAMAC 2019 showed that the 2018 recaptures do not comply with this pattern observed across earlier recapture years (WD 20). In addition, when looking at recapture rates in different geographical areas, the decrease with number of years at liberty is also no longer visible. Therefore, there is also no firm indication from the data that tag loss is a widespread problem.</p>
#4	Release years of RFID tagging experiments to be included in the assessment	Use data from tags released in 2013 and after	<p>The number of factories scanning catches has increased at the start of the RFID program. Before 2014, there were less catches scanned, originating from a more restricted geographical area (see WD 16, WD 26)</p> <p>Recent recaptures from tags released in 2011-12 give a markedly higher perception of cohort size at release time than recaptures from the start of the RFID program (e.g. 2012 and 2013). This could be linked to tag loss/malfunctioning, but also to the changes in the geographical origin of the catches scanned, in case of non- complete mixing.</p> <p>2013 is the first release year for which all recaptures were scanned with the current coverage in factories scanning for tags.</p> <p>The way the tag data are used in the stock assessment one assumes representative mixing of tagged fish within a year class, i.e. the recapture rate within a year class should be the same regardless of catch area. In WGWIDE 2018 this assumption was tested by introducing the RFID tag data into the SAM model as 4 different time series, based on recaptures in 4 different areas; West of UK/Ireland (Q1), North Sea (Q3-Q4), Norwegian Sea (Q3-Q4) and Iceland (Q3-Q4). This exercise demonstrated that the scaling parameter “survival” varied between the time series, being lowest in the Norwegian Sea area, suggesting that part of fish tagged at the spawning grounds off Ireland did not migrate up in this area, leading to lower recapture rates. This observation led to the conclusion that recaptures from release years 2011-2012 were from a different catch distribution, and estimation of abundance in release years 2011-2012 may therefore not be comparable to estimates from later release years 2013 and onwards. To reduce the potential bias effects of related to temporal and spatial changes in scanning, it was suggested to exclude release years 2011-2012 from the assessment.</p>
#5	Treatment of survival estimates between the steel and RFID tag time-series	Model with 2 different survival rates	The assessment run with a single survival rate provided an acceptable fit to the observations (some small patterns visible in the residuals, but also present in the residuals of the assessment with separate survival rates). There were, however, large negative process error deviations on abundances observed between 2011 and 2015, which indicates that the model consistently removed fish biomass from the stock at the time the RFID data come into the model. The model with a single survival rate also had a severe retrospective problem. Finally, the model with two separate survival rates give parameter estimates which are significantly different.

in the assess-
ment



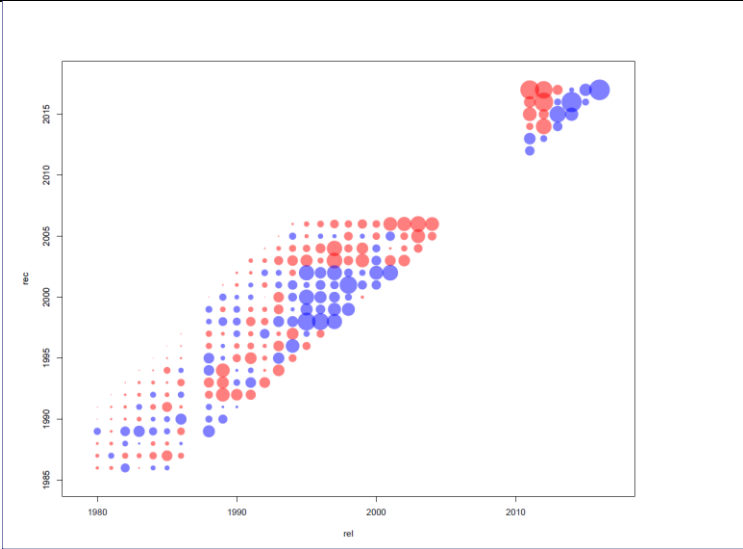
The reason for these two very different survival rates estimated for the steel tags and the RFID tags (around 40% and 10% respectively) was discussed at length during IBP MAC 2019. Experiments conducted recently have shown that the old way with manual jigging of mackerel and release at starboard side, gives the same recapture rates as the new way with automated jigging machines and fish releases through pipes at the port side. This suggests that there should not be any substantial deterioration of survival for the new tagging program.

Using the term “post release survival” for this scaling parameter probably introduced some confusion. This parameter appears in the model formulation to represent the disappearance of the tagged fish between release and the first recapture year. Any reduction in the efficiency of the detector systems at the factories, would lead to reduced recapture rates, and this would ultimately result in a lower “survival parameter” estimated. However, recent large scale experiments of the efficiency of the factories scanning landings of mackerel for RFID tags, do not suggest that the efficiencies are lower than in the steel tag times series (WD Issues regarding the updated version of RFID data 2019), so this could not explain the differences in estimates “survival” between the times series.

One difference between the steel tag tagging and the RFID tagging is related to the position of the inserted tag. During steel tagging, the tag was inserted into the muscle of the mackerel if gonads were large and ripe or running, whereas in less ripe or spent fish tags were injected into the abdomen. During RFID tagging all tags are injected into the abdomen regardless of maturation status, due to food safety issues. Experiments suggest that fish tagged in abdomen have higher tag loss than fish tagged in the muscles, and the tag loss typically happens at high rates during the first spawning event. RFID tag loss have been reported to be very high up to 40% for females in trout. Since RFID tags are more likely to be expelled at the first spawning after tagging, one should automatically expect a lower “survival”.

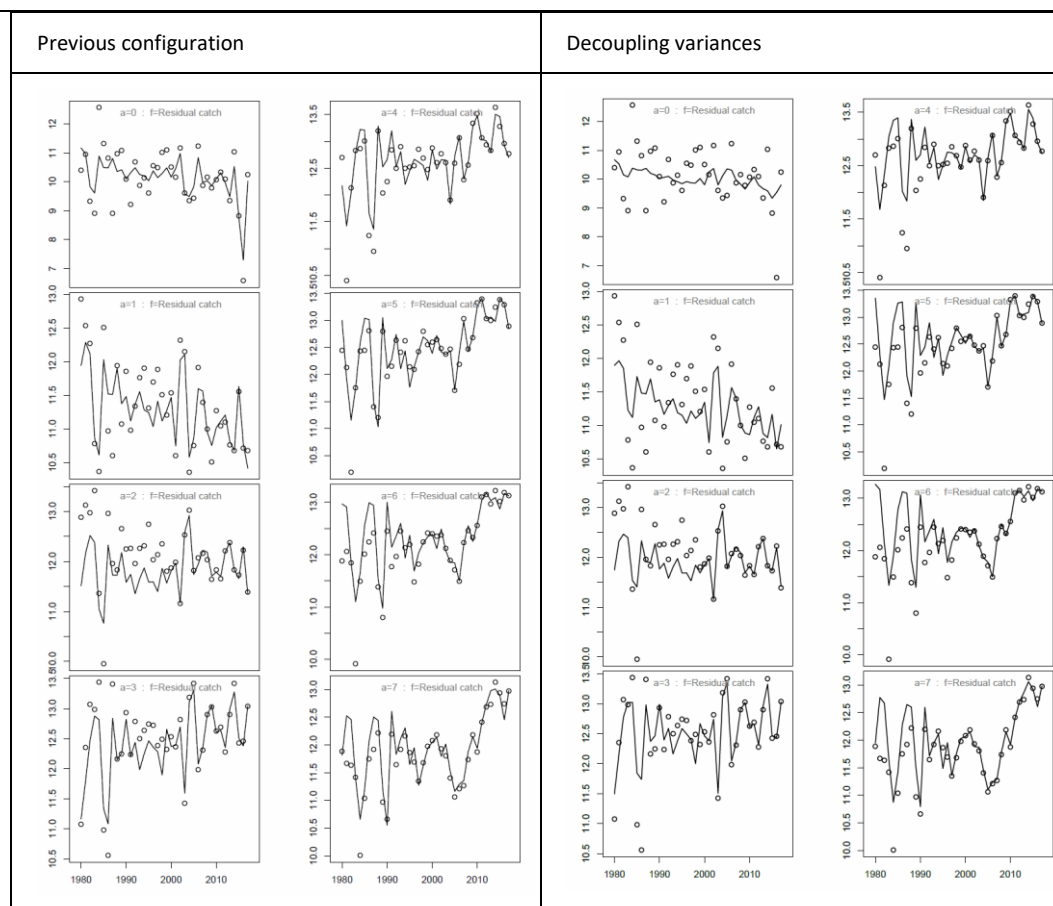
Besides that, a systematic bias in any of the variables in the tagging data set would be confounded with the “survival” parameter. For example, when running the model with 4 different survival rates for the steel tags and a 5th rate for the RFID, high survival is estimated for the period prior to 2000 (0.438 and 0.477), declining afterwards to similar values in the latest steel tags (0.081) and in the RFID tags (0.10)(WD 8). One of the explanations proposed was that there was a consistent underestimation of the volume of the catches scanned in the early years (before stricter control was implemented). This negative bias in the number scanned was compensated by an over estimation of the survival rate. Likewise, any bias on the numbers recaptured (e.g. underestimation due to poor scanning efficiency in the factories), will be confounded with the estimated value of survival rate (in this case lead to an underestimate).

#6	Treatment of tagging data as single observations, aggregated by recapture year or as index of abundance	As single observations	<p>Tagging data can be included in different ways in the assessment. Currently, the tagging data is included as single observations of recaptures by year and age, assuming these data follow a negative binomial distribution. The drawback of this approach is that the number of data points added each year is increasing because recaptures are generated over several different cohorts. Until the recapture data would cover all the relevant age classes, this would lead to increasing weight in the assessment relative to other sources. In addition it is difficult to estimate the over-dispersion (i.e. uncertainty in the tagging data) parameter in SAM when a short, small dataset is used.</p> <p>A second approach would be to aggregate recaptures across recapture years (WD Magne). This reduces the number of data points and results in a constant increase in the number of data points with each additional year of data (similar to a survey index). Data from old tags would have a low contribution to the expected number of cumulated recaptures as their number scanned are low (old fish) compared to earlier recaptures. Likewise, their contribution of the observed sum of tags recapture is low (few recapture). This should in principle minimise the bias due to tag loss/malfunctioning. A drawback to this approach is that it is unclear whether a negative binomial distribution is still appropriate for the sum of recaptures and that the assessment had a strong retrospective pattern.</p> <p>The third approach would be to calculate an index of abundance by age (WD Vatnehol). Tag information and other data sources would be used in a consistent manner, facilitating model interpretation. However, this method gives somewhat different result depending on the mathematical formulation that is being used. In addition, there were strong temporal pattern observed in the residuals for the RFID tag index, indicating that a correlation structure should be used for the errors on this index.</p> <p>After full consideration of the different options, the conclusions was that we should stick with using tagging data as single observations, because the other potential methods had more drawbacks in the model than the single observations. The number of observations to be used in the model could be modified by selecting certain relevant ages and years out.</p>
#7	Correlation structure of swept area survey	Use AR1 correlation structure as in base case assessment	The decision made at the 2017 benchmark (ICES, 2017a; Sánchez <i>et al.</i> , 2017) was revisited using the same criteria as in 2017 (WD 9). The conclusion is unchanged That is, trial runs using auto-correlated observation errors of the IESSNS show that the data are not independent, and that using a model assuming independence would be statistically incorrect.
#8	Correlation structure of the tagging data	No correlation implemented	Different configurations were tested for the inclusion of correlation in the tagging data (random effect on the numbers recaptured structured by release year, by recapture years, by combination of recapture and release years (Presentation A. Nielsen Correlation in tagging part, WD 28).



These different runs all presented a strong retrospective pattern, which was not considered acceptable. In addition, inspection of the values of the random effects (for the model with random effects on release year only) showed strong temporal pattern, suggesting that some variations in the population were absorbed by these random effects (although another possible explanation was a time varying bias in the tagging data or in the survival).

#9	Variance of F random walks	Estimate 3 parameters: for age 0, for age 1 and for age 2 and older.	Both changes (#9 and #10) in configuration were investigated to allow the model to treat the two first ages differently from the older ages. The rational for doing that is that the catches of ages 0 and 1 are very small, and likely to be of poorer quality than older ages which are partially or fully recruited to the fishery. Indeed, age groups 0 and 1 mostly correspond to discards, which are not estimated as accurately as landings, and for which only some of the countries provide estimates. Therefore it seemed more appropriate to model the catches of these age groups with separate observation variances and fishing mortality random walk variances (both likely to be higher than for older age groups).
#10	Observation variance of catch data	Estimate 3 parameters: for age 0, for age 1 and for age 2 and older.	The model with separate random walk variances only estimated a very high F variance for ages 0 and 1, which allowed the model to fit very closely the catch data for these age groups. The F variances for the older ages was then estimated lower that for the model with a single parameter, and the overall observation variances for the catches was smaller than in the base case.

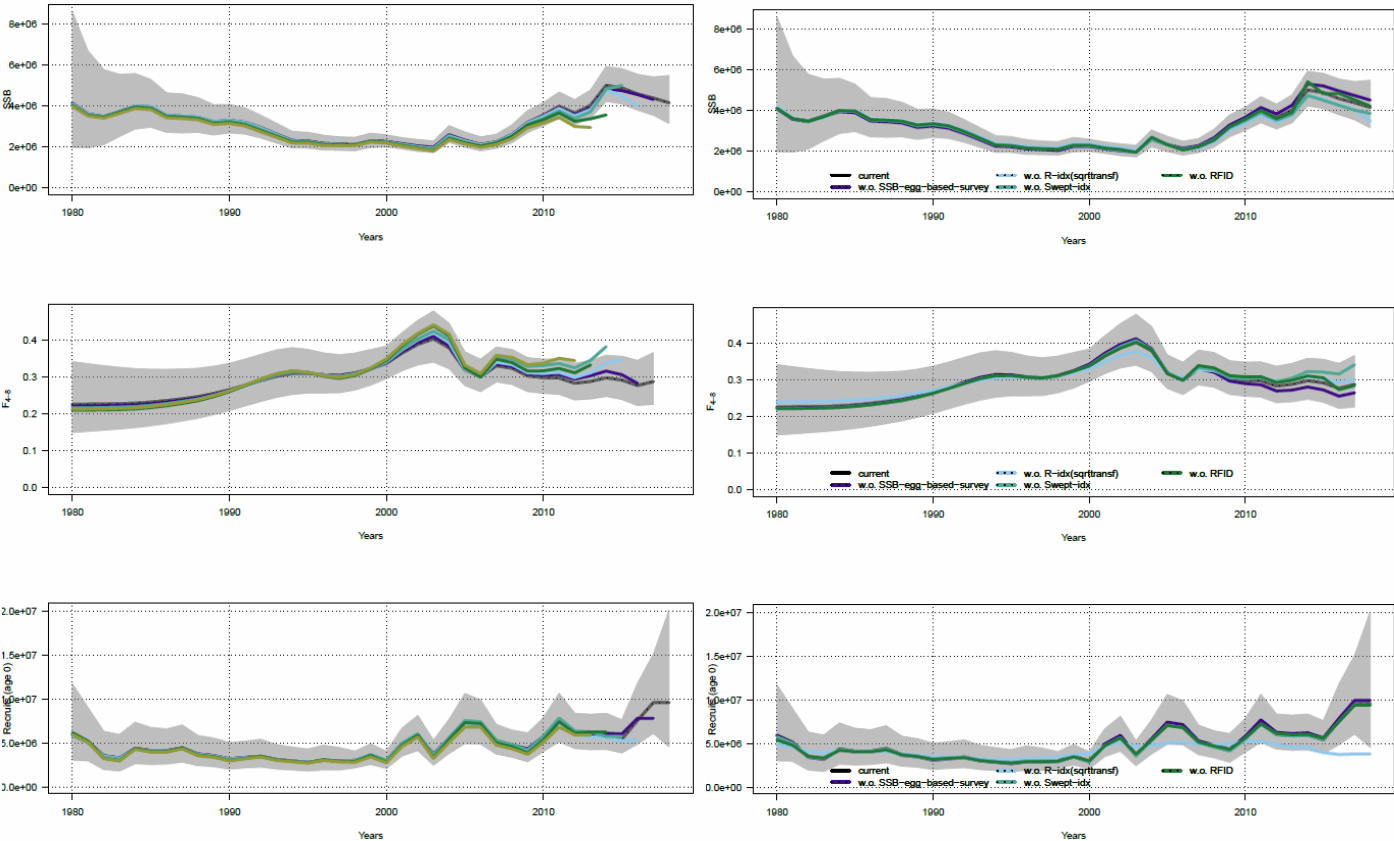


The model was then run with separate observation variances for the catches of 0 and 1 year olds. The observation variances for catches age 0 and 1 were estimated to be high, and estimates for the subsequent ages was lower than in the base case. By allowing the model to produce a poorer fit to the catches of ages 0 and 1, the model produced a better fit for the rest of the age-groups, which represent most of the population. A consequence of this improved fit to the catch information is the lesser influence of the egg and swept area surveys, and of the RFID data, as shown by the smaller sensitivity to the removal of any of these data sources compared to the base case assessment.

The final run, which was adopted as the final assessment, combined the decoupling of the F random walk variances and of the catches observation variances. This run was similar to the run with only observation variances decoupled, but with slightly higher F variances for ages 0 and 1 than for older ages. Residual plot for this runs showed some small improvement compared to the base case run (now residuals of similar size for all age-groups) and had slightly more consistent retrospective plots. Further tests indicated that the model had properly converged (jitter runs) and was robust (simulation study).

In the three scenarios evaluated, the recruitment was no longer informed by the catches of young fish, but instead the model closely followed the variations from the recruitment index (with a much lower observation variance than in the based case). This largely modified the perception of the historical recruitment dynamics for this stock. This should not be taken as an indication that the recruitment index is a near perfect predictor of recruitment; it simply reflects the fact that when recruitment is no longer forced to follow the catches, it completely follows the only other source of information available.

Leave-on-out runs as well as retrospective runs show larger consistency from year to year when dropping individual datasets or trimming off years of data.



#11

Reference points
New reference points were calculated

See section 5.3 for a full description of the procedure followed to calculate reference points. Final reference points were estimated as:

	Framework	Reference Point	Value	Technical Basis
	MSY Approach	MSY Btrigger	2.50M t	B _{pa}
		F _{MSY}	0.23	Stochastic simulation
	Precautionary Approach	B _{lim}	1.99M t	B _{loss} from 2019 interbenchmark assessment (2003)
		B _{pa}	2.50M t	B _{lim} x exp(1.645 x σ_{SSB}), $\sigma_{SSB}=0.14$
		F _{lim}	0.46	The fishing mortality that, on average, leads to B _{lim}
		F _{pa}	0.37	F _{lim} x exp(-1.645 x σ_f), $\sigma_f=0.14$

7 Reflection of reviewers

ICES Inter Benchmark Assessment of Northeast Atlantic Mackerel

External reviewers: Noel Cadigan, Lisa Ailloud and Gary Nelson

General Comments on Workshop Process

The review panel members felt that their roles and responsibilities in the benchmark were not clearly defined. Although access was given to the SharePoint where analytical documents were stored, there was not an organized set of documents that directly provided information needed to thoroughly review the sampling, analysis and modelling methods for the inter-benchmark assessment.

The external reviewers did feel that the attitudes and conversations among stock assessment members were cordial and there were no strong politically-motivated agendas pushed by anyone. The external reviewers appreciate the efforts of the working group members to respond to all requests put forth.

Comments on Stock Assessment

TOR 1. Review the use of tagging data in the assessment by considering and conclude on a definite model configuration with respect tagging

The external reviewers thought the use of tag data in the model is appropriate

a) Bias in tag data collection

- i. In 2013, there was an expansion of fisheries plants used to detect tags. Stock assessment analysts believe this may have increased bias in rate of tag returns. As a solution, the committee elected to remove the release data from 2011 and 2012. Although considered a pragmatic short-term solution, the review panel recommends that other analytical methods (e.g., spatial analyses, comparison of cohort size estimation based individual recaptures without the new plants that came on-line in 2013, etc.) should be used to determine if bias actually exists.
- ii. There was evidence of tag loss/non-mixing/migration/tag malfunction such that the tag density is not a constant rate over time which suggests some type of bias in recaptures. The stock assessment committee attempted to remove this potential bias by using just the returns from the first two years of recaptures. Although considered a pragmatic short-term solution, conducting long-term double tagging experiments (i.e., sea pens, etc.) is recommended to resolve this issue. In the future, incorporating tag loss/migration functions in the stock assessment model should be fully explored to allow the remaining years of recapture years to be included. To investigate potential non-mixing, analyses of the tag data using the instantaneous rates model of Hoenig *et al.* (1998) or the simpler method of Latour *et al.* (2001) should be considered. Software for Hoenig *et al.* method is available at <https://www.nefsc.noaa.gov/nft/IRATE.html> or from Gary Nelson.
- iii. The stock assessment committee chose to remove release ages <5 because they felt their sampling does not fully represent the age structure of the catch and that the mixing assumption was more likely violated at the young ages. The review panel considers this a source of uncertainty but not necessarily bias.
- iv. Ageing error – The external reviewers suggest conducting simulation analyses exploring the impact of ageing error and plus groups on the model results. Also, reviewers suggest exploring the use of age error matrices to deal with potential age differences among laboratories. There is an important difference between the way ages were determined for

the steel tag series vs. how ages are being determined for the RFID tags. With steel tags, fish age was determined at recovery, by reading daily otolith rings. The age at release was then calculated by subtracting the time at liberty from the age at recapture. Hjartåker (2017) points out that this exercise resulted to fish being assigned ages of 0 or negative at release, indicating a potential negative bias in the age readings. The potential for a negative bias in the age readings of larger individuals was also an issue that was highlighted in the last ICES ageing workshop report. With RFID tags, the age at release is estimated using an age length key built from the sampling of fish for lengths and ages at the time of tagging. Each fish is assigned a distribution of ages and, from that distribution, assigned an age a (how? Mean?). When the fish is recaptured, the age at recapture (a) is calculated as the age at release plus the time at liberty. The numbers at age a scanned at the factory is then determined by ageing a sample of fish (either by random sampling or length stratified sampling). The fact that bias appears to increase with increasing age is likely to cause a positive bias in the estimated numbers at age a scanned since older fish are likely to be misidentified as age a fish. Overestimating the number of fish scanned at age a would translate to a lower survival rate of tagged fish from that cohort, causing a pattern similar to what is being interpreted as tag loss. Simulation work could help determine if this is a potential issue. We also recommend that more validation work be done in older fish (8 plus) using chemical markings to get a better understanding of the aging bias pattern in older fish.

b) Sensitivity of the assessment to tagging data

The external reviewers believe the stock assessment group fully explored the sensitivities of tag data sources in the model.

c) The mathematical implementation

The external reviewers found no flaws (failed to reject the mathematical implementation). As more tagging is completed in the future, the reviewers recommend that the type of Poisson over-dispersion (i.e., type of NB) be investigated more thoroughly.

d) Aggregation level and inclusion of tagging data to be used in the assessment

Data were aggregated across plants which the review panel considered appropriate. The inclusion of the RFID and steel tag data in the stock assessment was deemed appropriate.

TOR 2: Review new information on other data sources relevant for the assessment

- a) Updated recruitment indices
- b) Update methods on dealing with invalid catch data in the assessment
- c) Bias in indices
- d) Estimated proportion of fisheries before surveys

These items were not presented at the second WK, so review by the panel was not possible.

TOR 3: Re-evaluate and update the assessment model configuration

The model configuration selected as the baseline is the best available and decisions made seem reasonable. There were many sensitivity analyses related to the treatment of tagging data, but less treatment of other uncertain data sources like natural mortality and more-recent potential unreported catches on young ages.

- a) Parameter bindings
- b) correlation structures
- c) Inclusion of data
- d) Sensitivity runs

TOR4: Re-examine and update MSY and PA reference points according to ICES guidelines

The procedures used to develop the reference points follow the ICES Advice Technical Guidelines 12.4.3.1 ICES fisheries management reference points for category 1 and 2 stocks.

A reduced recruitment series from 1998–2016 was used because prior values had more uncertainty about them. One reviewer suggests that it may be better to fit the stock-recruitment model within the SAM and avoid external fitting. The choice of age class to fit the SR model is questionable. The age 0 recruitment estimated from SAM seemed fairly unreliable (and depend on whether the model chose to fit the recruitment index or the age 0-1 catch well). Estimates of year class strength at age 3 could be substantially different than at age 1 because of the SAM process error in addition to some small catches. Perhaps a better approach is to derive ICES RPs based on a SR model using age 3 for recruitment. It also seems to be a problem that the SAM process error, which is ~20% CV, is not included in the EQSIM analyses.

Literature

- Hjartåker, I.D. 2017. Sources of bias in the RFID tag-recapture data used in the stock assessment of North East Atlantic Mackerel. *Msc thesis, University of Bergen*. 80pp.
- Hoenig, J. M., N. J. Barrowman, K. P. Pollock, E. N. Brooks, W. S. Hearn and T. Polacheck. 1998. Models for tagging data that allow for incomplete mixing in newly tagged animals. *Canadian Journal of Fisheries and Aquatic Sciences* 55: 1477–1483.
- Latour, R., J. M. Hoenig, J. E. Olney and K. H. Pollock. 2001. A Simple Test for Nonmixing in Multiyear Tagging Studies: Application to Striped Bass Tagged in the Rappahannock River, Virginia. *Transactions of the American Fisheries Society* 130:848–856.

Annex 1: List of participants

Name	Institute	Country (of institute)	Email
Anders Nielsen	DTU-Aqua	Denmark	an@aqua.dtu.dk
Andrew Campbell	Marine Institute	Ireland	andrew.campbell@marine.ie
Aril Slotte	Havforskningsinstituttet	Norway	aril.slotte@hi.no
Claus Reedtz Sparrevohn	Danmarks Pelagiske Producentorganisation	Denmark	crs@pelagisk.dk
David Miller	ICES Secretariat	Denmark	David.Miller@ices.dk
Gary Nelson (<i>reviewer</i>)	Commonwealth of Massachusetts, Division of Marine Fisheries	USA	gary.nelson@state.ma.us
Gjert Endre Dingsør	Fiskebat	Norway	gjert@fiskebat.no
Guðmundur J. Óskarsson	Marine and Freshwater Research Institute	Iceland	gjos@hafro.is
Höskuldur Björnsson	Hafrannsóknastofnun	Iceland	hoskuldur.bjornsson@hafogvatn.is
Ian Gatt	Scottish Pelagic Fishermen's Association	UK	ian.gatt@scottishpelagic.co.uk
Jan Arge Jacobsen	Havstovan, Faroe Marine Research Institute	Faroe Islands	Janarge@hav.fo
Jens Christian Holst	Ecosystembased	Norway	jens@ecosystembased.com
Jens Ulleweit	Theunen Institute	Germany	jens.ulleweit@thuenen.de
Leif Nøttestad	Institute of Marine Research (IMR)	Norway	leif.noettestad@hi.no
Lisa Ailloud (<i>reviewer</i>)	International Commission for the Conservation of Atlantic Tunas (ICCAT)	Spain	lisa.ailloud@iccat.int
Lisa Readdy	CEFAS	UK	lisa.readdy@cefes.co.uk
Magne Aldrin	Norsk Regnesentral	Norway	magne.aldrin@nr.no
Martin Pastoors	Pelagic Freezer-Trawler Association	The Netherlands	mpastoors@pelagicfish.eu
Neil Campbell	Marine Scotland	UK	Neil.Campbell@gov.scot
Niels Hintzen (<i>chair</i>)	Wageningen Marine Research	The Netherlands	Niels.hintzen@wur.nl
Noel Cadigan (<i>reviewer</i>)	Marine Institute, Centre for Fisheries Ecosystems Research	Canada	Noel.Cadigan@mi.mun.ca
Rose Mari Berge	Fishing company	Norway	rose-ma@online.no

Samuel Subbey	Institute of Marine Research (IMR)	Norway	samuel.subbey@hi.no
Sean O Donoghue	Killybegs Fishermen's Organisation Ltd.	Ireland	kfo@kfo.ie
Sindre Vatnehol	Havforskningsinstituttet	Norway	sindre.vatnehol@hi.no
Sonia Sanchez	AZTI	SpainP	ssanchez@azti.es
Stanislovas Jonusas (<i>observer</i>)	European Commission Directorate for Maritime Affairs and Fisheries DGMARE	Belgium	Stanislovas.jonusas@ec.europa.eu
Steve Mackinson	Scottish Pelagic Fishermen's Association	UK	steve.mackinson@scottishpelagic.co.uk
Teunis Jansen	DTU-Aqua & GINR	Denmark & Greenland	tej@aqua.dtu.dk
Thomas Brunel	Wageningen Marine Research	The Netherlands	Thomas.Brunel@wur.nl

Annex 2: Agenda IBP meeting 11–13th of December

10:00 Tuesday 11th – 17:00 Thursday 13th

ICES Headquarters, Copenhagen, Denmark

Tuesday

- Welcome & introduction by ICES (10:00 – 10:30)
- Conflict of Interest [notification, declaration, consequences] (10:30 – 11:00)
- Preliminary ToR [discuss, revise if necessary, accept] (11:00 – 13:00)
- Lunch (13:00 – 14:00)
- WDs by ToRs [present, discuss, way forward, impact] (14:00 – 17:30)

Wednesday

- WDs by ToRs continued [present, discuss, way forward, impact] (09:00 – 13:00)
- Lunch (13:00 – 14:00)
- General discussion on focus IBP (14:00 – 15:30)
- Conclusions and final ToRs (16:00 – 16:45)
- Discuss how to achieve progress [if necessary in subgroups] (16:45 – 17:30)

Thursday

- Work on ToRs (09:00 – 13:00)
- Lunch (13:00 – 14:00)
- Time-line agreements [incl additional Web-meetings] (14:00 – 14:30)
- Agreeing on process [include vs. exclude in IBP, role of participants, etc] (14:30 – 15:00)
- Documenting progress (15:00–17:00)

Annex 3: Agenda IBP meeting 4-7th of March

Agenda IBPNEAMackerel

4–7 March

IJmuiden, Haringkade 1, 1976CP, IJmuiden, The Netherlands

ICES professional secretary: David Miller (david.miller@ices.dk)

ICES Chair: Niels Hintzen (niels.hintzen@wur.nl)

External Chair: Gary Nelson (gary.nelson@state.ma.us)

Reviewers: Noel Cadigan (noel.cadigan@mi.mun.ca) and Lisa Ailloud (lisa.ailloud@ic-cat.int)

Meeting times: 09:00 – 10:30, 11:00 – 12:30, 13:15 – 15:00, 15:30 – 17:00 (2x coffee break of 30min and 45min lunch break)

General sessions, data sessions, modelling sessions, working sessions

Monday 4th of March:

09:00 – 10:00

Round of introduction
Facilities at Wageningen Marine Research
Meeting time-line
ICES Code of Conduct

10:00 – 10:30, 11:00 – 12:00

Presentation on catch-at-age data, uncertainties and sensitivity runs
Decision on catch-at-age data

12:00 – 12:30, 13:15 – 15:00, 15:30 – 17:00

Presentation on tagging data processing
Presentation on tagging data aggregation
Presentation on tagging data mathematical implementation in the model
Presentation on tagging sensitivity runs in the assessment

Tuesday 5th of March:

09:00 – 10:30, 11:00 – 12:30

Summary of presentations on tagging
Decision on mathematical implementation
Decision on types of data aggregation for sensitivity testing

13:15 – 15:00

Working session

15:30 – 17:00

Presentation on sensitivity runs

Wednesday 6th of March:

09:00 – 10:30

Presentation on sensitivity runs

Decision on preferred tagging data inclusion

11:00 – 12:30

Working session (parameter binding), leave-on-out, etc

13:15 – 15:00

Presentation on final model

15:30 – 17:00

Report writing

Working session on reference points

Thursday 7th of March:

09:00 – 10:30

Presentation on reference points

11:00 – 12:30

Report writing

13:15 – 15:00

Reviewers feedback

Summary of all decisions made

Recommendations

15:30 – 17:00

Closing remarks and agreeing on time-line and responsibilities to finish tasks

Report writing

Annex 4: Recommendations

The following recommendations were agreed by participants and reviewers of the interbenchmark.

Recommendations can be addressed to one or more of the following: other Expert Groups, ICES Secretariat, ICES Data Centre, ACOM, or SCICOM.

Recommendation	Addressed to
RFID tagging data	
Support the existing proposal by WGWIDE to hold a Tagging Data workshop prior to WGWIDE. In particular, the workshop should (i) investigate tagging data to understand better the differences in the distribution of ages covered by the tagging programme vs the ages caught by the fishery and what this implies regarding how tagged fish mix in the stock and its implications for assessment, including the choice on using of ages 5+ age and 2013 as start year for release). (ii) investigate alternative ways to use the tagging data to estimate total mortality (Ref WD 27 -, e.g. in terms of inferred total mortality (ref. WD 27y – Deriving Auxilliary stock information)	ICES Secretariat, WGWIDE
Recommend that tag scanners have regular efficiency tests and routine servicing so that any corrections for changes in scanning efficiency (affecting numbers scanned) can be made.	WGWIDE
Biological parameters	
Questions remain over the consistency and reliability of ageing among mackerel age readers. This issue was raised in relation to apparent younger fish for a given size being recorded by Scotland in comparison with Norwegian and Irish sampling from fisheries taking place at the same time and place. In an effort to improve the data for the mackerel assessment, the benchmark supports the existing recommendations of WKRMAC2 (2018)	WGBIOP
Advise on latest information on estimates for natural mortality at age in mackerel and density dependent effects such as cannibalism	WGSAM
Evaluate whether time-varying maturity ogives are relevant for assessment and advisory purposes	WGWIDE
Indices of stock abundance	
Reflect on Working Documents regarding sources of bias in the IESSNS swept area survey (WD6- Holst Mackerel bias in the trawl method) - and mackerel egg survey (WD5- Holst Mackerel bias in the egg method)	WGMEGGS, WGIPS
WGIPS is requested to report to WGWIDE on any issues relating to stock containment by the IESSNS survey because this affects the interpretation of the catchability ($q=1$) assigned to the survey in the stock assessment model.	WGIPS
Request WGWIDE and ICES Data Centre to reflect upon the discussion of quality assurance procedures documented in this report in section 5.2.	WGWIDE, ICES Data Centre
Stock assessment modelling	
Consider use of alternative indicators of stock biomass in the stock assessment model (e.g. exploitable biomass) that may relate to evidence from the fishing on trends in biomass.	WGWIDE
Include an irregular grid AutoRegression as an option for the F model in SAM	WGMG

Investigate and simulation test how datasets are internally weighted in SAM	WGMG, WGWIDE
Access to Data	
Include requirement for data call to get Scottish biological data for use with tag programme	ICES Data Centre
Provide online access to i) DATRAS data products "CPUE by age and haul", "CPUE by length and haul" and "ALK" for all demersal surveys in Q4 and Q1, and ii) quality assurance reports.	ICES Data Centre

Annex 5: Future research requirements

Research Questions	Justification
RFID tagging data	
What is the short-medium term mortality of mackerel tagged with RFID tags?	Questions remain regarding apparent differences in mortality of mackerel tagged with steel and RFID tags and how this influences the stock assessment.
To what extent does tag loss or tag malfunction occur and how does this change over time, with the procedure for tagging fish (e.g. in cavity/ muscle) and in male and female fish? Can double tagging experiments be performed to investigate this?	Evidence of reducing tag recapture rates with age point to possible shedding of tags and or malfunction of tags over time. The consequence for assessment is possible bias in estimates in abundance, which is one reason why the benchmark chose to limit recapture data to include only tags with up to 2 years out at sea.
Does the length/age distribution of tagged fish match the age profile of fish in the vicinity at that time.	There is a need to understand if the selection pattern of the RFID tagging experiments reflects the population available to be tagged at that time and place.
Does the age distribution of scanned tagged fish that is derived from age-length keys accurately reflect the ages of the tagged fish?	There is a need to validate the assumed ages of recaptured fish by trying to recover samples of tagged fish from the factories or from biological sampling.
Biological parameters	
What are the consequences for representation of mackerel stock structure of using age-length keys to determine the age distributions of released and recaptured mackerel?	Given difficulties with ageing issues, particularly for older age mackerel, the use of age-length keys to assign ages to recaptured fish may cause a bias in numbers at age estimates.
How do changes in mackerel growth affect estimates of weight-at-age and what is the sensitivity of stock assessment to such changes?	Changes in growth rate necessitate that length and weight data should routinely be collected during biological sampling to see how these change with age as the stock evolves. At present there are cases where weights at age is derived from historical length-weight relationships that do not reflect recent conditions. Self-sampling from industry vessels could provide annual weight-length keys by month.
How do changes in mackerel distribution and growth affect estimates of natural mortality of mackerel.	Estimating natural mortality is difficult and hard to verify. The recent expansion in distribution area and change in weight-at-age could be associated with changes in natural mortality. Investigating the role of mackerel as prey and predator in multi-species models could provide further insights.
Indices of stock abundance	
Is the mackerel stock distribution adequately covered by surveys?	There are concerns that when scientific surveys do not fully contain the stock distribution, important information can be missing that may bias stock assessments.
What information can industry contribute to evidence changes in the distribution and abundance of mackerel? Examples discussed during the benchmark included information on in length and weight at age, acoustic estimation,	Fishermen have an acute awareness of changes in mackerel abundance and age structure of the stock, which would benefit stock assessment by being made available in a quantifiable ways.

<hr/> measurements of the spatial extent of fish distribution (e.g. school marks, distance and clustering), spatial extent of fishing. <hr/>	
Stock assessment modelling <hr/>	
How can stock assessment models be evolved so that they are better equipped to take account of spatial dynamics of stock, survey data, and fishery and improvements in the resolution of available biological and fishery data?	Numerous issues discussed during the benchmark were related to understanding of the spatial structure of the stock and fishing. Spatial assessment models may improve the representation of the mackerel stock.
<hr/>	

Annex 6: Minority Statement and Response to the Minority Statement by the IBP

This statement does not reflect the opinion of IBPNEAMac but of one member only. A response from IBPNEAMac to the issue raised is given below.

Minority statement by Jens Christian Holst:

Having followed the very good and constructive process of ICES IBPNEAMac 2019 throughout, I found at the end that I cannot agree in the model agreed by my colleagues in the benchmark group. On this basis I have been allowed by ICES ACOM to write a minority statement into the report. I appreciate this opportunity as I find it important for an open scientific body like ICES, in particular in a benchmark process.

There are several signs of a strong growth of the NE Atlantic mackerel stock starting around 2006-7. Amongst other, larger and larger amounts of mackerel were observed in the international survey in the Nordic Seas, IESNS, in May from 2008. The distribution of mackerel increased out over its prior known area of distribution with mackerel found in summer from the southern point of Greenland, pass Jan Mayen Island to The Longyear Town in Spitzbergen and into the southern Barents Sea. Significant, earlier unseen, mackerel fisheries started around Iceland and in the Greenland zone in summer. The international summer survey in the Nordic Seas, IESSNS, indicate a tripled distribution area and a doubled density from 2007 to 2016 according to Nøttestad and Utne (2016) i.e. a six doubling of the stock.

Given equal tagging mortalities between the two tagging methods used in the metal and RFID tagging experiments carried out in the mackerel stock also indicate a strong growth in the stock size. The assumption of equal tagging mortality between the two tagging methods is strong and valid in my view, given the results of the tagging experiment carried out by IMR, Norway, in 2018, showing no difference in tagging mortality between the two methods.

The scientific quantitative observations on a strong growth in the mackerel stock referred above are in close correspondence with the qualitative observations done by the professional mackerel fishermen and also by the public's observations along the shores of the NE Atlantic during the period described above.

The model agreed by the group does in my view not produce an assessment that corresponds with the growth and present size of the mackerel stock as indicated above. In my opinion there is a level error, maybe in the order of more than one, produced by the model agreed by the group and the size of the actual mackerel stock in the ocean.

Some of my main concerns about the agreed model is that ignores the results of the tagging mortality experiment and estimates two separate tagging mortalities for the two tagging methods. The estimated values of the two tagging mortalities are around 0.6 for the old tagging method and 0.9 for the new. A SPALY (Same Procedure As Last Year) run estimating one common tagging mortality results in a spawning stock about 2.5 times higher than that of the agreed model in the final year, giving in my view a valid and strong indication that the actual size of the mackerel stock is severely underestimated by the agreed model.

Furthermore, three independent tagging mortality experiments have been carried out on mackerel (by Hamre, Lockwood and Slotte), indicating tagging mortalities between 0% and 20%. Alt-

though tagging mortality experiments cannot simulate the real tagging mortality of 'in situ' releases, the difference between the results of these three experiments and the tagging mortality estimated by the agreed model and adopted by the benchmark group, are simply too large to be acceptable in an assessment in my view.

The agreed model uses a constant natural mortality M at 0.15 over all ages and years. The strong growth in distribution area combined with the parallel strong decrease in growth parameters like weight at age and widespread observations of cannibalism and hunger, indicate that an increase in M in later years may very well be part of explaining the inconsistencies between the data and the agreed model and the need to exclude valid experimental results and data.

The benchmark group, in my view, put a too large emphasis on the search for the best possible model fit to the data, i.e. finding a lowest possible AIC, while disregarding the highly valid result of a common tagging mortality coming from the 2018 tagging mortality experiment and by leaving out for instance a large portion of the RFID tag data.

Rather than disregarding valid experiment results and cutting data, the benchmark group should in my view have looked closer for possible errors in the data and different model formulations that would have resulted in an assessment better in correspondence with the larger stock as indicated above.

On this background I recommend ICES to initiate a focused study into various factors which may help entangling some of the inconsistencies between the data and the agreed model. Such a work should amongst other contain a modelling exercise where different scenarios of increased M in parallel with the last decade decrease in growth is tested along with estimating one common tagging mortality and including the entire RFID dataset. A qualitative focused study into describing the possible increase in M should also be initiated.

Response to minority statement by the IBP:

The minority statement of Holst focusses on a number of topics, summarized as:

- a. Estimated tagging mortality between steel tags and RFID tags: Holst argues that several experimental trials have demonstrated that release survival rates should be no different between the steel and the RFID tags.
- b. Rate of increase and absolute level of stock size: Holst argues that the increase in the Mackerel stock is not well reflected in the assessment and underestimates the absolute stock biomass. He suggests that the IESSNS survey shows a 600% increase from 2007 – 2016 and that when assuming similar tagging survival rates between steel and RFID tags one would estimate a 2.5fold higher biomass.
- c. Natural mortality assumption: Holst argues that natural mortality in the recent period can likely be higher than currently assumed in the assessment.
- d. Data inclusion in the assessment and model parameterisation: Holst argues that a large portion of RFID data is dropped and that model selection was based on statistical features rather than reflecting on the results from the tagging experiments.

A response to each of these points is given below.

- Ra. The IBP does not dispute that release survival rates between steel and RFID tags can well be similar. The parameter named 'survival rate' in the assessment however represents a

number of processes (in addition to actual post-release survival) which impairs a one-to-one comparison with the experimental trials.

The reason for these two very different “survival rates” parameter estimates for the steel tags and the RFID tags (around 40% and 10% respectively) was discussed at length during the IBPNEAMac 2019. Experiments conducted recently have shown that the old way with manual jigging of mackerel and release at starboard side, gives the same recapture rates as the new way with automated jigging machines and fish releases through pipes at the port side. This suggest that there should not be any substantial deterioration of survival for the new tagging program. Using the term “post release survival” for this scaling parameter likely introduces confusion. One difference between the steel tag tagging and the RFID tagging is related to the position of the inserted tag. During steel tagging, the tag was inserted into the muscle of the mackerel if gonads were large and ripe or running, whereas in less ripe or spent fish tags were injected into the abdomen. During RFID tagging all tags are injected into the abdomen regardless of maturation status, due to food safety issues. Experiments suggest that fish tagged in abdomen have higher tag loss than fish tagged in the muscles, and the tag loss typically happens at high rates during the first spawning event. RFID tag loss have been reported to be very high up to 40% for females in trout. Since RFID tags are more likely to be expelled at the first spawning after tagging, one should automatically expect a lower “survival”.

Besides that, a systematic bias in any of the variables in the tagging data set would be confounded with the “survival” parameter. For example, when running the model with 4 different survival rates for the steel tag period (dividing the steel tag data up in blocks of years) and a 5th rate for the RFID, high survival is estimated for the period prior to 2000 (0.438 and 0.477), declining afterwards to similar values in the latest steel tags (0.081) and in the RFID tags (0.10)(WD Brunel3). One of the explanation proposed was that there was a consistent underestimation of the volume of the catches scanned in the early years (before stricter control was implemented). This negative bias in the number scanned was compensated by an over estimation of the survival rate. Likewise, any bias on the numbers recaptured (e.g. underestimation due to poor scanning efficiency in the factories), will be confounded with the estimated value of survival rate parameter (in this case lead to an underestimate). Furthermore, in the time-period of the steel tags, no contrasting data was available to reliably estimate tagging survival rate while this is the case in more recent years. Finally, a test has been conducted that showed that using two distinct survival rate parameters was appropriate.

These points provide clear arguments on why the estimated survival rate, potentially confounding several different processes, can be different between the steel and RFID tags.

- Rb. The final accepted assessment by the IBP shows a rapid increase in population biomass since the mid-2000s where the mature part of the population more than doubled. The assessment estimates this based on several different data sources, such as catches and the cohort signal therein, the IESSNS survey, the tagging data and the egg survey. There are contrasting views among these datasets where some suggest further increases in stock size and others suggest a declining trend. The six fold increase in Mackerel according to the IESSNS survey since 2007 depends largely on the relatively low 2007 estimate. The spatial coverage of this survey in its initial year was very limited compared to the following years and it is therefore that the benchmark made the decision in 2017 to exclude the year 2007. From 2010 to 2017 the biomass estimate from this survey indicates

- a doubling of the stock while the assessment indicates a downward trend from 2014 onwards. This illustrates the differences in perception by each of the data sources. Arguments why tagging survival rate should be kept separate are provided under Ra. In addition, a model run in which survival rates for the steel and RFID series was assumed similar resulted in a statistically unacceptable fit to the data. Strong patterns in residuals in the tagging data fit, residuals in the catch data fit and strong bias in retrospective perception are apparent in that run. This run was therefore not considered to be a reliable and appropriate estimate of the Mackerel stock.
- Rc. Alternative assumptions on natural mortality were not considered by the IBP. These were not part of the Term of References nor were analyses brought forward during the IBP to discuss the topic. The group considers that natural mortality is generally very difficult to estimate and verify but is open for science founded alternatives to be sensitivity tested in the assessment in the future.
- Rd. The group carefully scrutinized the different data sources that go into the assessment and have made substantial improvements on data quality checking and validation of the results. The group has gone to great lengths to identify any potential bias in the RFID tagging data and has decided to drop those data-points that could not be validated or where concerns raised by the group could not be addressed appropriately. This has led to maintaining around 85% of the recaptures which is represented by 25% of the data points that go into the assessment (hence dropping 75% of the RFID data points). Arguments for dropping these data are provided under Ra and Rb as well as the summary section of the IBP report and include issues related with spatial coverage of tag recapture registration, tag-loss bias and representative mixing of young tagged fish. The group followed a strict protocol to arrive at its final model configuration. All decisions related to selection of data were made based on *a priori* knowledge on the methods by which the data were collected and the potential bias in this process that could distort the quality of the data. Assessment model evaluations were used to evaluate if there would be statistical support for these *a-priori* findings. Furthermore, stock assessments are based on advanced statistical models. In statistics it is common to reflect and select on model fit vs observations as expressed in residuals. The general fit of the model to the data (as expressed by the patterns in residuals) was considered the second stage of model and data selection. A widely used selection criteria in statistical model selection is the AIC criteria which can be used to illustrate improvements in model fit to the data. Lower AIC values usually indicate in smaller and more balanced residuals, two desirable features of a model fit to data. In the final stage of model configuration, in which only assessment model parameter configurations played a role, the AIC criteria was used for final model tuning. The group argues that data and model selection were carried out according to best practice and highest scientific integrity.

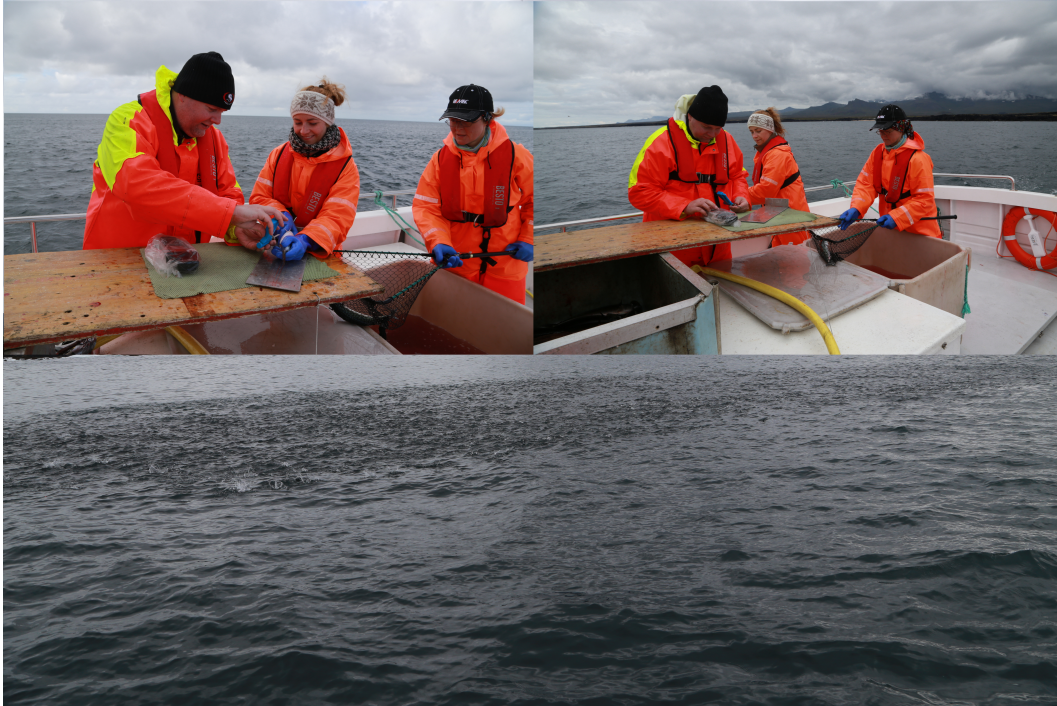
Annex 7: References

- ICES, 2015. Report of the joint ICES-MYFISH Workshop to consider the basis for FMSY ranges for all stocks (WKMSYREF3), 17th – 21st November 2014, Charlottenlund, Denmark. ICES CM 2014/ACOM:64. 156 pp.
- ICES, 2017a. Report of the Benchmark Workshop on Widely Distributed Stocks (WKWIDE), 30th January – 3rd February 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:36. 196 pp.
- ICES, 2017b. Report of the Workshop on Management Strategy Evaluation for Mackerel in Subareas 1-7 and 14, and in Divisions 8.a-e and 9.a (Northeast Atlantic) (WKMACMSE), 28-29 August 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:48. 2016 pp.
- ICES, 2017c. ICES fisheries management reference points for category 1 and 2 stocks. ICES Advice 2017, Book 12. http://ices.dk/sites/pub/Publication%20Reports/Advice/2017/2017/12.04.03.01_Reference_points_for_category_1_and_2.pdf
- Sánchez, S.; Brunel, T. and Nielsen, A. 2017. North East Atlantic Mackerel assessment – sensitivity analysis. Working Document for the Benchmark Workshop on Widely Distributed Stocks (WKWIDE-2017). 30th January-3rd February. Copenhagen, Denmark.

Annex 8: Working Documents

NEA Mackerel

Analysis of tagging data



Höskuldur Björnsson MRI Iceland
December 10th 2018

Short definition to avoid confusion. In this text tagging mortality refers to the mortality introduced by the tagging operation, bleeding, predations by birds etc. Tagloss is then the gradual loss of tags with time behaving in the same way as additional natural or hidden mortality.

RFDT tagging data are an important data set in current Mackerel assessment. The data are age-disaggregated, both at tagging and recapture with the recaptured fish not taken from the pile and age disaggregated. In the following report the author is just learning to understand the data base and the results are mostly a summary of plots done during that procedure.

One nice thing about the mackerel tagging program is that the recapture effort is known. A number of mackerel factories have scanners for the catch and the weight and number of mackerel landed in those factories is known. In addition length and age of the landed fish is sampled regularly allowing landings by age to be compiled for the same factories.

Proportion of the catch that is scanned has been quite high since 2012, increasing from 10% 2012 to 20% in 2017 (figure 1 and table 1). If 20% of the stock are caught each year the recapture rate in the year after tagging should be approximately $0.2 \times 0.2 \times e^{-TM} \times e^{-TL} \times e^{-M}$, where TM is tagging mortality, TL tag loss and M natural mortality in one year. 0.2 is approximately the proportion of the catches that is scanned each year. Looking at tagging from 2013-2016 only 0.4 – 0.7% of the tags have been recaptured in the year after tagging (table 2, the 3 factors TL , M and TM account for 80-90% reduction in the recaptures. The assessment estimates $\approx 90\%$ tagging mortality that is apparently right order of magnitude. When looking at the results proportion surviving is often a better measure to look at, the difference between the numbers 0.9 and 0.95 is not large but the difference between 0.05 and 0.1 seems larger.

Looking at recaptures from the tagging experiments from 2014-2016 when scanning rate had increased 1.6% of the tagged fish has been recaptured but it must be remembered that relatively few of the 2018 recaptures have entered the database. Comparing Icelandic and other tagging experiments in those year the proportion recaptured is 2.5% in the Icelandic experiments but 1.5% in the other ones. This comparison is interesting as the Icelandic experiments are conducted by smaller boats and much closer to the coast. The difference could though be caused by size of fish as considerably larger fish is tagged in Icelandic waters (average length 37.2 vs 35.8 cm). Later analysis that indicate higher recapture proportions of larger fish can though not be explained by geographical distribution as the Icelandic tags are only 4% of the total.

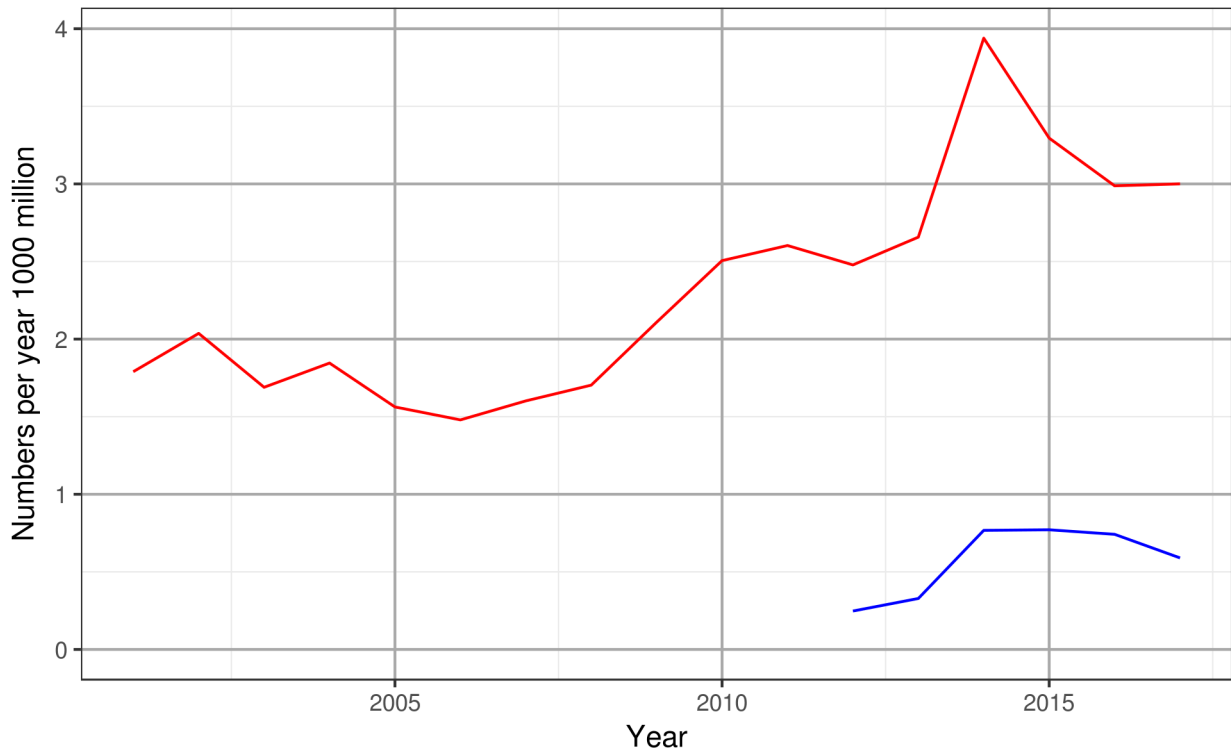


Figure 1: Number caught (red line) and number scanned (blue line)

Table 2 shows the number tagged each year, number recaptured 2012-2018, percent recaptured 2012-2018, percent recaptured the year after tagging and the recaptures each. What can be noticed is rapid increase in the proportion caught in the first year from 2012 - 2015 (2011 tagging experiment is not comparable as the fish is much smaller (figure fig:ledist). The number and proportion of the catch that has been scanned increased from 2012-2014 but has been relatively stable since then (figure 1).

	ryear	cno	nscan	p
1	2012	2.48	0.25	0.10
2	2013	2.66	0.33	0.12
3	2014	3.94	0.77	0.19
4	2015	3.30	0.77	0.23
5	2016	2.99	0.74	0.25
6	2017	3.00	0.59	0.20

Table 1: Number caught number scanned and proportion of catch that has been scanned each year

	Ntags	N2012-18	%2012-18	%R_Ty+1	2012	2013	2014	2015	2016	2017	2018
2011	49898	297	0.60	0.07	36	55	55	58	43	21	29
2012	32137	283	0.88	0.18	31	58	60	67	34	21	12
2013	22791	337	1.48	0.39	0	26	89	109	61	31	21
2014	55184	940	1.70	0.58	0	0	112	321	277	139	91
2015	44711	617	1.38	0.49	0	0	0	123	221	180	93
2016	48840	770	1.58	0.77	0	0	0	0	183	374	213
2017	59963	536	0.89	0.62	0	0	0	0	0	165	371
2018	40009	209	0.52	0.00	0	0	0	0	0	0	209

Table 2: Number tagged 2011-2018 and number recaptured for 2012-2018, Percent recaptured 2012-2018, percent recaptured year after the tagging year and number recaptured for each year 2012-2018

An indicator of scaled recaptures from a tagging experiment would be $\frac{Nr_{ty,ry}}{Nt_{ty} \times Nscanry}$. What does this index tell us? The number scanned are from the fishable stock but as time passes new recruitment is introduced into the fishable stock and lower proportion of the scanned fish is from the tagged agegroups.

If recruitment to the fishable stock was reasonably constant and the tagged fish close to fully selected to the fishable stock, the slope of those the catch curves obtained this way would be an indicator of total mortality (Z) plus annual tagloss. Age disaggregating the scanned will help in correcting for variable recruitment. Use of age-length keys, both at tagging and recaptures will introduce additional noise.

The indicator of scaled recaptures is shown in figures 2 and 3. The trends in scaled recaptures has been towards higher recaptures in the first years but more rapid drop. One explanation could be good recruitment to the fishable stock 2012-2014 compared to the following years . The possibility of better survival from the most recent tagging experiments can not be excluded.

Looking at the tagging experiments 2011-2014 and the recaptures in 2015 and 2017 (scaled by number scanned) the numbers reduce by factor of 3 indicating Z of 0.55 per year that might be an indication of F in the range 0.25-0.3 for the years 2015-2016. (using tagloss per year of 0.1 – 0.15). This number is though dependent on the recruitment to the fishable stock not being too variable.

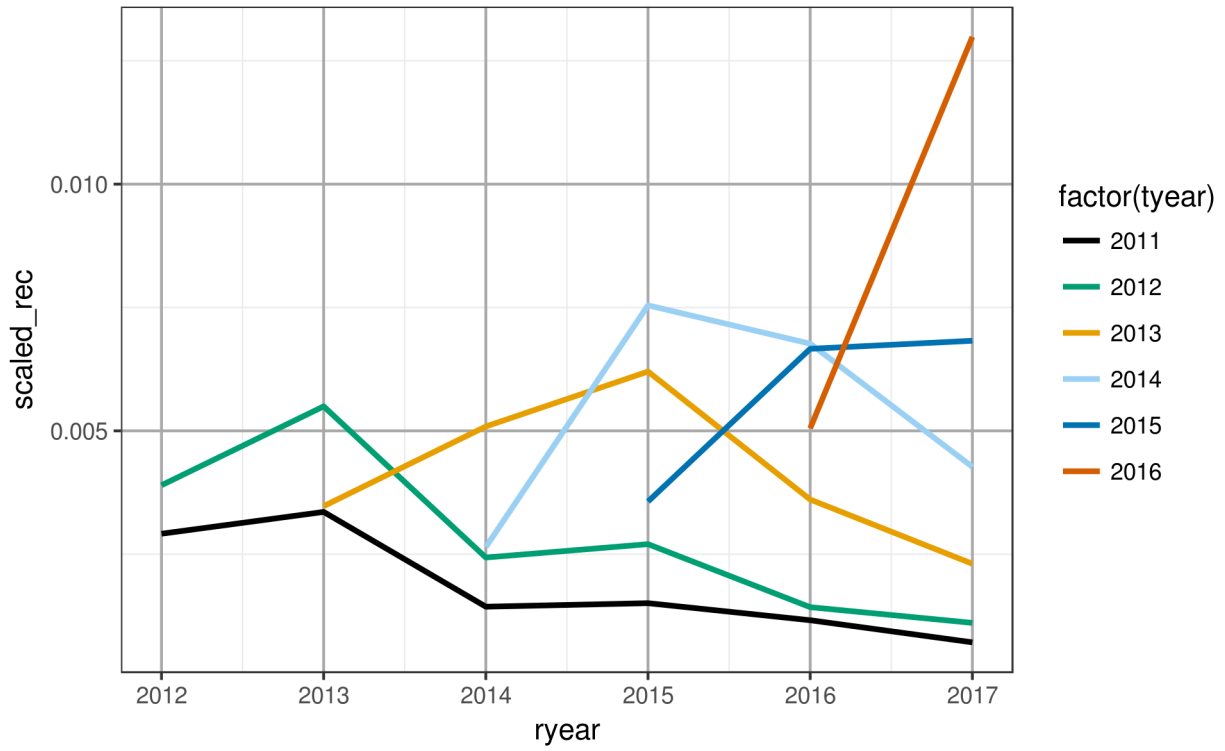


Figure 2: Scaled recaptures for each tagging year and recapture year.

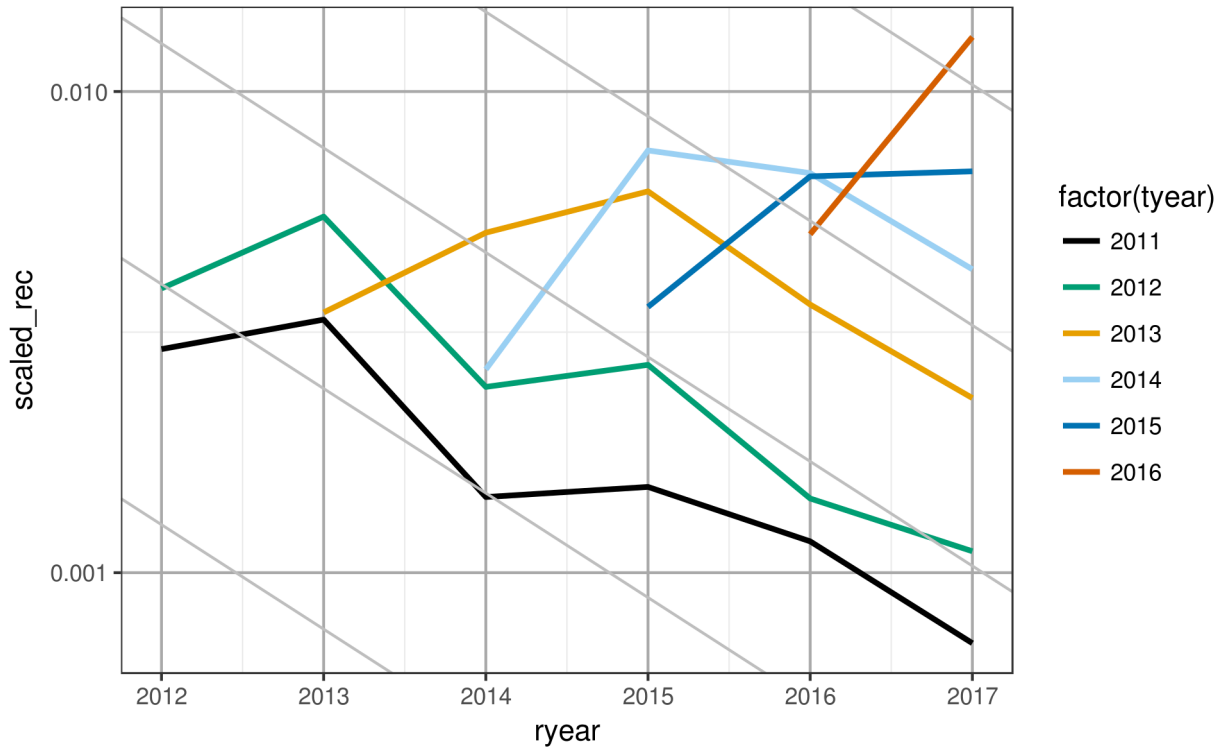


Figure 3: Scaled recaptures for each tagging year and recapture year plotted on log scale. The grey lines correspond to a slope of 0.5

One disadvantage of catchcurves from tagging experiments is that additional unknown variable (annual tagloss per year) is introduced. The question is what that parameter does with the estimated tagging mortality. Introducing this parameter in the SAM model (simple change) leads to tagging survival (in RFTD tags) changing from 0.12 to 0.16 but the estimated tag loss mortality per year is 0.14 (table ??). The objective function changes from 2786 to 2769 and ssb in 2018 changes from 2.40 to 2.84 million tonnes.

	run	parameter	value	stderr
1	No tagloss	logitReleaseSurvival	-0.47	0.10
2	No tagloss	logitReleaseSurvival	-2.20	0.08
3	No tagloss	logitRecapturePhi	1.57	0.27
4	No tagloss	logitRecapturePhi	1.85	0.63
5	Tagloss	logitReleaseSurvival	-0.38	0.12
6	Tagloss	logitReleaseSurvival	-1.82	0.12
7	Tagloss	logitRecapturePhi	1.65	0.28
8	Tagloss	logitRecapturePhi	2.59	1.02
9	Tagloss	logtagLoss	-3.67	0.56
10	Tagloss	logtagLoss	-1.99	0.18

Table 3: Parameter estimates from a SAM model run without and with annual tagloss estimated

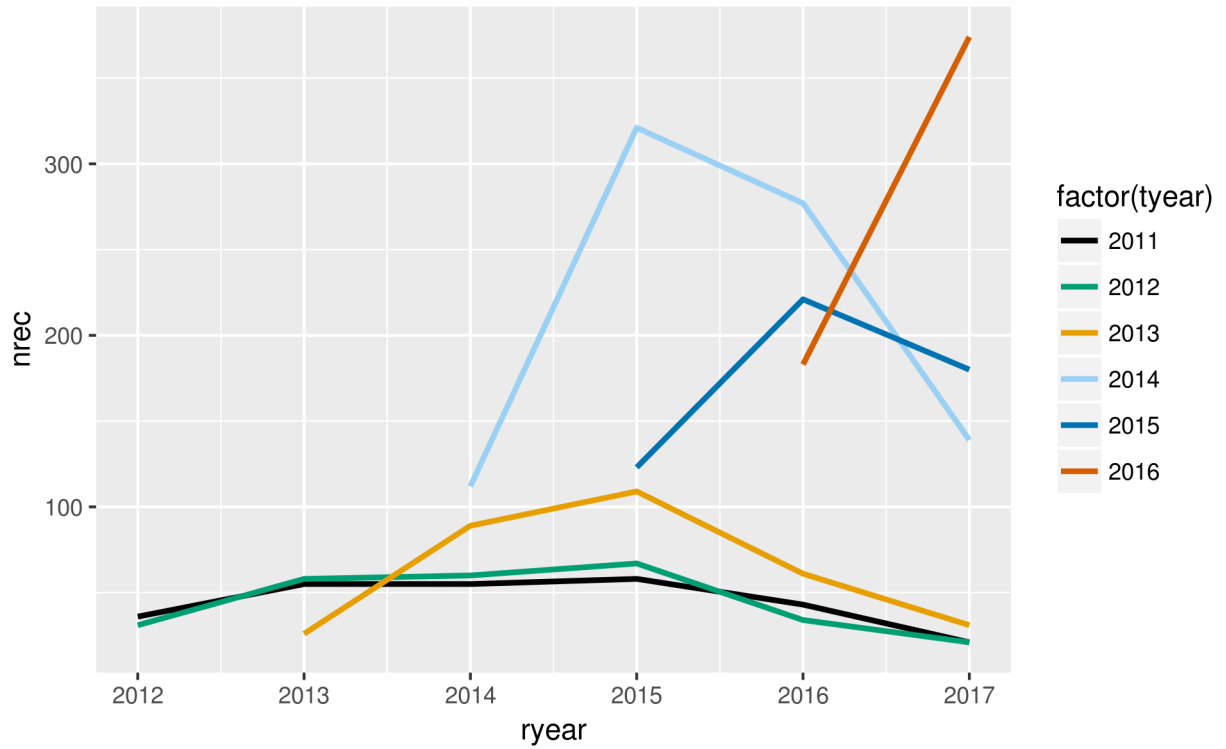


Figure 4: Number recaptured for each tagging year and recapture year.

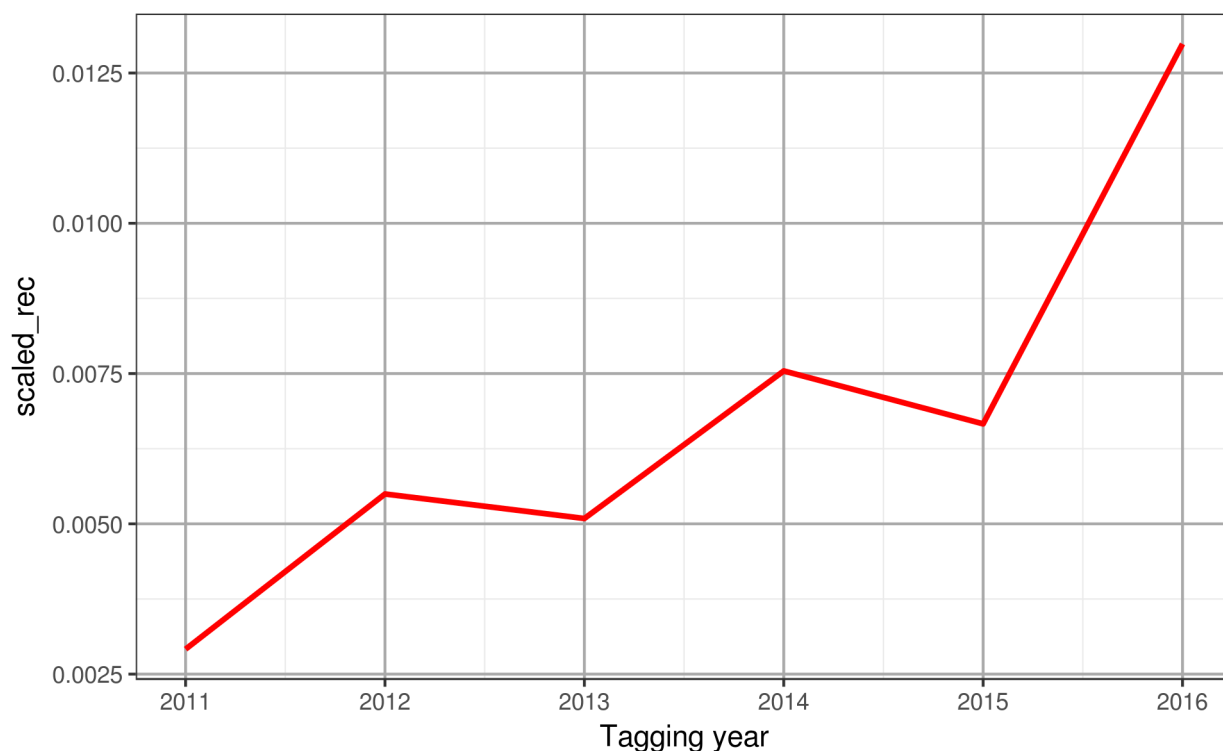


Figure 5: Scaled recaptures in the year following the assessment year

Looking at scaled recaptures in the year following the tagging year for the tagging years 2011-2016 continuous increase can be seen. An increase like shown here would in an assessment model be interpreted as reduced stock size/increased F or as described earlier poor recruitment to the stock. This might be correct but if what is happening is improved handling of the fish when tagged or improved detection the assessment would be wrong.

One thing that affects the proportion detected is the age/size composition of the tagged fish, larger fish tagged would normally lead to more recaptures (this is taken care of by age disaggregating the tagged fish). Quick look indicates that length distributions other years than 2011 are reasonably similar (figure 6), though not identical and the tendency is towards larger fish. Age distribution is probably not variable as mackerel grows slowly after the first 2-3 years.

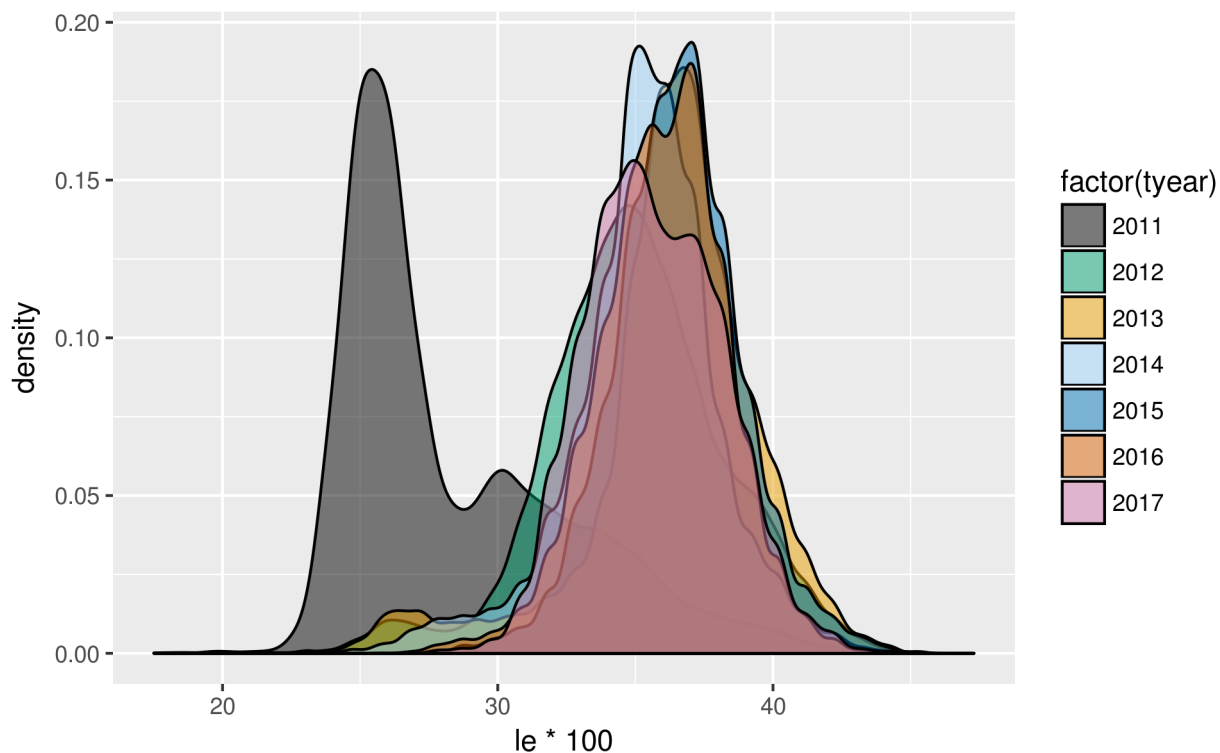


Figure 6: Length distribution of tagged fish 2011-2017

Scaled recaptures of fish from different tagging years in the year following the assessment year is shown for different size classes in figure 7. Apparently the recapture rate of large fish is highest, though not the largest fish where it is again lower (38.5cm and larger). This picture does of course have to be corrected for the size of the scanned fish. It does though lead to suspicion that tagging survival is size based. Time trend in the recapture rate for all but the smallest fish is also interesting. The picture for recaptures 2 years after the tagging year is less clear (figure 7)

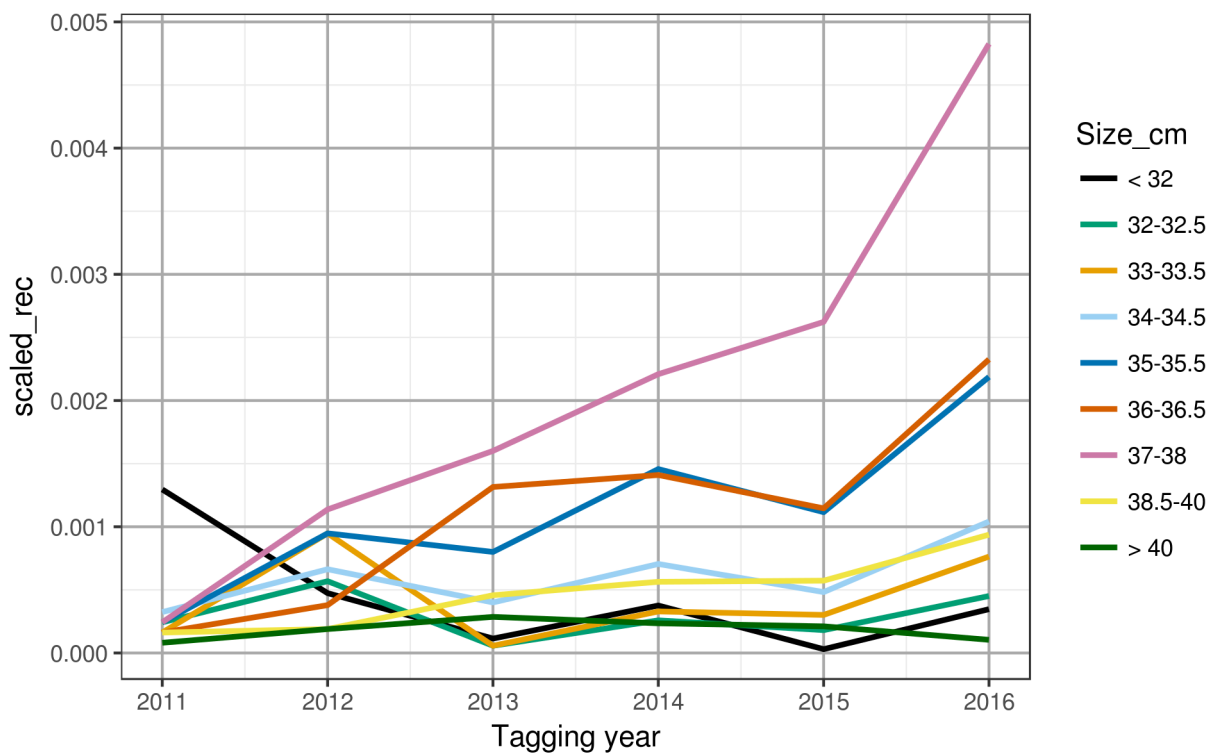


Figure 7: Scaled recapture 1 year after tagging for different size classes

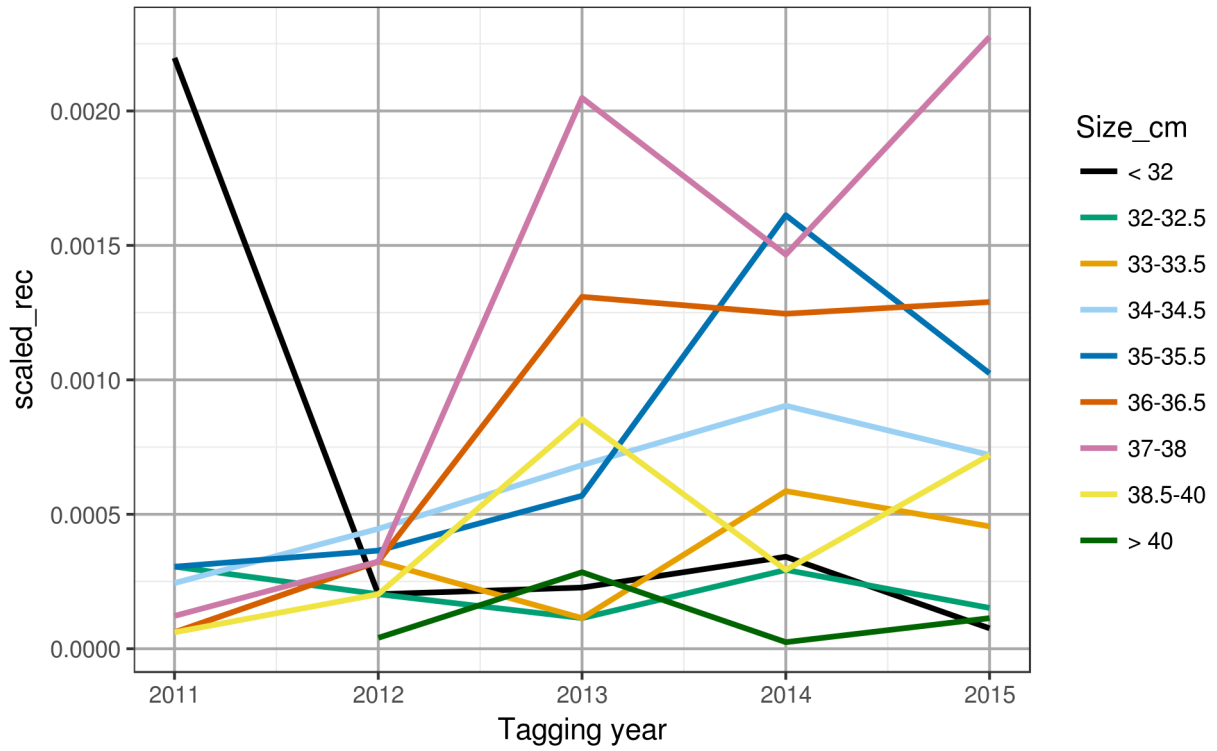


Figure 8: Saled recapture 2 year after tagging for different size classes

1 Length and age distribution of recaptured fish

Length distribution of scanned fish have been quite variable (figure 10). The most apparent problem is in 2016 where fish is in some factories measured to the nearest 1/2cm. Those fishes will here be randomly allocated to the nearest cm.

A problem with the data base is that in the table BioSamplesCatches some AssayID occur 2-4 times meaning that the same fish is used in more than one cell. Probably OK if it is in the same year (needs to be checked).

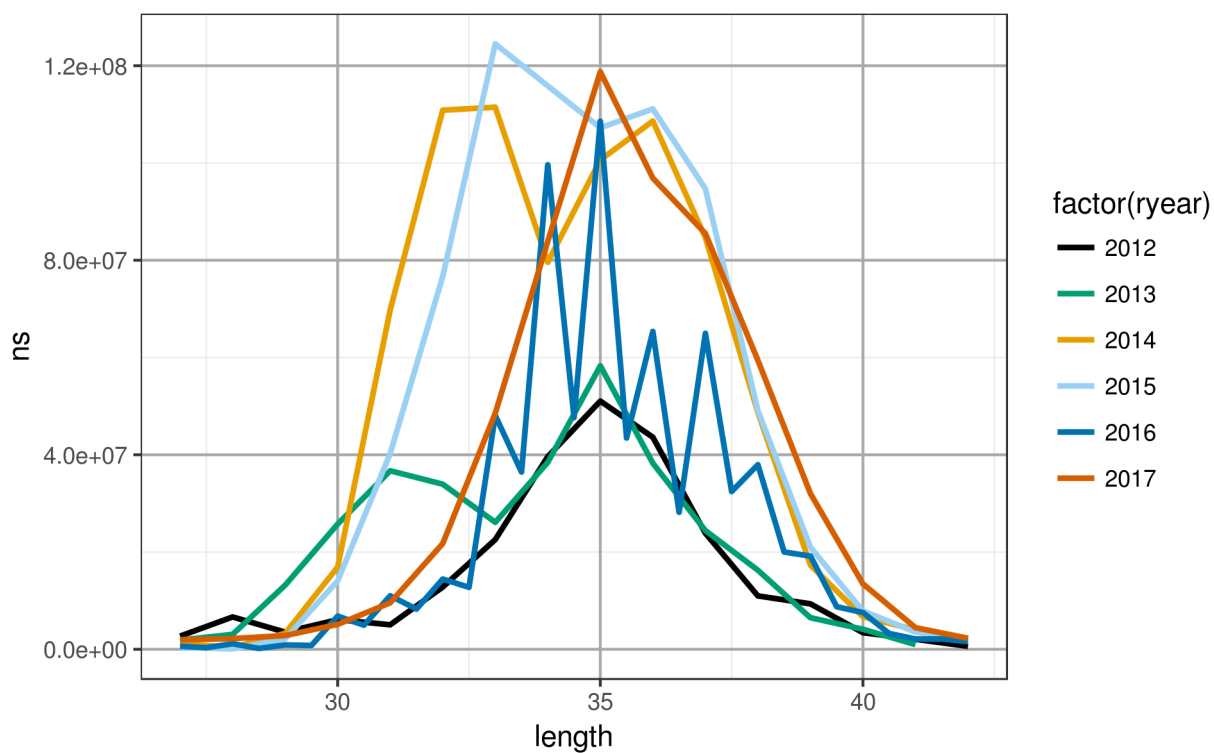


Figure 9: Length distributions of scanned fish scaled up with number scanned

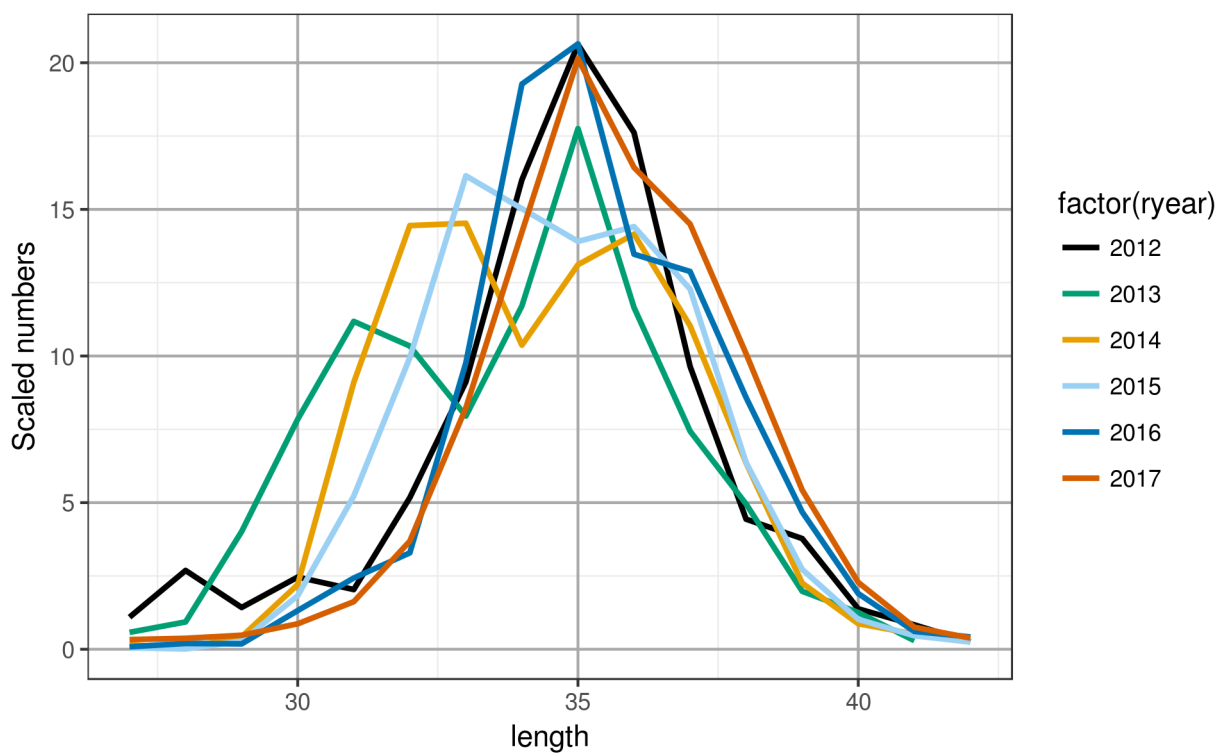


Figure 10: Length distributions of scanned fish scaled up with number scanned, summed up to 100 every year

Scaled length distributions from screened tags 2012-2017 are shown in a different way in figure 11 where they are compared with the average. The most notable deviation from the average are indications of recruitment in 2013-2015. These length distributions should be compared with the length distributions used to compile catch in numbers.

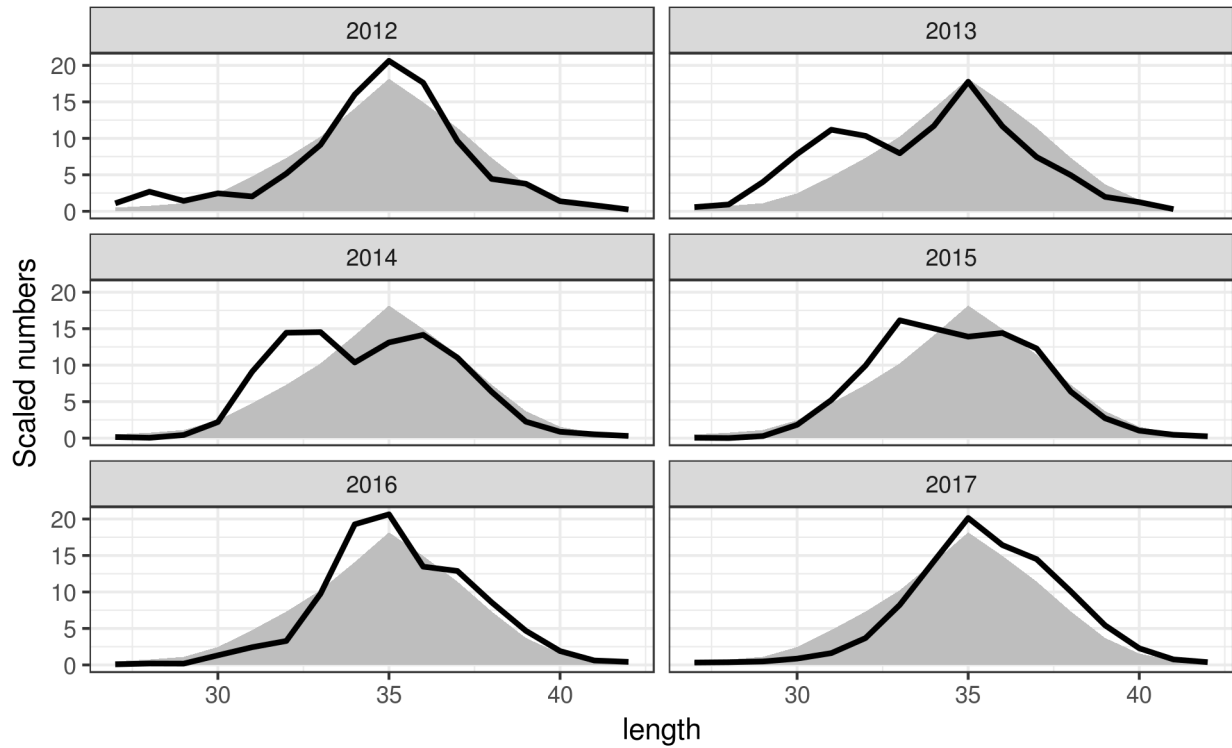


Figure 11: Length distributions of scanned fish scaled up with number scanned, summed up to 100 every year. The shaded grey values show the average over the years

As modelling of the mackerel is done, age distribution of the scanning and how it compares with catch in numbers is a larger issue. Is it good enough to assume that the selection of the scanning is the same as of the fisheries?

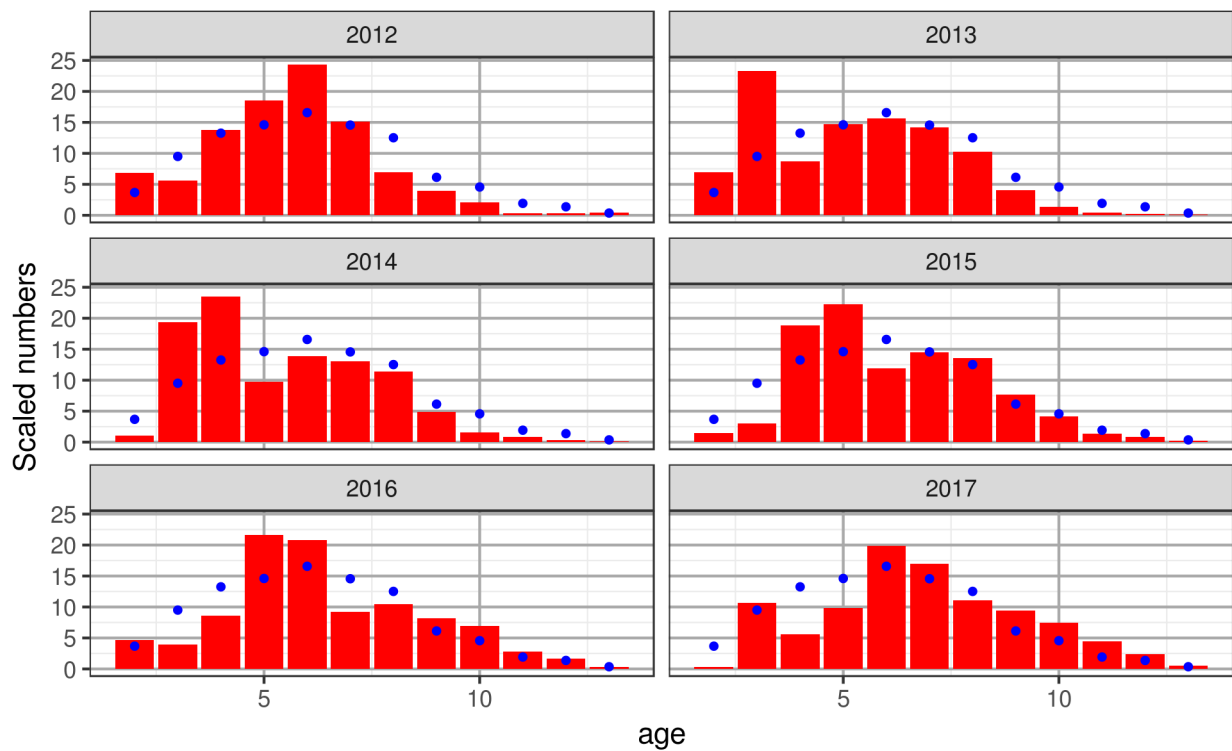


Figure 12: Age distributions of scanned fish scaled up with number scanned, summed up to 100 every year. The points show the average over the years

The age distributions shown in figure 12 does not match the age distribuitin of catches in the annual

assessment very well. This needs to be investigated further but the proportion of fish scanned has been over 20% in recent years. The number of samples is reasonably high or 25000 in 7 years.

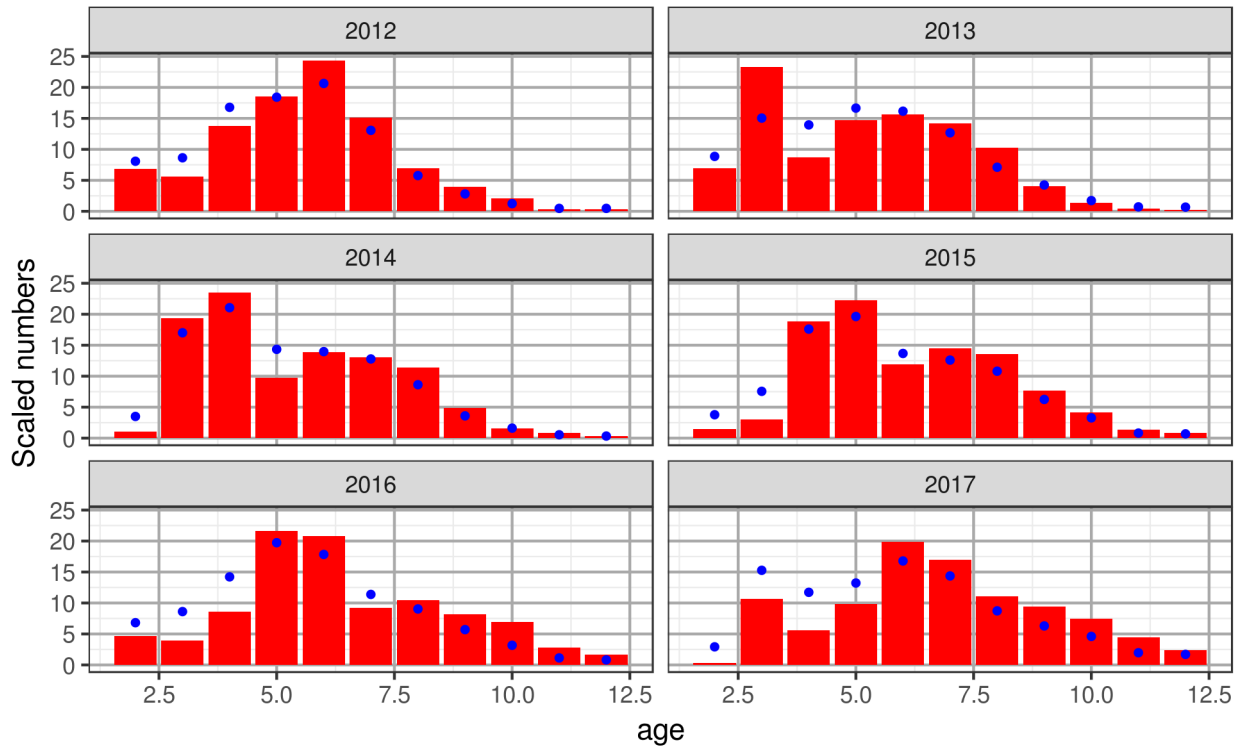


Figure 13: Age distributions base on the scanned fish (bars) and from catch in numbers (points)

2 Spatial distribution of tag recaptures.

Tagging of mackerel have since 2011 been conducted in 2 locations, west of Ireland (97% of the tags) and west of Iceland (3%). The tagging west of Ireland is conducted early in the year. The recapture location in the year of tagging has been similar for both the tagging areas (figures 14, 15 and 16). The average distance between tag and recapture location is the same for fish caught in the tagging year as those in later years. This does not prove complete mixing as the fish can only be recaptured where catches are scanned (figure 17). But taking account the number of fishes caught in the tagging year is rather high (table 2) the mixing of tagged individuals in the stock seems to be relatively fast (max 1 year, probably less).

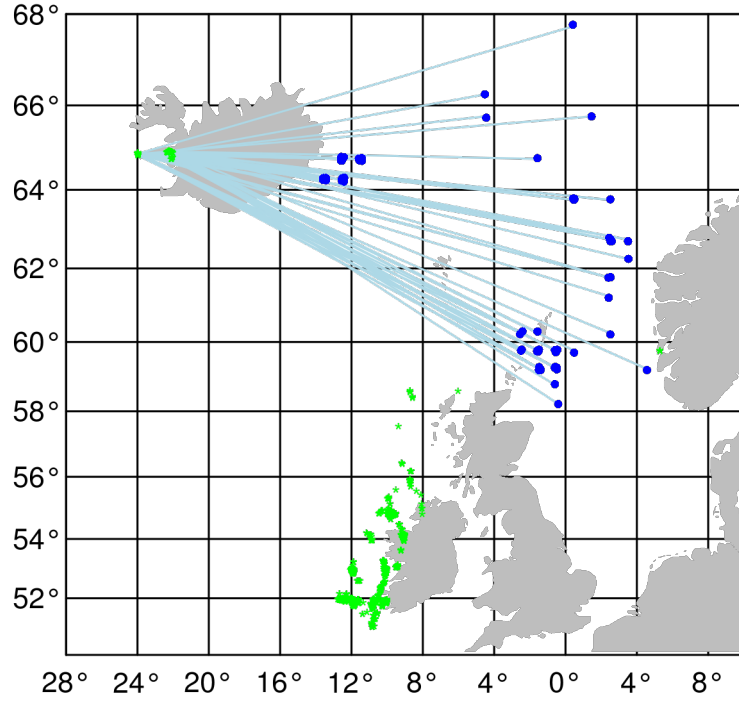


Figure 14: Same years recaptures from the tagging experiments in Iceland. The blue points show the recaptures

	0	1	2	3	4	5	6	7
Nyearsout	580	543	519	512	501	480	371	391

Table 4: Average distance against number of years out for recaptured tags

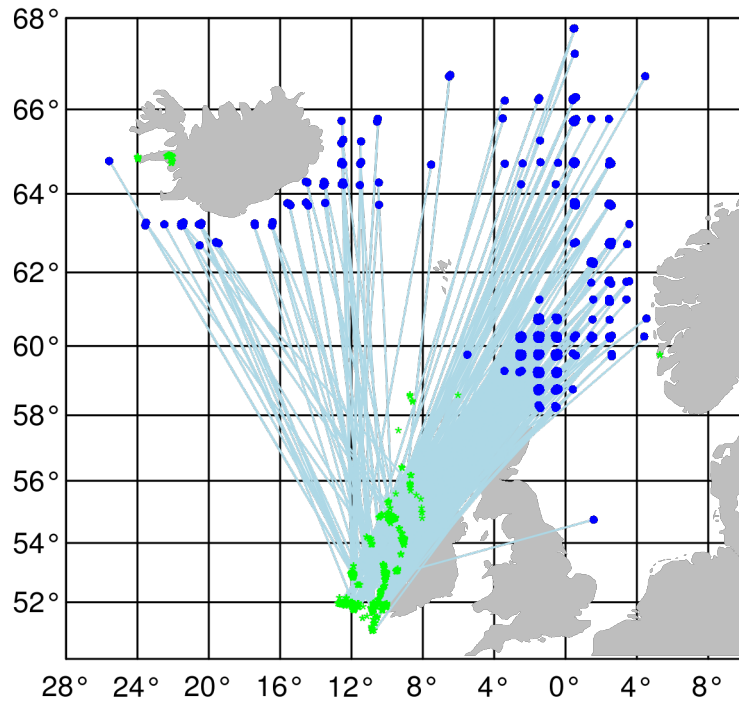


Figure 15: Same years recaptures from the tagging experiments. The blue points show the recaptures

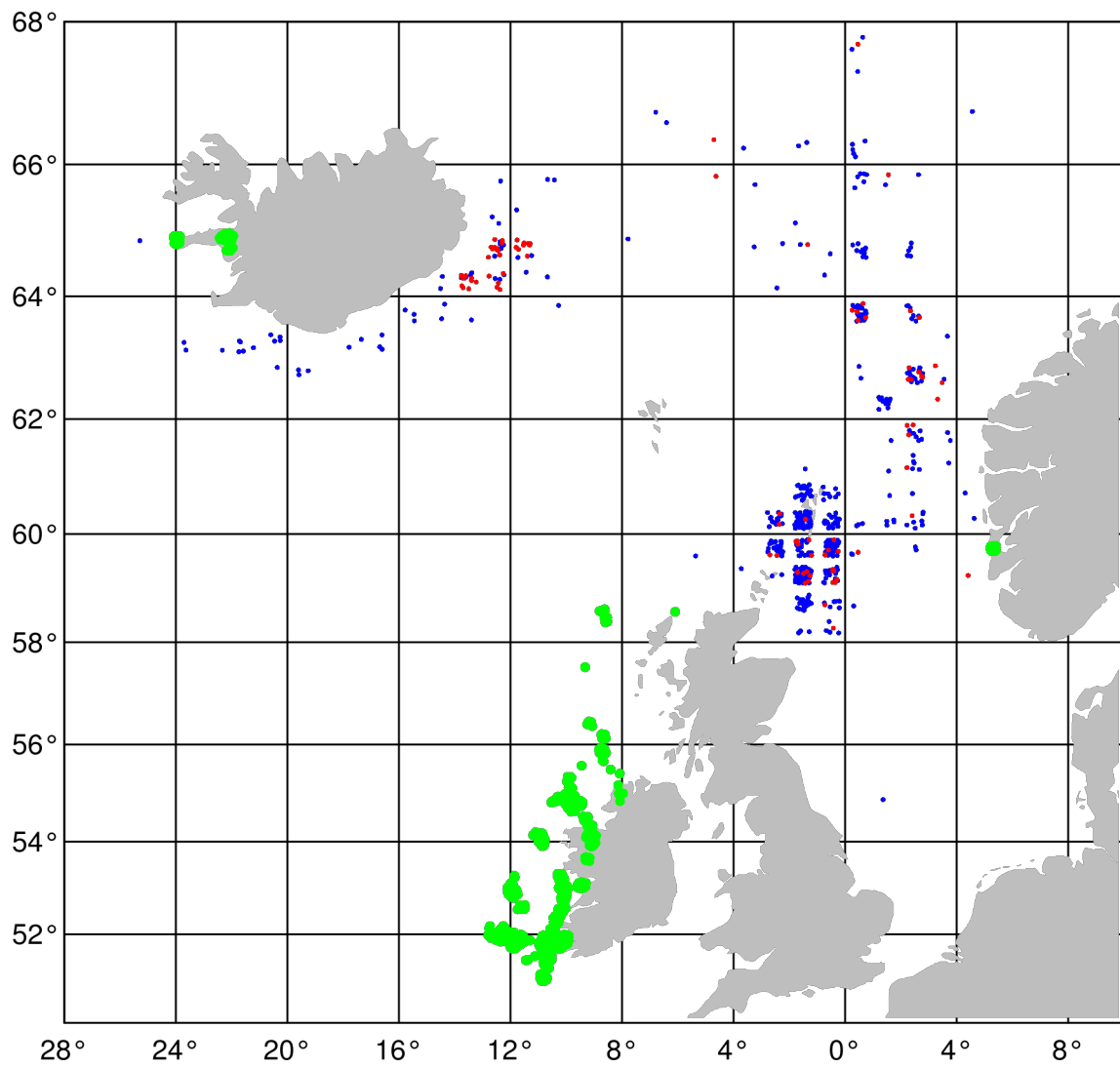


Figure 16: Location of recaptures in the tagging year from tagging experiments near Iceland (red) and Ireland (blue). The green points show the tagging areas.

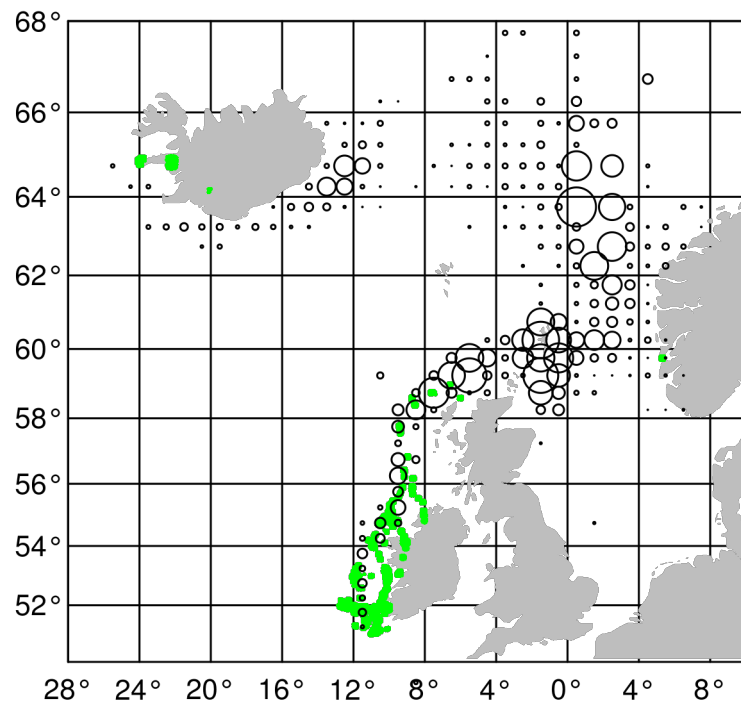


Figure 17: Distribution of scanned catches 2012-2017

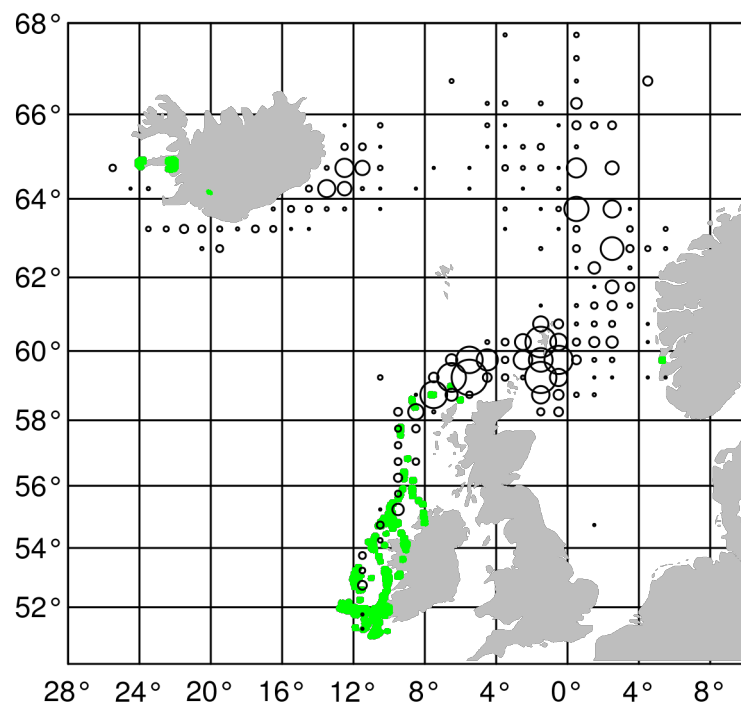


Figure 18: Number of recaptures by square 2012-2017

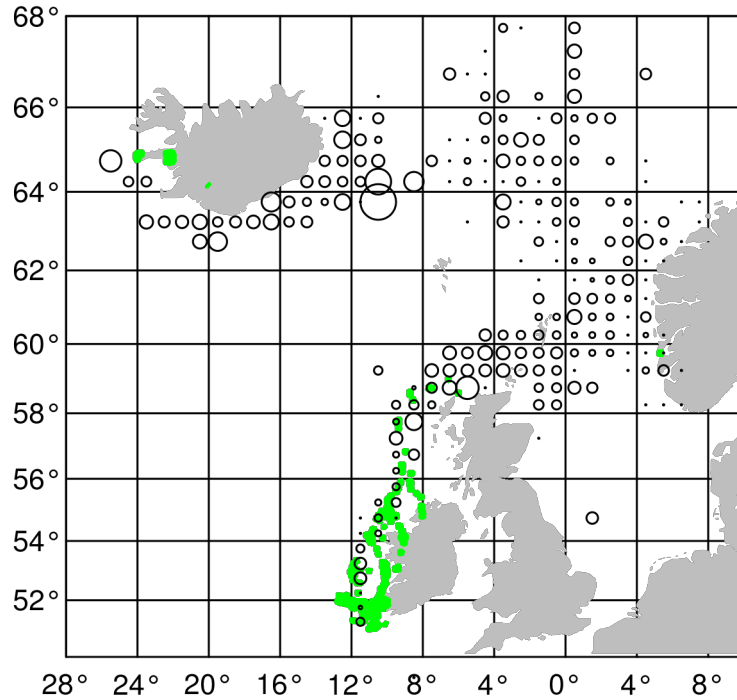


Figure 19: Distribution of number of recaptures divided by screened catches by Ices rectangle

3 Additional factors

Two factors that can affect survival are registered for 70% of tagged fish. The columns area called **release_waves** and **release_birds**. Their effect on proportion recaptured does not seem to be large (table 5)

	1	2	3
1		Birds	
2	Lots	Medium	None
3	0.987	0.988	0.985
4		Waves	
5	Lots	Medium	None
6	0.013	0.014	0.011

Table 5: Proportion recaptured for different levels of birds and waves when the fish is released

@

4 How to proceed

There are some problems with the use of tagdata in the assessment, most of them related to very high tagging mortality that varies with size, between years etc. The factors causing the mortality are known but have not been quantified and they will always be variable. Comparison with the older steel tags is interesting as the survival rate (plus all other confounded factors) is much higher there. Still it looks like a number of factors in how the tagging is conducted has improved (see Dankerts short report).

As the tags are modelled now, each tagged fish is aged via age-length keys and the tagged fish summed up by age group and tagging year. Each scanned fish does therefore not have an unique age or rather an agedistribution. When this fish is scanned it does again have the age distribution. For each recapture year the scanned number at age is compiled from the number of scanned fish and age samples from the factories that scan the fish. The recaptured fish from each tagging year is certain proportion of the scanned fish, the same as the proportion of the stock that was tagged and survived the tagging in a given year (tagloss not included). The comparison between observed and predicted numbers is done age by age for each tagging year and recapture year, This leads to quite many comparisons that are done by using negative binomial distribution that unlike a Poisson distribution has a possibility to estimate overdispersion.

The problem with this approach is that correlation in the data is neglected. Most of the correlation is between agegroups in the same tagging year introduced by interannual difference in survival from tagging experiments. By ignoring this correlation the number of degrees of freedom of the tagging data is increased and their weight in the likelihood increased. If correlation was to be modelled some kind of multivariate normal distribution would have to be used. What kind of noise is introduced in the data by double use of age-length keys is difficult to say. Ignoring correlations leads to increased weight of the tagging data in the likelihood function. Ignoring correlation in the IENSS survey has relatively large effect on results of this years assessment increasing SSB 2018 from 2.4 -3.2 million tonnes or so. The problem we have is all about weight of contradictory data. The only good solution is to "believe in something", "strong belief" has always simplified the life of people.

What could be done is to use more aggregated measures but with most of the code similar to what is used now. Based on the age distribution and number of tagged fish the total predicted numbers each year is compiled for each combination of tagging year and recapture year. When the predicted numbers are calculated they are either based on the total number scanned each year and that the number caught at age predicted by the model or that the number scanned per age is compiled as it is now based on samples from the scanning factories.

Number of tags with catcthid 3348 Number of tags with recapture year 3989

Some comments on modelling tag data in SAM for the mackerel assessment and a general comment on the “process error” in SAM

Magne Aldrin, Norwegian Computing Center

Introduction

In the first part of this note I discuss some aspects of how tag data have been modelled in the mackerel assessment, and sketch some alternative ways to model it. In the second part I discuss how the “process error” in SAM should be interpreted and argue that it should be included in the observation equations.

Tag data versus other data types

When several and quite different data sources are included for estimating a model, the importance of each data source depends on its contribution to the total likelihood, which again depend on the actual formulation of the model. Therefore, the exact model formulation may be more important when several data sources are included than if only one data source is used.

For instance, if you have linear regression problem where the errors (residuals) are correlated and non-Gaussian (non-normal), you will still get unbiased and consistent estimates of the regression coefficients if you assume Gaussian, uncorrelated errors and fit the model by maximum likelihood. (The uncertainty estimates will be wrong, however.) On the other hand, if two data sources are available with information on the same phenomena, ignoring correlations within one data source, but account for it for the other, will give too much weight to one of the sources.

In the 2017 and 2018 mackerel assessments, correlations between ages within the same year were included for the survey index. Even this was an improvement of the model, it probably implied that the survey index data were down-weighted compared to other data sources, since the correlations were positive. The tag data, on the other hand, were treated as uncorrelated, and these tag data consist of many observations. I haven't analyzed those data, but it is reasonable to expect that they are correlated in several ways, both a) between the various recapture years of the same age group tagged the same release year and b) between age groups tagged the same year, and c) perhaps also between observations within the same recapture year.

There is also another problem with the model for the recapture data. The RFID recapture data used in the assessment are not integer valued. The reason is that the length of a fish is measured, and then the fish may be partly allocated to several ages by an age-length key. However, in SAM, the recapture data are assumed to be negatively binomial distributed, and the data are rounded to the nearest integer values before the likelihood are computed. This is a problem for small recapture values, less than 3, say. In the RFID data available at stockassessment.org from the 2017 assessment, 18% of the observation are larger than 0 but less or equal 0.5 and thus rounded to exact 0, 14% are between 0.5 and 1.5 and rounded to 1 and 9% are between 1.5 and 2.5 and rounded to 2. The consequences of this rounding procedure is unclear, but at least it violates the assumption of the negative binomial distribution of the RFID data.

Alternative ways of including the tag data in SAM

One may use the tag data to construct two abundance indices, one based on the steel tags and another based on the RFID tags. These can then be included in SAM as indices in exactly the same way as the IESSNS survey index, including correlations between ages within the same year. In this way, SAM would probably balance the influence of the three different data sources well.

However, this solution is not perfect. First, it is not obvious how different recaptures (from one, two, three years after release etc.) should be optimally weighted together. Second, it ignores possible correlations between indices for different years, due to recaptures in the same year. It is probably possible to construct a model for the raw RFID tag data with length data on each fish and combining these with the age-length key within SAM, including the appropriate correlations, but I am not sure if it is worth the effort.

The process error in SAM – what is it?

Some of the working documents for the assessment meeting have discussed the potential effect of the process error in SAM. However, I miss an interpretation of what this is. In the population model described at page 97 in the original SAM paper of Nielsen and Berg from 2014, the abundance of a cohort change from one year to another by

1. F , fishing mortality (rate), modelled as unknown and random
2. M , natural mortality (rate), assumed to be known and 0.15 for mackerel
3. PE , process error

This process error must be either mortality or migration. Since fishing mortality is already accounted for, it must either account for random variation in the natural mortality rate (do anyone believe that it is exact 0.15 for all age groups all years?) or for migration into or out from the system, but the difference between these interpretations does not matter.

What does it mean to set the process error (i.e. its variance) to 0? That mean that the natural mortality rate is 0.15 always. If that is not true, the only way the SAM model can adjust for it is to change the fishing mortality rate. This can be seen by comparing Figure 1 with F_{bar} in Figure 3 in the working document “analyses of the process error ...” by Thomas Brunel, uploaded at the sharepoint. In SAM, a positive process error PE means that $(M - PE)$ is less than M . My interpretation of Figure 1 is that the real natural mortality is less than 0.15 most of the time, but larger than 0.15 in the years after 1990 and after 2000. Therefore, when PE is set to zero and $M=0.15$, the estimated F is increased in the periods where the real natural mortality is less than 0.15, see Figure 3.

My conclusion is that the process error term should be kept in the population model. Setting it to 0 doesn't solve the potential problem of assuming uncorrelated process error.

This brings me to a dispute I (and my colleagues) have with Nielsen and Berg. We mean that since PE goes into the population exactly the same ways as F and M , it should also be included in the observation equations at page 98 in the Nielsen and Berg (2014). The observation equations are where the observed/estimated catch are linked to the modelled catch and when an observed survey index from a given time in a year is assumed proportional to the modelled abundance at the same time. We have written a short note on this, and Nielsen and Berg have written an answer, see references below.

I am not sure if it is worth using time on discussing this more general SAM modelling issue during the mackerel benchmark, but I mentioned it since several people seems to be interested in the effect of the process error (=migration or variation in natural mortality).

References on SAM and process error:

Nielsen, A. and Berg, C. (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, Vol. 158, p. 96–101.
<https://doi.org/10.1016/j.fishres.2014.01.014>

Aldrin, M., Aanes, S. and Subbey, S. (2019). Comments on incongruous formulations in the SAM (state-space assessment model) model and consequences for fish stock assessment. *Fisheries Research*, Vol. 210, p. 224–227. <https://doi.org/10.1016/j.fishres.2018.08.001>

Nielsen, A. and Berg, C. (2019?). Response to: Comments on incongruous formulations in the SAM (state-space assessment model), To appear in *Fisheries Research*,
<https://doi.org/10.1016/j.fishres.2018.09.008>.

RFID tagging data sensitivity run on the WGWIDE 2018 assessment.

Claus R. Sparrevohn

Background and purpose

This working document presents a series of sensitivity analyses, done in order to better understand the performance of the RFID tagging data series in the 2018 mackerel assessment. The background for this work is, that at the 2018 WGWIDE meeting in Thorshavn concerns about the influence from the RFID tagging series on the performance of the assessment was raised. One of the concerns that was raised in Thorshavn was that it appeared as if the impact of including RFID data was somewhat higher the expected, based upon the experience from the interbenchmark carried out in 2016. The reason for this fully understood during the meeting as further exploitation was not carried out during WGWIDE, partly due to lack of time. All that was achieved during WGWIDE was some preliminary runs excluding/adding specific recapture years. This document provides the full analyze of adding and excluding specific year(s) and compare how the SSB estimate behave.

Assessments presented throughout the document are all based upon the 2018 WGWIDE run (WGWIDE2018.V2) such the run called the “baserun” is identical to assessment which provided the basis for the advice issued the 28th of September 2018. Neither catch data or any of the fishery independent surveys has been changed except for the RFID data series, where RFID data has been excluded from the run according to the list below:

- Recapture years:
 - Forward from 2012 to 2017,
 - Backward from 2017 to 2012 (2012 did not converge),
 - Excluding single years (2013-2017, 2012 and 2015 did not converge)
- Release years:
 - Forward from 2011 to 2016,
 - Backwards from 2016 to 2011,
 - Excluding single years (2011-2017)
- Recaptures out:
 - less than 2 to 7 years
 - More than 1 to inf. (did not converge above 2 years)

All results are compared to the 2018 WGWIDE agreed assessment, referred to as the baserun, and identifiable in the plots by having a thick line. On all figures the SSB is only shown from 2000 and onward. Since no changes in SSB was observed prior to this period, independent on the RFID data used, this was done to improve the visual presentation.

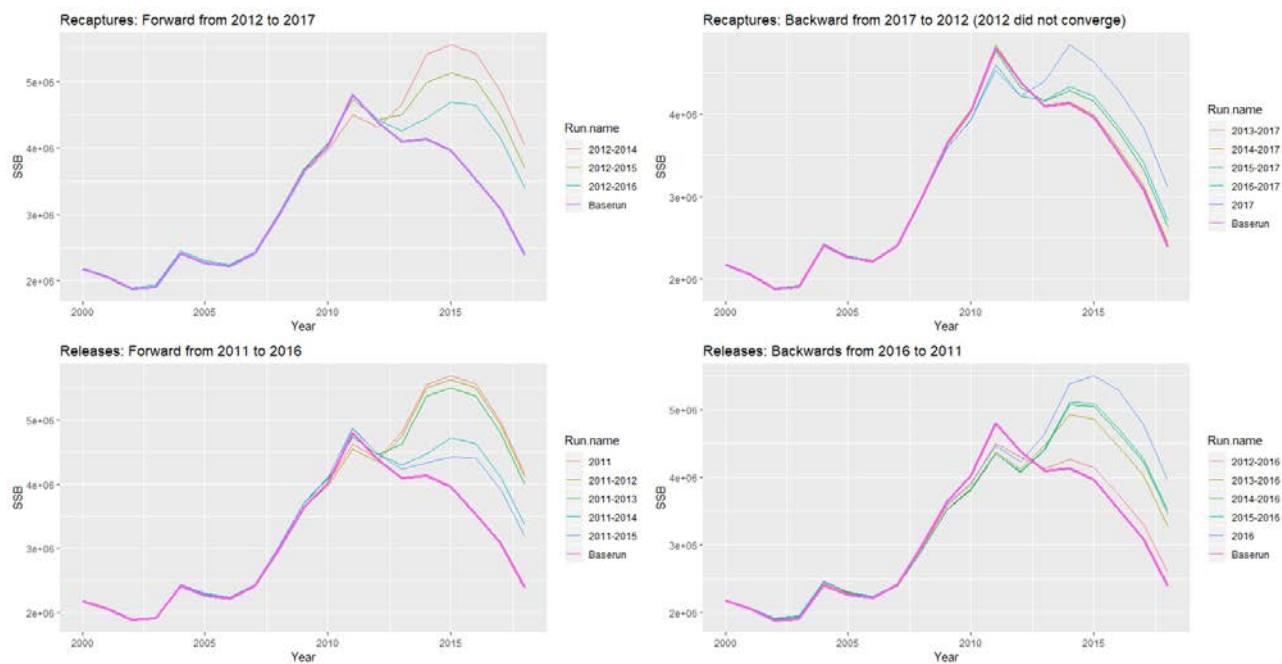


Figure 1. Thick line is the base-run and each of the thinner lines represent modified run.

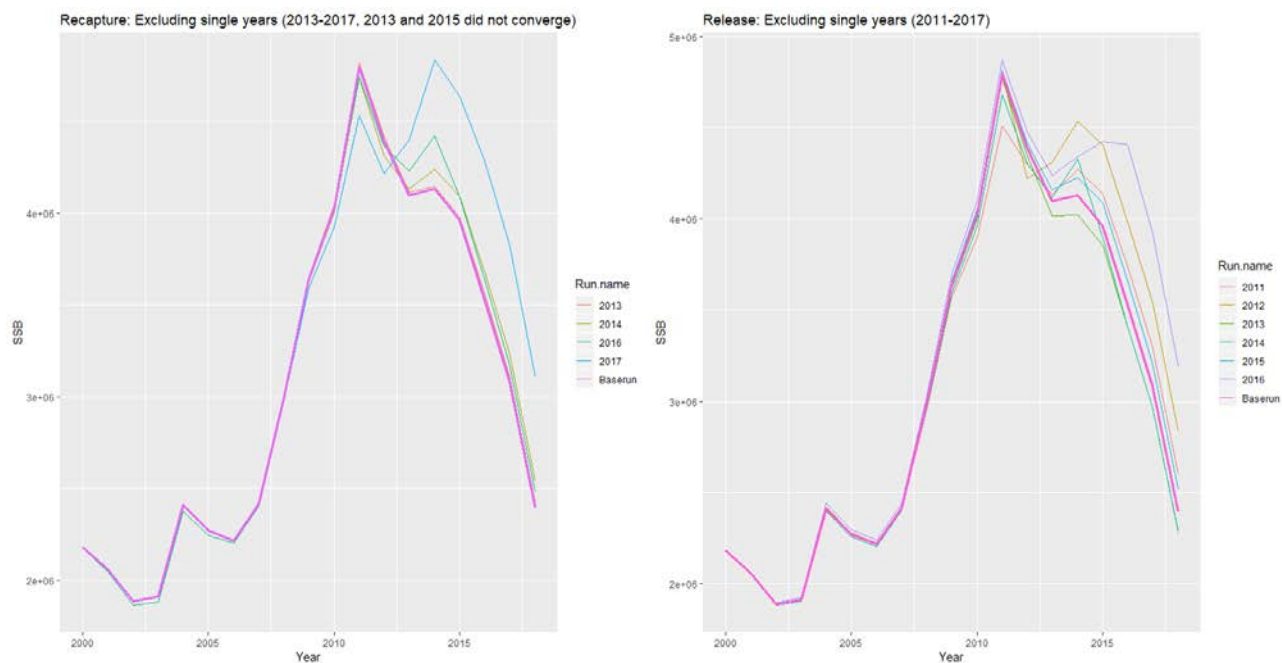


Figure 2. The thick line shows the base run for comparison. Excluding single recapture years from the run did cause the model not to converge in some cases.

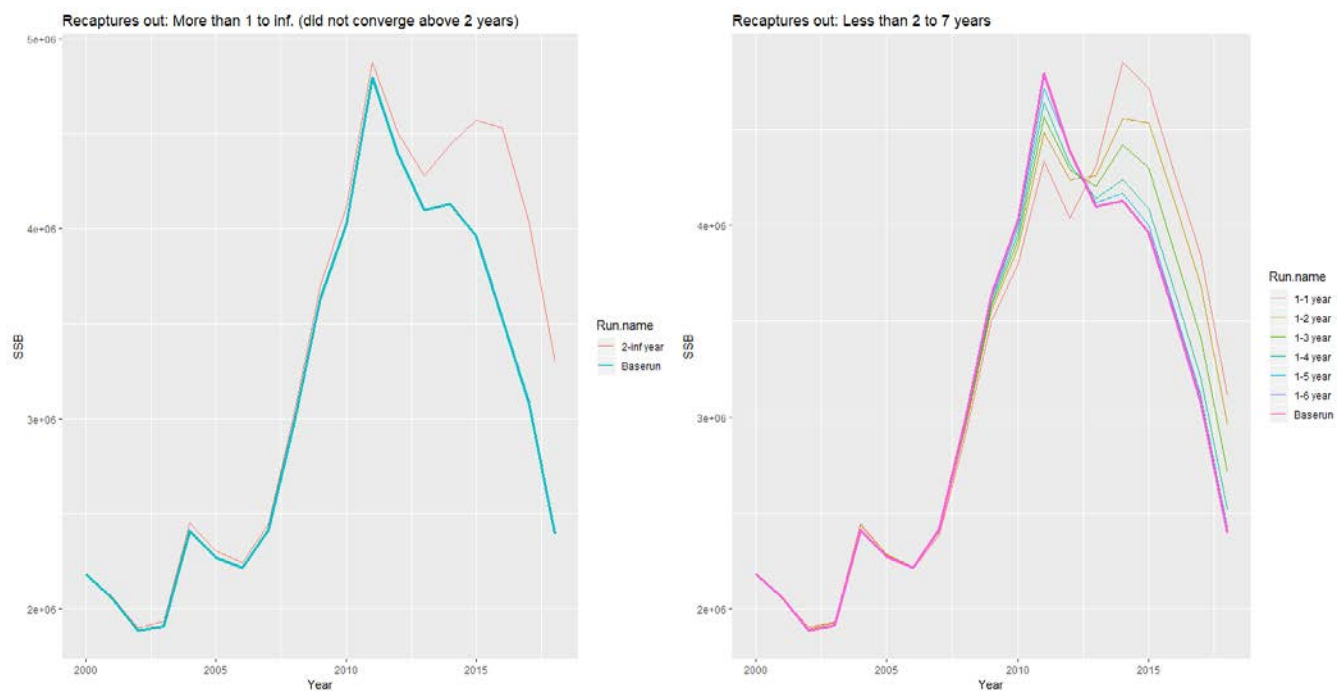


Fig. 3 The effect of number of years between release and later recapture.

Some thought and suggestions

I should be kept in mind that what is really estimated by the tagging data is the cohort size at tagging time. Thus, release in 2011 will only provide inform on the cohort size in 2011. The likelihood for this estimation is the sum of likelihoods (not the sums of recaptures combined into one likelihood) for the following recapture years.

It should also be kept in mind that scanning a number of mackerels for a given cohort, without detecting any tags, does also add to the likelihood.

For all explorations, except excluding single release years, limiting the years of tagging (recaptures or releases) creates a trend in the SSB, such that the more tagging information the lower the SSB estimate.

Two what extend the trend in SSB when limiting the dataset is a result fewer data and hence a lower weight to this dataset, or if it is due to the trend in data is unclear.

NEA mackerel



Alternative assessment

Höskuldur Björnsson

August 31st 2018

1 Input data

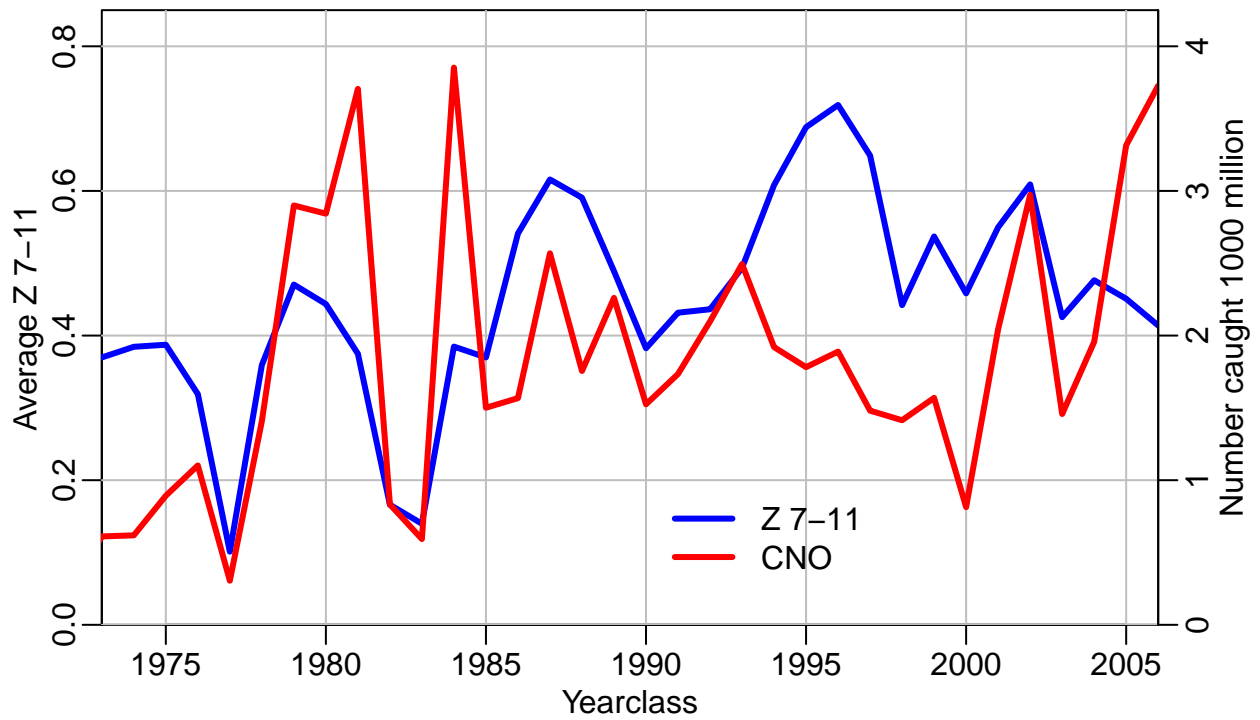


Figure 1: Z_{7-11} and total number caught for a number of yearclasses

Analysis of catch in numbers indicates that Z_{7-11} was lower before 1986 than later (figure 1). The most notable thing about the Z values are the low values for yearclasses 1977, 1982 and 1983 that are all small (the numbers caught is very small). As substantial amount of age 0-2 is caught in this period the total numbers for yearclasses 1973 - 1976 are too low, but the data start in the year 1980.

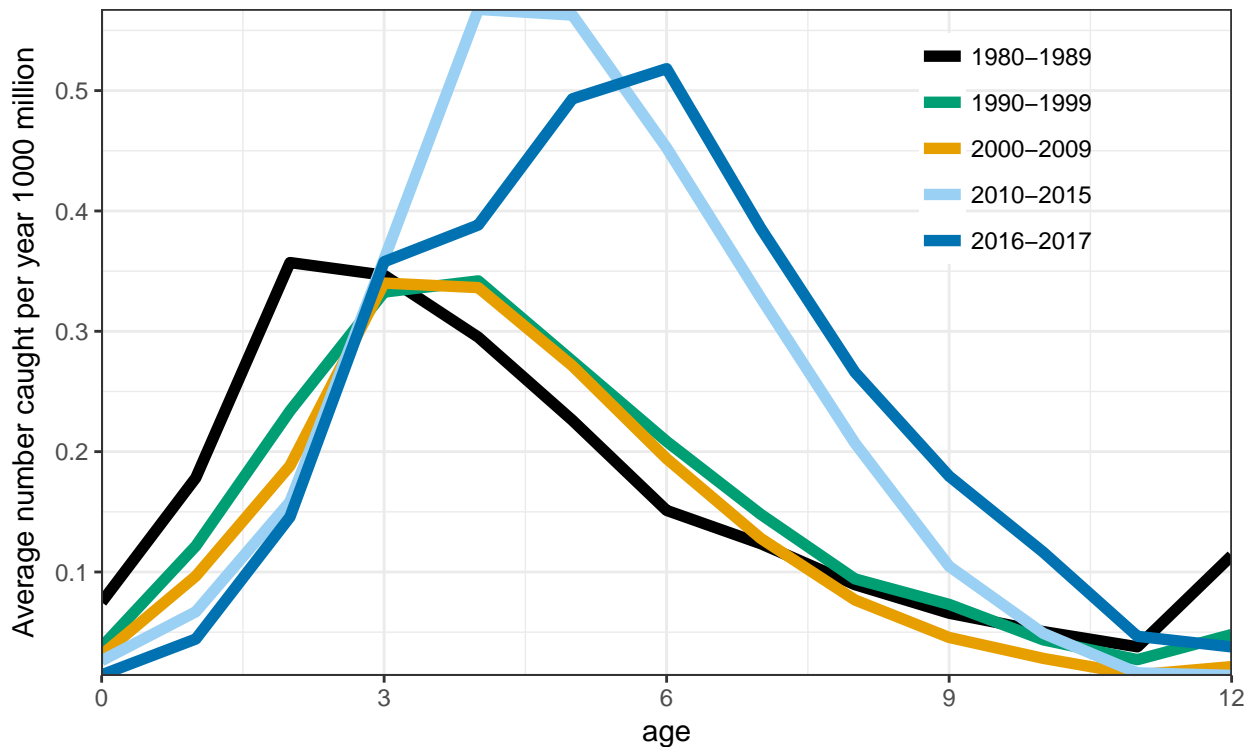


Figure 2: Average catch in numbers by age split by period

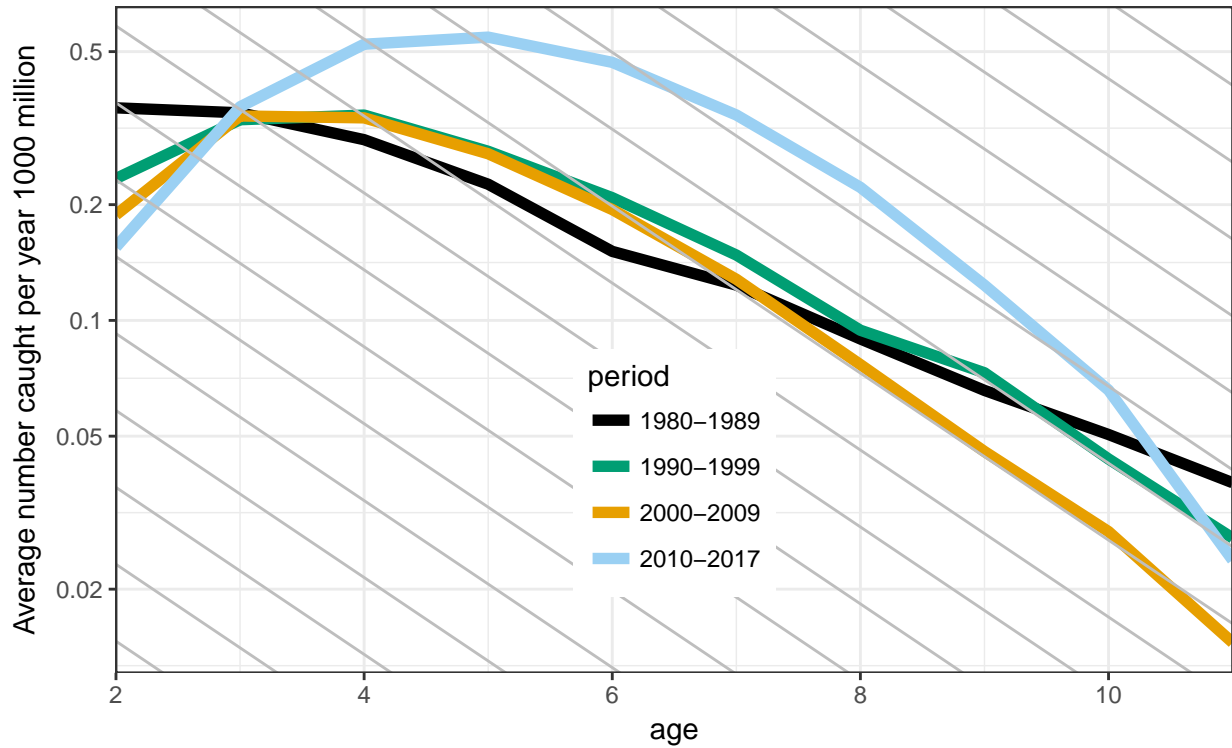


Figure 3: Average catch in numbers by age and period. The grey lines correspond to $Z=0.5$

Looking at catch in numbers by age split by periods (figures 3 and 2) indicates change in selection with much more young mackerel caught in the first periods, and also much more of the plus group. Interpreting the plot directly as change in selection is not trivial as they might also represent changes in yearclass size. After 2010 the yearclasses entering the fishery might be larger than those leaving. The lack of old fish after 2000 does still suggest higher fishing mortality after 1995 than before. But the results do at least indicate that in a separable model the selection should be allowed to change once or twice during the period 1980-2015. In most of the runs presented the selection is allowed to change once, in 1996.

Catch curves suggest that Z is close to 0.4. (figure 4)

2 Pelagic survey

Indices from the pelagic survey 2010-2018 show relatively much old fish in recent years but reduced number of younger fish compared to the years 2013-2014. Age 1 is very high in 2015, 2017 and 2018 that might indicate change in distribution and/or reasonable recruitment. The index at age 2 seems to be a better measure and might be used as recruitment indices.

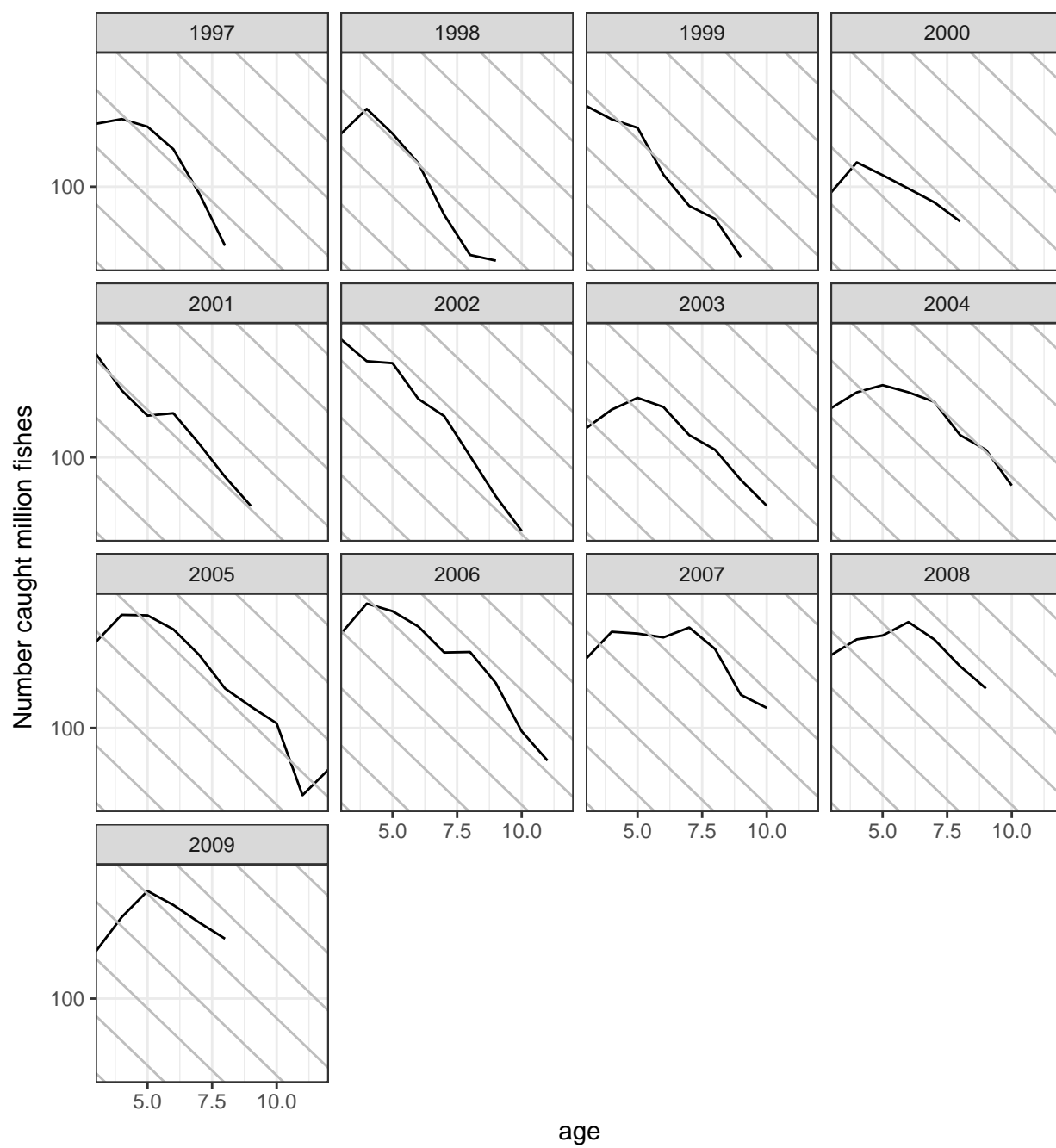


Figure 4: Catch curves from commercial catch. Grey lines correspond to $Z=0.4$

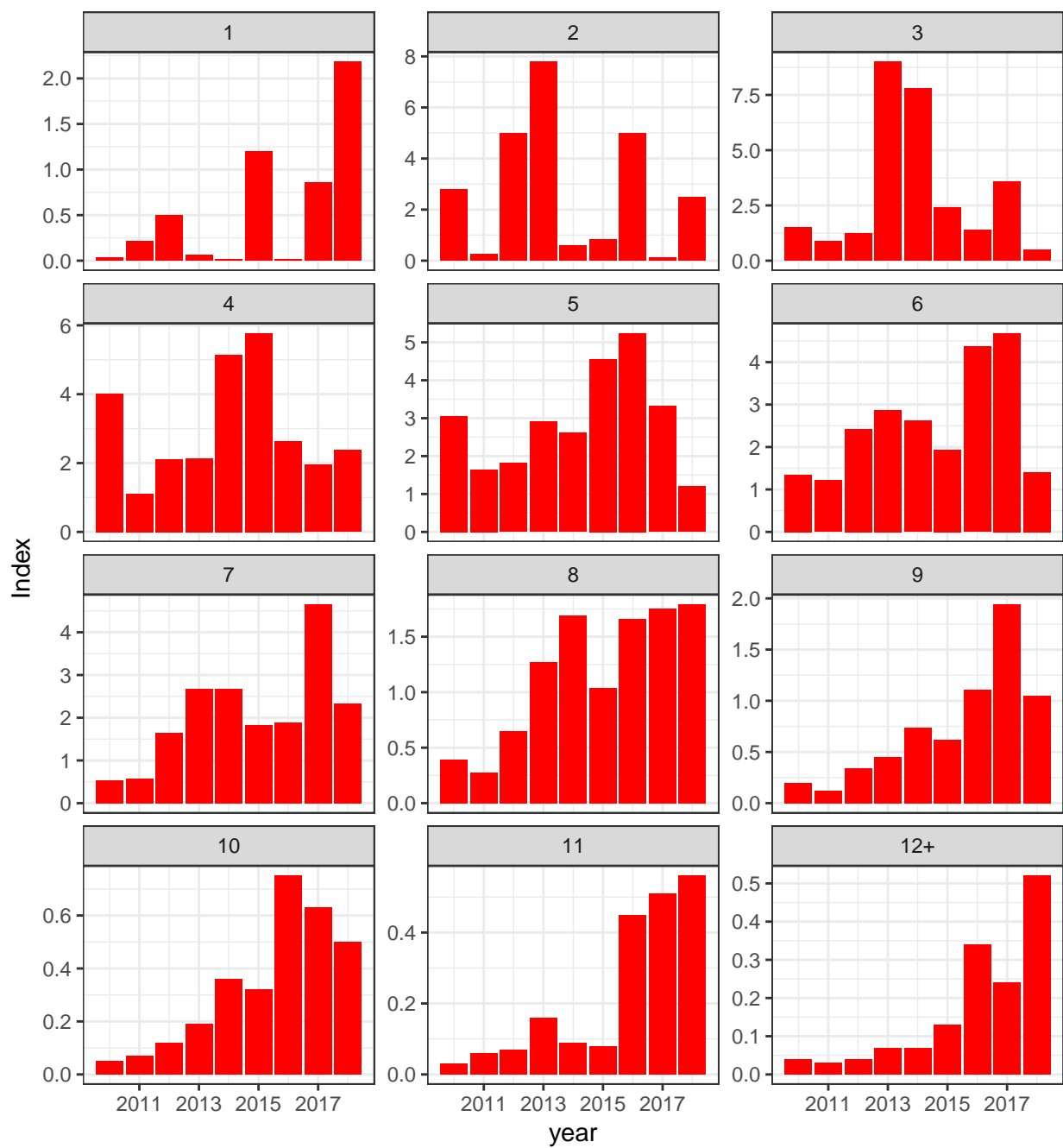


Figure 5: Index by age in the pelagic survey 2010-2018

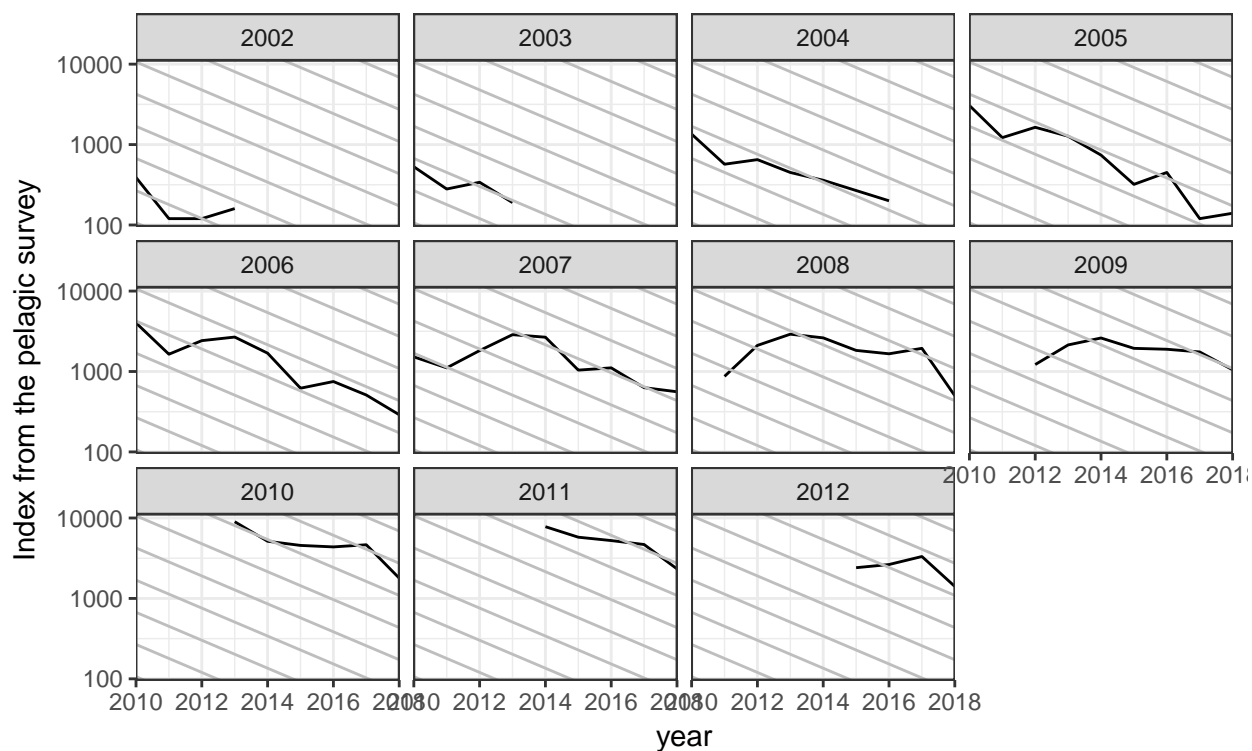


Figure 6: Catch curves from the pelagic survey. The grey lines show a slope of 0.4

Catch curves from the pelagic survey do suggest that the Z is less than 0.4 (estimated average Z of ages 6-9 2010-2017). This observation is though confounded with how much catchability increases with age and that the coverage of the survey has not increased much.

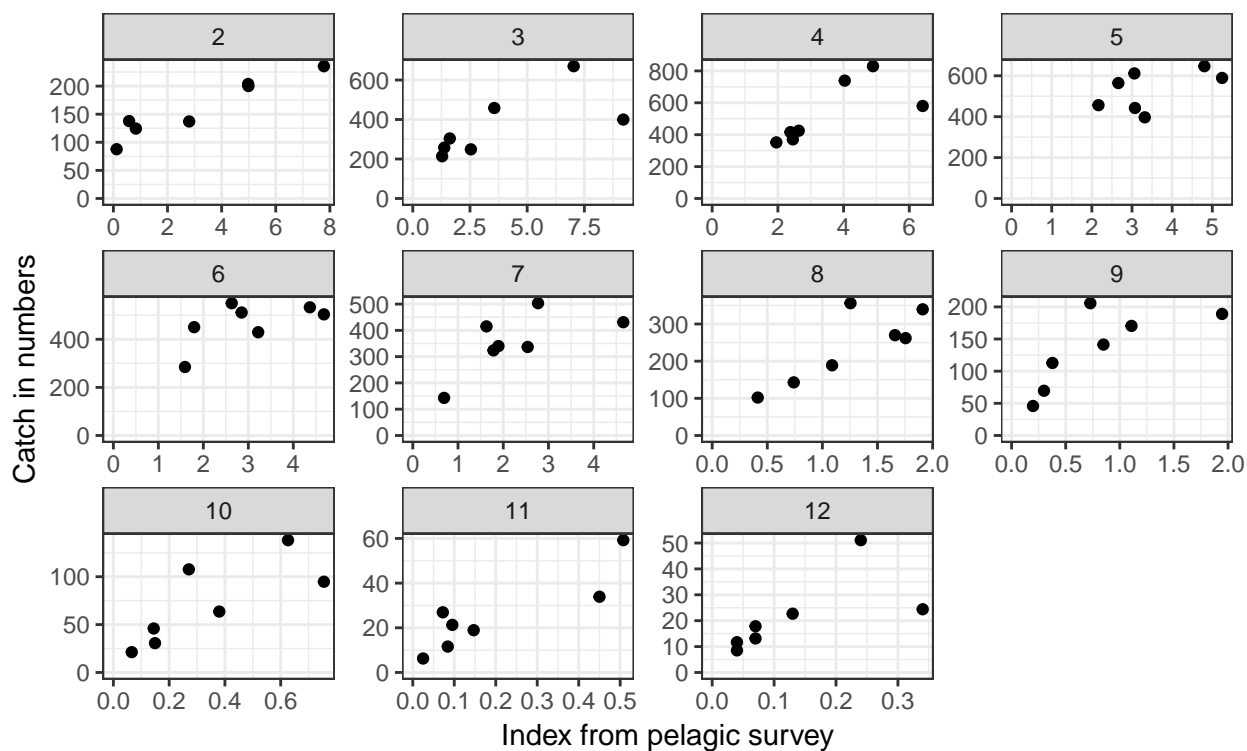


Figure 7: Catch in numbers by age vs indices from the Pelagic survey for the years 2010 and 2012:2017.

Catch in numbers and index from the pelagic survey fit well for the older age groups but not as well for the younger age groups where contrast in data is less, especially in the catches.

3 Assessment

4 model runs and settings were compared, the first 3 from based on the separable model used for HCR evaluations last year.

1. Egg survey and Pelagic survey 2-12+
2. Egg survey Pelagic survey 2-12 and recr index
3. Egg survey and Recr index
4. The SAM model presented at the meeting.

The settings like how CV changes with age and the age when the Q in the pelagic survey is the same can be investigated more. The estimated parameter in the correlation in the pelagic survey was 0.48 between adjacent agegroup (AR1 type model), less than in the SAM run. Can possibly be caused by age 2 that is incorporated here.

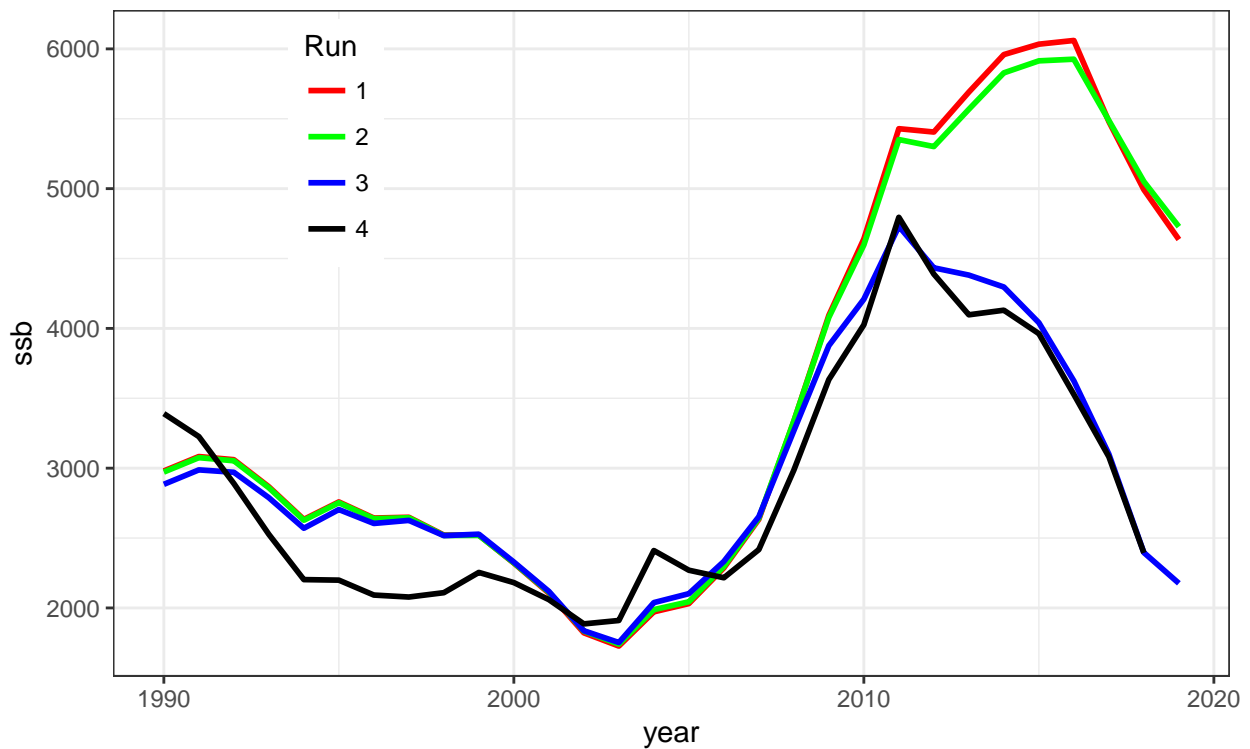


Figure 8: Spawning stock for the 3 model settings and from SAM (black)

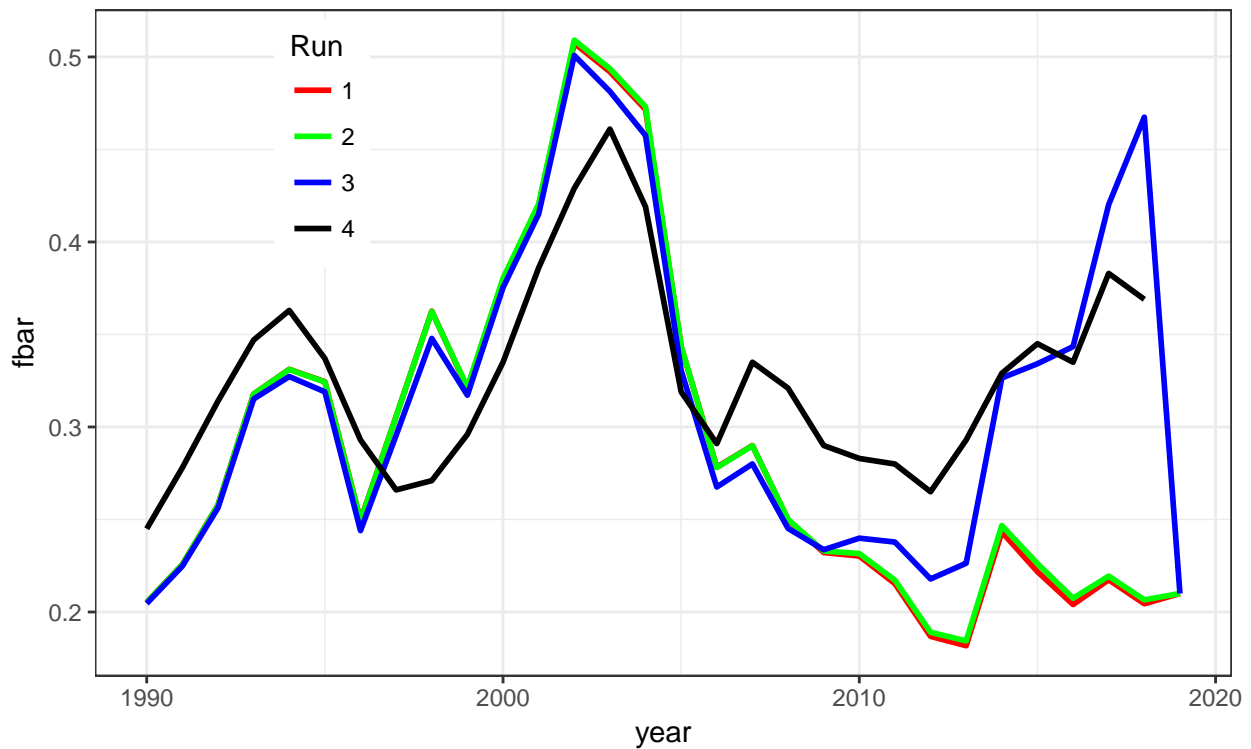


Figure 9: Average Fishing mortality for the 3 model settings and SAM (black)

Compared to the separable model not using not using the pelagic survey (run 3) the SAM model gives similar spawning stock in recent years (figure 9) but often lower F (much more stable F in spite of the same stock size). Compared to results obtained 2 years ago the separable model estimates higher CV on the egg survey and somewhat lower on the pelagic survey meaning it takes more notice of the pelagic survey.

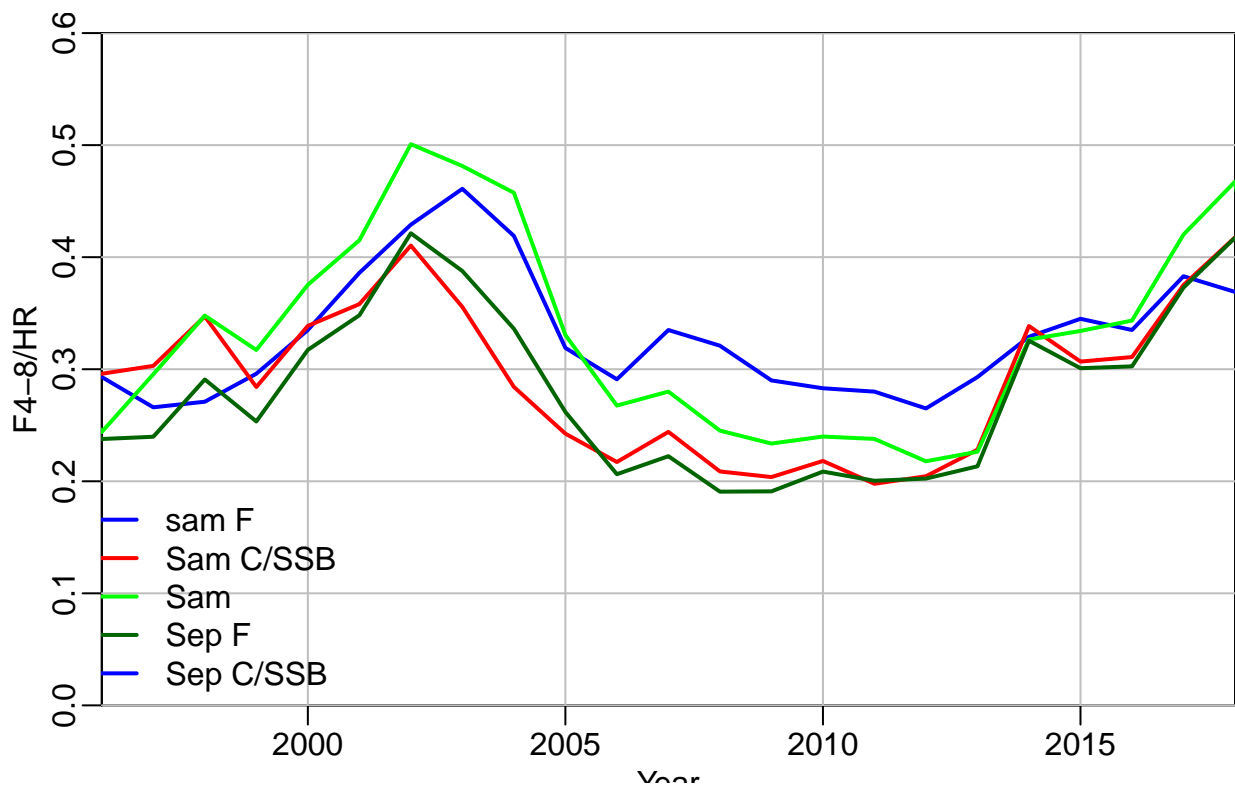


Figure 10: Comparison of different measures of fishing mortality in SAM and SEP not using the pelagic survey.

Figure 10 shows quite different F from the SAM and SEP models (run 3) but Catch divided by SSB is the

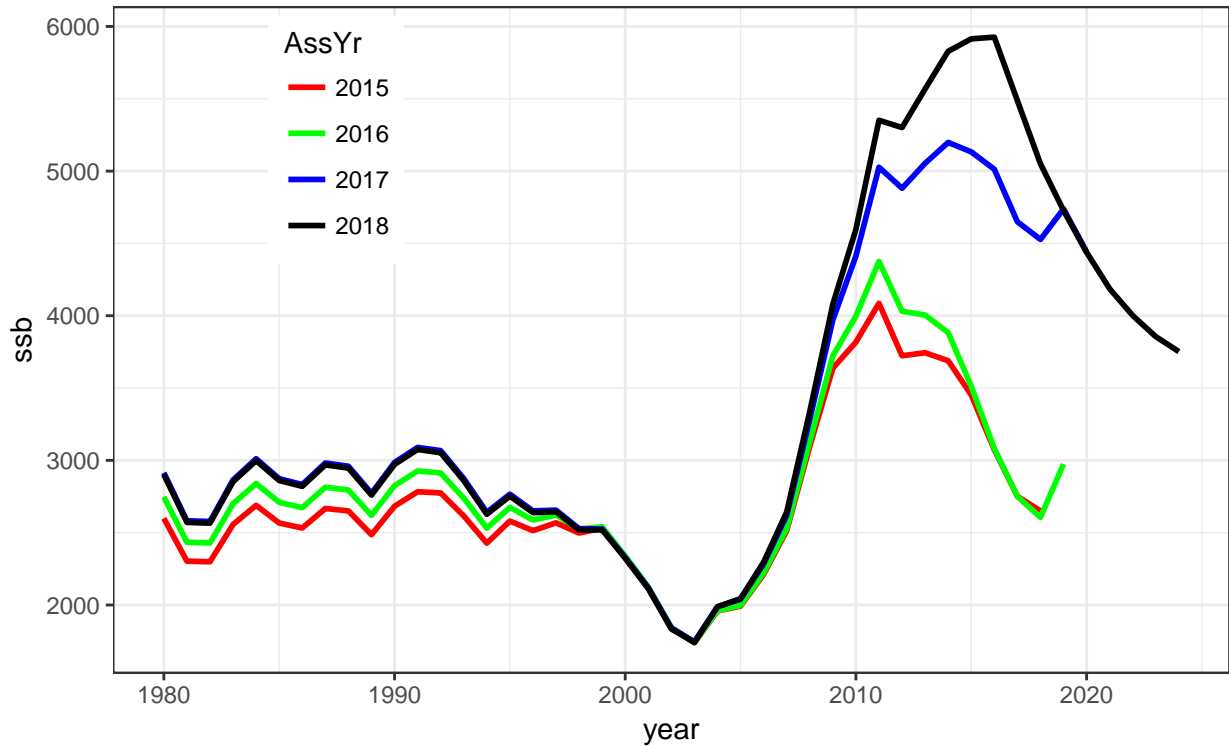


Figure 11: SSB from runs terminating in 2015-2018

same as SSB is similar. One possible reason is that SAM does in the end not use the "correct catch", what is shown here is correct catch/ssb

Retros from the separable model show constant upwards revision since 2015 (figure 11). What has been happening is that the pelagic survey is coming in and the egg survey going out. In 2015 and 2016 the estimated CV of the egg survey hits a lower bound (0.13) while in 2017 and 2018 the estimated CV is 0.26 and 0.32. The estimated CV and ρ for the pelagic survey have reduced at the same time. In the HCR simulations done last year but based on same data as the 2016 assessment a lower bound of 0.25 was put on the CV in the egg survey.

The estimated SSB in 2010 when the pelagic survey started has not changed much in the runs terminating in 2015-2018 (figure 11) due to normal convergence of a simple age based assessment model.

Behaviour of the model can be understood when looking at figure 15. Since 2015 the biomass in the assessment from 2000-2010 has not changed very much in the model. The egg surveys indicate that the stock in 2010-2013 was 50% higher in 2000-2005 and then it drops back down to the 2000 level in 2016. The pelagic survey is variable but shows the stock today at higher level than in 2010 and 2012 and most of the 2013-2017 values are considerably higher.

One interesting feature of the results is that the predicted survey biomass of the pelagic survey is close to the "real biomass" i.e $q = 1$ (figure 13). The 2007 and 2011 surveys that are not used in the assessment are shown there as points but they are apparently much lower than prediction. The picture of the pelagic survey as an absolute biomass estimate is not as nice if estimated q vs age is plotted, it is below 1 for younger and above 1 for older. (figure 14). The shape of the selection curve is also similar to selection of the fisheries (would be more similar if age of flat selection was the same.)

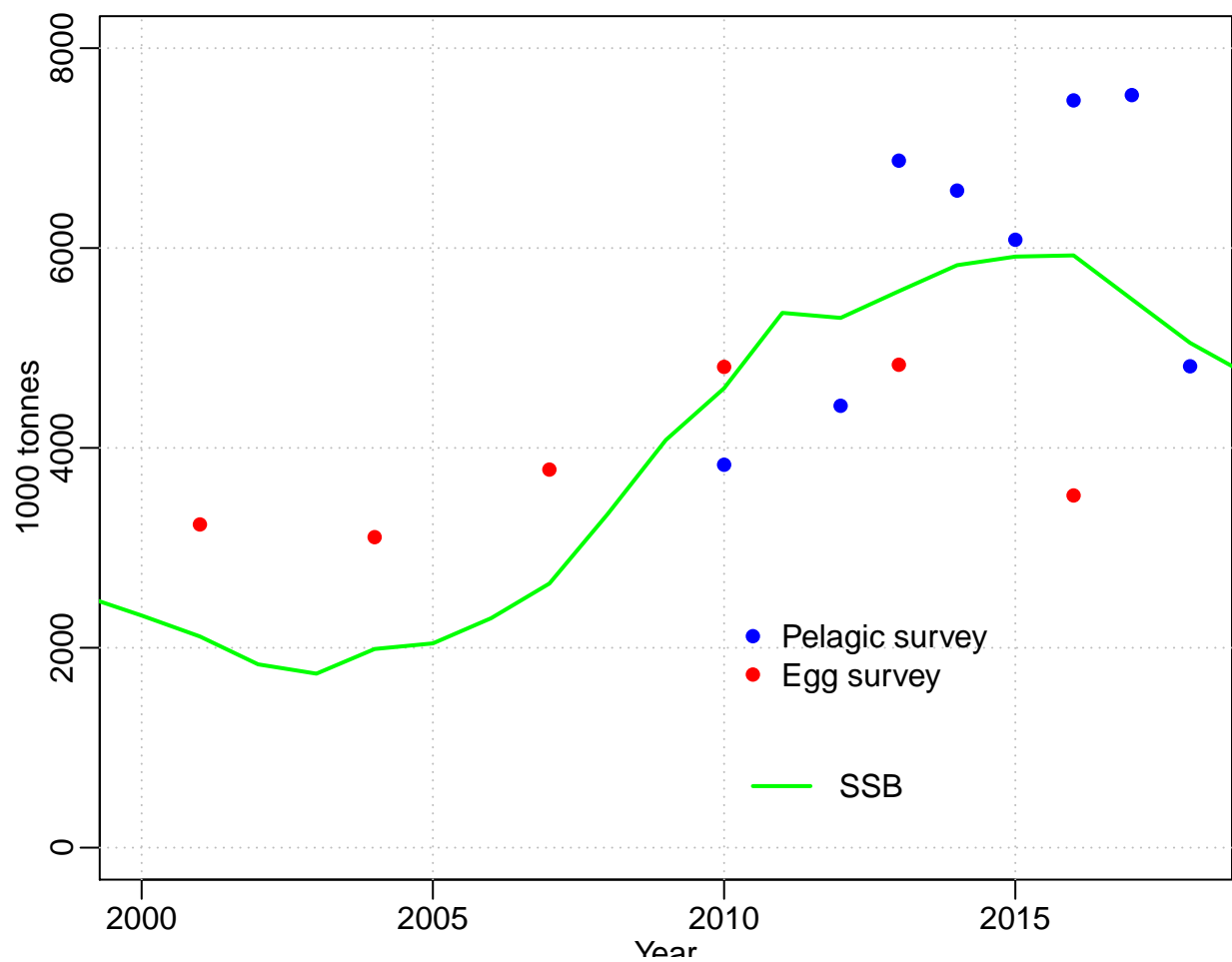


Figure 12: Spawning stock and eggsurvey index and aggregated index from the Pelagic survey.

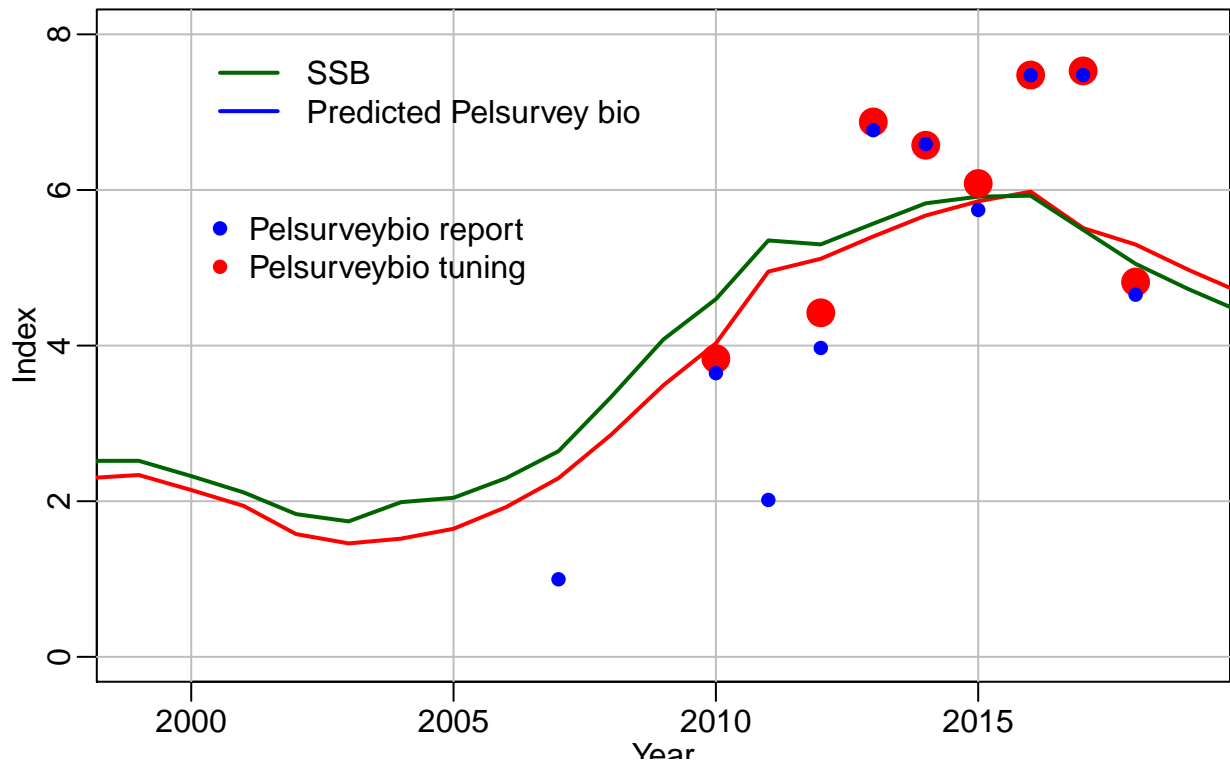


Figure 13: Observed and predicted survey biomass from the pelagic survey and Spawning stock. The red points show the points used in tuning but the blue points are from the report showing also the years 2007 and 2011

One interesting feature of the model setup with the pelagic survey is that using lower M gives better fit to data (not uncommon and does not necessarily say anything about M). If lower M is used the biomass from the pelagic survey does on the other hand start to exceed stock biomass.

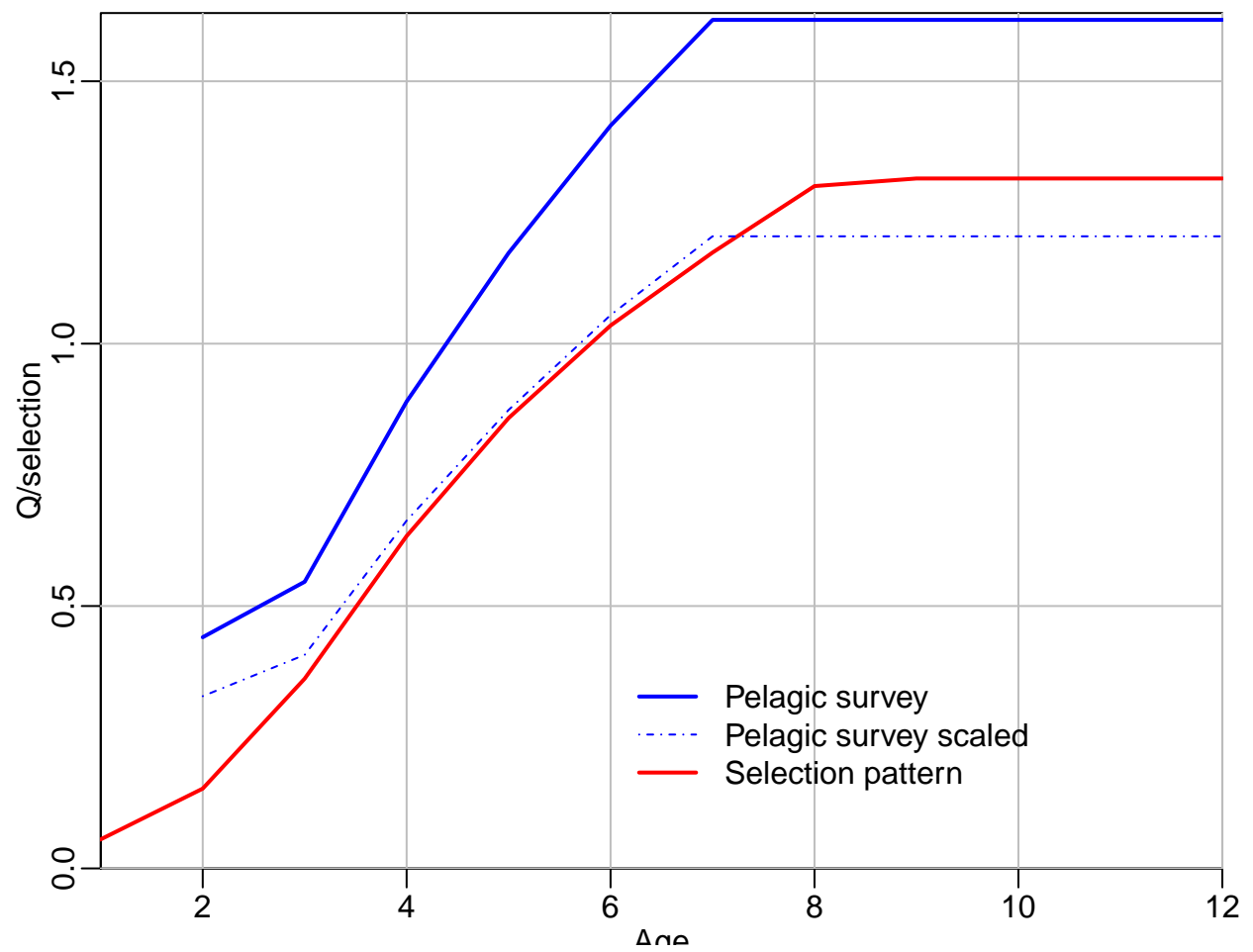


Figure 14: Estimated q vs age from the pelagic survey and selection of the fisheries

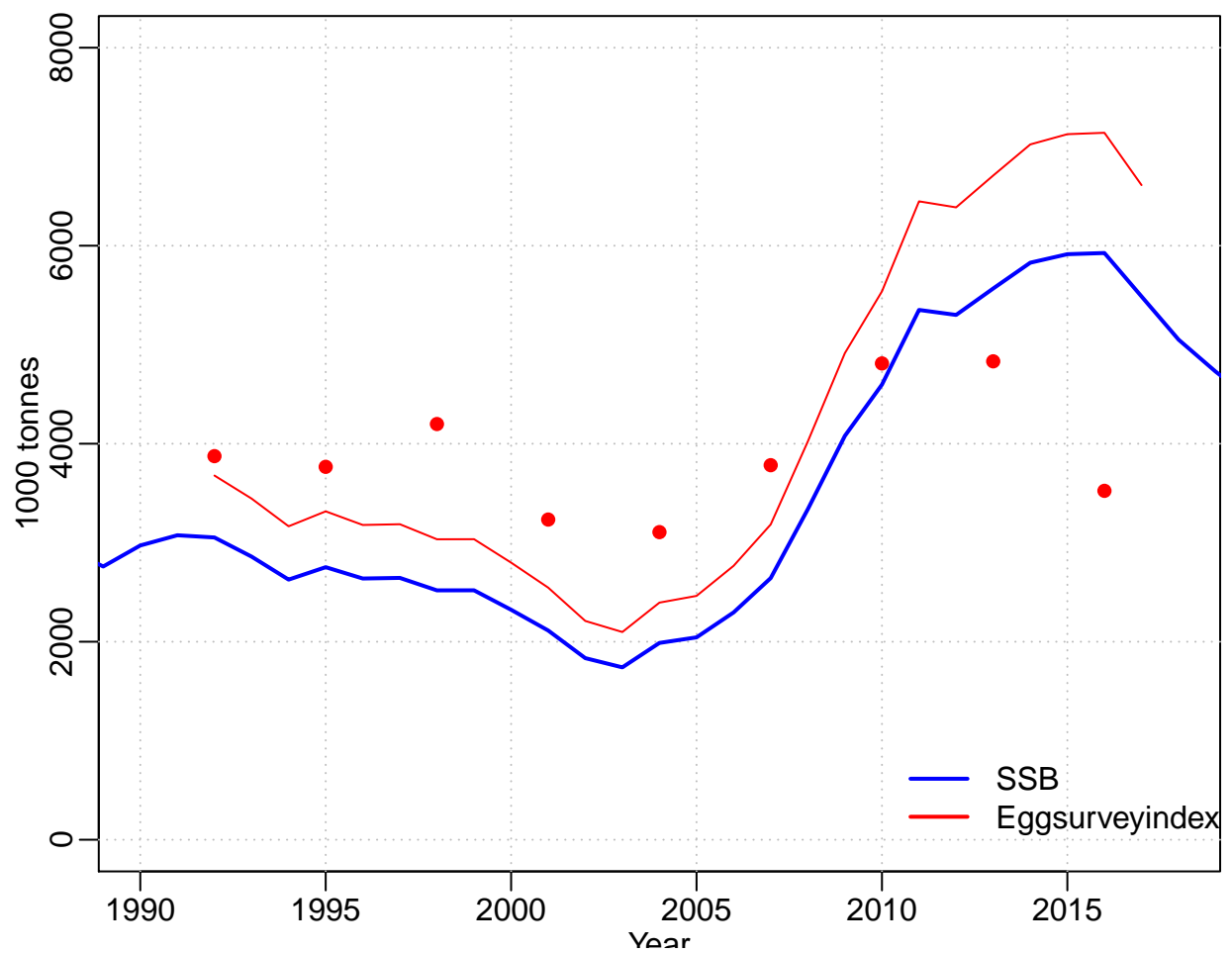


Figure 15: Spawning stock, observed and predicted eggsurvey index

Working document to the inter-benchmark workshop on the assessment of NE Atlantic mackerel, IBPNEAMac 2019.

On a possible serious bias in the egg biomass indexes on NE Atlantic mackerel.

Jens Christian Holst

Ecosystembased

Introduction

The egg-, survey- and tagging (equal tagging mortalities) biomass indexes used in the NE Atlantic mackerel assessment indicate very different stock trends. While the egg index shows a fairly stable and low stock since its start in 1992, the survey index and tag index, given equal tagging survival for steel and RFID tags, indicate a strong growth in the stock from around 2006-2007 and onwards.

A tagging experiment carried out by the IMR west of the British Isles in 2018 indicate no significant difference in tagging mortality between the old and the new tagging methods for mackerel, thus strongly indicating that equal tagging mortality between the steel and RFID tags is a valid assumption.

The survey and tagging indexes are in strong correspondence with the observations of pelagic mackerel skippers throughout the NE Atlantic; the mackerel stock has been growing strongly during the last decade. It is also in strong correspondence with Nøttestad and Utne (2016) who state that “During the same period (2007-2016) the distribution of mackerel has increased from 1 million km² to over 3 million km², in addition the density has doubled from 1,5 tonnes / km² to > 3 tonnes / km²” (Translated from Norwegian by JC Holst). This consequently indicates a six fold increase in the mackerel stock during the ten years period leading up to 2016 according to Nøttestad and Utne.

Even if the biomass series are used as indexes in the assessment model, their absolute values are important for the resulting stock assessment as for instance demonstrated in the leave one out run graphs. It is therefore important to try to minimize bias in the biomass index series through improved knowledge and understanding of various sources of sampling bias.

The egg spawning stock estimation method was first described by Lockwood, Nichols and Dawson (1981). In short, the method estimates the spawning stock of mackerel by first estimating the total yearly production of Stage I mackerel eggs. Then this number is divided by the average realized fecundity, giving an estimate of the number of spawning females. Then the estimated number of spawning females is multiplied with the estimated mean weight of all spawning females and finally this biomass is multiplied by two to account for the males.

Consequently, the method is very much indirect and its accuracy depends on a series of parameters being estimated unbiased in order to get an unbiased estimate of the true size of the spawning stock.

In this working paper I will in particular look at a possible serious bias in the estimated number of eggs spawned by using the observed average density of Stage I eggs for estimating the spawning stock.

Do the estimated density of stage I eggs give an unbiased estimate of the total number of eggs spawned?

Lockwood's method assumes that the Gulf net hauls gives an unbiased estimate of the total egg production in the spawning stock using the observed density of Stage I eggs. But will the estimated density of Stage I eggs obtained with Gulf nets with a geographic spacing of 30*30 minutes give an unbiased estimate of the total number of mackerel eggs spawned ?

In ICES WGMEGS 2008 the exponential decay model was used as an alternative method to estimate the production of horse mackerel eggs. The idea behind using this method is that when an exponential decay model is fitted to the observed density of newly spawned horse mackerel og mackerel eggs, the estimated y-axis intercept will be an estimator of the initial density of eggs, i.e. the average density of eggs over the entire spawning area at the very moment of spawning.

It is important to note that the example from WGWECS 2008 is about horse mackerel eggs. Afterwards, we will look at the relevance of this to the estimation of the mackerel spawning stock.

From ICES WGMEGS 2008 (page 73):

"7.4 Egg production estimate for southern horse mackerel in 2007"

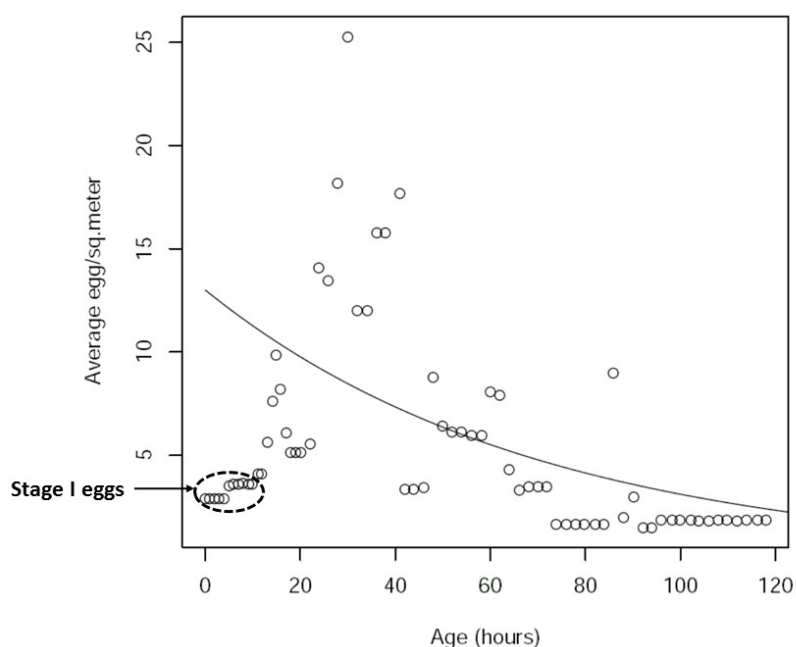
To obtain the egg production (P) for the total stock area, the exponential decay model:

$$N_t = P_0 \cdot \exp(-Z \cdot t)$$

where N_t is the mean number at age t of eggs per square meter and P_0 is the production of eggs per square meter, was fitted by non-linear regression to the mean number of eggs/sq.meter per

age class (Figure 7.4.1). Mean number at age was used instead of the raw data in order to smooth the highly variable data, and to avoid that outliers could have a high leverage on the fitting of the model. **Also, the eggs younger than 12 hours of age were not included in the model fitting. Due to their high aggregation during about the first 12 hours after spawning, the eggs of these ages are rarely caught (Figure 7.4.1).** Therefore, their apparently low abundance “pulls” down the initial part of the mortality curve, which results in artificially low mortality (Z) and production (P0) estimates.”

2007



“Figure 7.4.1. (WGMEGS 2008). Mortality curve fitted to the mean number of eggs/sq. meter at age. Eggs aged less than 12 hours were not used for fitting the model.”

*The original figure has been modified by encircling the approximate age range of Stage I eggs. Stage I eggs will be from 0 hours and up to between 4 to 12 hours in horse mackerel depending on the temperature in the drift route of the eggs.

Concerning possible bias in the estimated number of spawned eggs it is alarming to note that the authors excluded eggs younger than the age of 12 hours in their analysis. Stage I eggs are from age 0 hours and up to from 4 to 12 hours old in horse mackerel depending on temperature, meaning that the Stage I eggs were left out of the analysis because “**Due to their high aggregation during about the first 12 hours after spawning, the eggs of these ages are rarely caught.**” But Stage I eggs are the eggs used in Lockwoods method to estimate the spawning stock both in horse mackerel and mackerel!

Looking at the observed distribution of egg density by age it is obvious that the highest density is reached between approximately 25 to 40 hours (Figure 7.4.1 above). At this stage the eggs

have obviously become more available for being caught than just after spawning. Most probably this is due to a more even spread of the eggs over the survey area at this higher age and the Gulf nets starts catching them more efficient than before they were spread out from probably concentrated spawning hotspots. The large distance between the sampling positions and the very small water mass sampled in a Gulf net haul combined with spawning in concentrated hotspots seems a probable explanation to this phenomenon.

Can the available egg data at all be used to estimate the initial egg density?

In figure 2, two additional exponential decay lines have been hand drawn into the plot in *figure 7.4.1.* above. Given the observed distribution of average egg density by age it does not seem possible to estimate the initial egg density, we simply do not have the necessary data because the stage I eggs are more or less no present in the samples. Which of the three lines estimates the Y-intercept best given the available data? Maybe the upper blue line, intercept C = 37 eggs/sq meter is the most realistic one as it fits best with the highest densities observed between 25 to 40 hours? Or maybe the lower, intercept A = 13 eggs/sq meter, is the most realistic and the one fitted by WGMEGS 2008. Or B at 25 eggs/sq meter?

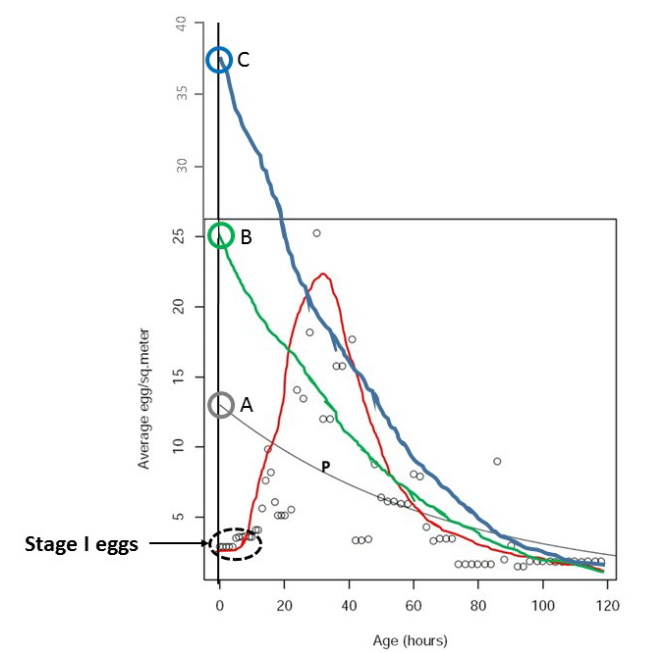


Figure 2. Figure 7.4.1 WGMEGS 2008 modified. Red line is approximate distribution of average egg density by age. A, B and C are the Y-axis intercepts for three exponential decay models, grey, green and blue.

Comparing the various estimates of the initial egg density at age 0 hours it is obvious that the method suggested by Lockwood et al (1981) using the observed density of Stage I eggs gives an

unrealistic and strongly negatively biased estimate of the initial egg density and thus the spawning stock. The three estimates derived from the exponential decay model in figure 2 gives estimates of the egg production from 4,3 to 12,5 times higher than that indicated by the estimated density of Stage I eggs (Table 1).

Table 1. Estimates of initial egg density derived from the measured density of Stage I eggs and from the three different exponential decay models shown in figure 2.

Estimator	Estimated initial egg density	Ratio of Stage I egg density
Observed density Stage I eggs	3	1,0
Exponential decay model Y intercept A	13	4,3
Exponential decay model Y intercept B	25	8,3
Exponential decay model Y intercept C	37,5	12,5

Discussion

The data and analysis presented above concerns horse mackerel eggs. However, the same source of bias will apply to the estimation of the total amount of mackerel eggs spawned and thus the estimated mackerel spawning stock biomass index.

In WGMEGS 2008, page 91, this source of bias was discussed:

“For the estimation of daily egg production for both mackerel and horse mackerel, only the counts of stage I eggs are used. This is recognised as a conservative estimate of the total eggs spawned because of mortality which occurs during development. However until there is consistency in the identification of the other stages, between all countries, the other stages cannot be used for the estimation of mortality rates and backtracking to total eggs spawned.”

The working group consequently stated that the estimated number of eggs spawned from the Lockwood method is a conservative estimate because of the mortality which occurs during development, ie between the moment the eggs are spawned and the time of their mean age when sampled.

The 2008 WGMEGS furthermore stated that it will it will not be possible to estimate mortality rates and backtrack to the total eggs spawned until there is consistency in the identification of the other stages (than stage I).

As seen from table 1 the ratio of the estimated density of stage I horse mackerel eggs measured through the Gulf sampling and the estimated backtracked number of total eggs spawned using the exponential decay model will vary greatly according to what model is used. In the example above the ratio between the estimated number of eggs spawned is from 4,3 up to 12,5 times larger when using the exponential decay model for backtracking the actual density of eggs at the moment of spawning. It is uncertain how this will look in mackerel eggs due to the fewer stages in mackerel eggs. However, it will be important to further study and describe this source of potential bias in the mackerel stock assessment given the high weight the egg biomass index has to the spawning stock estimate due to the low standard deviation of the egg biomass index.

The large contrast indicated in stock size development between the egg biomass index, and the tag and survey indexes after about 2007 suggest that there can be some sort of systematic negative bias in the egg biomass index. The large increase in spawning period and spawning area observed in the mackerel stock after about 2006-7 can explain a significant, but unknown part, of the lack of increase in the egg spawning biomass index since around 2007. The bias described above may very well represent another significant part in explaining the lack of growth in the egg index after 2007, in particular if this phenomena by some reason is density dependent. It is therefore important for the ICES WGMEGS to thoroughly analyse this potentially large and negative source of bias in the mackerel egg biomass index to evaluate its potential in contributing to underestimation of the mackerel spawning stock. It will also be important to carry out investigations which can further clarify the underlying factors causing this bias. It is strongly suggested that such investigations are carried out during the WGMEGS 2019 egg survey.

References

ICES. 2008. Report of Working Group of Mackerel and Horse mackerel Egg surveys. ICES CM 2008/LRC:09.

Lockwood, S, Nichols, J., and Dawson, W. 1981. The estimation of a Mackerel (*Scomber scombrus* L.) spawning stock size by plankton survey. J. Plankton Research, No 3 and 2.

Nøttestad, L og Utne, K. R. 2016. Makrellens vandringer – historisk ekspansjon i Norskehavet de siste 10 år. NATUREN nr. 6. DOI: 10.18261/issn.1504-3118-2016-06-05

On possible bias in the IESSNS trawl biomass index on NE Atlantic mackerel.

Jens Christian Holst

Ecosystembased

Introduction

The egg-, survey- and tagging (equal tagging mortalities) biomass indexes used in the NE Atlantic mackerel assessment indicate very different stock trends. While the egg index shows a fairly stable and low stock since its start in 1992, the survey index and tag index, given equal tagging survival for steel and RFID tags, indicate a strong growth in the stock from around 2006-2007 and onwards.

A tagging experiment carried out by the IMR west of the British Isles in 2018 indicate no significant difference in tagging mortality between the old and the new tagging methods for mackerel, thus strongly indicating that equal tagging mortality between the steel and RFID tags is a valid assumption.

In this document, I present various sources of possible sources of bias in the trawl biomass index and argue for carrying out an experiment where purse seine sets are done in parallel with the trawl hauls to study the average density of mackerel coming out of the two methods.

IESSNS trawl biomass index

There are several possible sources of bias in the IESSNS trawl biomass index.

Variation in the mean depth distribution of mackerel between years

One important assumption for the trawl method to give an unbiased estimate of the true density of mackerel in a trawled area is that the depth distribution of the mackerel is constant between years, implying the trawl catches the same proportion of the mackerel in

the area every year. Based on observations from skippers on mackerel fishing vessels and the IESSNS survey vessels, there are strong indications that the depth distribution of the mackerel vary between years.

An example: For several years from about 1998 the IMR carried out trawl surveys for postsmolts of European salmon in the surface in late June-early July in the Norwegian Sea. After sea arrival, the postsmolts swim northwards from their respective rivers and the first postsmolts reach the international zone in the Norwegian Sea around June 15th. At about the same time also the mackerel arrived in the international zone during that period (Now the mackerel arrives much earlier).

On arrival at the southern border between the Norwegian and international zone around 20th June there would always be around 10-15 Russian pelagic trawlers fishing for mackerel along the border on the international side. Normally we would always see the buoys on the trawl wings at the surface behind the vessels. Then one year, probably in 2002, I could not see the buoys. I contacted the skipper of the Russian vessel by VHF and asked him why. He answered "The fishing is very difficult this year as the mackerel is swimming deeper than normal and we have to fish at 40 to 60 meters. This gives us much smaller catches".

In 2018 the mackerel abundance index of IESSNS was down 40% from 2017. At the same time the temperature in the upper 20 meters in the Norwegian Sea was down 0.5 -2.0 °C. What effect this sudden drop in surface temperature had on the depth distribution of the mackerel, or on the vertical distribution of the food of the mackerel, and consequently on the depth distribution of the mackerel, is not known. It may very well be that the total amount of mackerel in the survey area was the same in 2017 and 2018, but that a larger proportion of the mackerel was distributed at a depth below the sampling depth of the trawl in 2018 than in 2017.

The Norwegian mackerel fishermen made the general observation that the mackerel was deeper distributed in 2018 and thus more difficult to catch by trolling and with purse seines in 2018 as compared to earlier years. This applied both to the oceanic and coastal fisheries.

Bias due to mackerel situated below the efficient fishing depth of the Mulpelt 832

All of the five skippers I have spoken with who has participated in the IESSNS survey express strong concerns about the large amount of mackerel they observe on their echo sounders and sonars swimming under the efficient catch depth of the Mulpelt 832 trawl. Some of the skippers are of the opinion that the trawl only catches a small fraction of the mackerel which were in the total water column under the efficient catching depth of the trawl. As indicated above, the fraction of mackerel situated below the efficient catching depth of the trawl is most probably varying between years.

Bias due to mackerel swimming along with the trawl

The trawl will herd mackerel in front of it and the mackerel will swim along with the trawl for a period during the catch phase. This means the trawl will not catch all the fish in the trawled volume. Consequently, the catch should be multiplied with a number larger than one to compensate for the mackerel not caught due to it swimming along with the trawl. This is presently not done.

Vertical scaring effect of the vessel

During the IESSNS survey the propeller of the vessel towing the trawl will typically be at a depth down to about 5-6 meter and the trawl will typically catch down to 25-30 meters. Vessel generated sound is described to affect fish behavior down to about 200 meters. Trawling at 5 knots and with a power around 6000 KW produces a very strong sound and vibration in the water mass down to 30 meters. This vessel generated vertical scaring effect, thus bias, exerted on the mackerel in the IESSNS survey is unknown but must be anticipated to be large because of the shallow distribution of the mackerel fished for.

A suggested experiment to get a less biased estimate of the true density of mackerel in the IESSNS area.

As described above there are several potential sources of bias in the trawl method applied in the IESSNS survey. It will therefore be important to carry out studies and experiments which can be used to evaluate the size of the various sources of bias connected with trawl hauls.

Purse seine is a much less selective sampling tool than a trawl for several reasons. A large purse seine at 850 meter length and 260 meters depth will catch efficiently down to about 180 meters. Setting a purse seine this size with a small mesh size and using the right setting and hauling technique will give a much less biased estimates of the concentration of mackerel under a given area of the ocean as compared to a trawl haul. Amongst other a purse seine would catch over the entire depth inhabited by mackerel not only the upper 25-30 meters. When setting the purse sein, the vertical scaring effect of the vessel will cut through the mackerel concentrations in one part outside the net and one part inside the net, like slicing a bread, instead of scaring the fish away from the fished volume like in trawling.

By having one or two purse seiners following the survey vessels and setting their seines at the same time and in the neighborhood of where the trawl haul is carried out throughout the survey, one would get an additional and less biased data set estimating the average density of mackerel over the entire survey area. Studying the relationship between the densities of mackerel estimated from the purse seine sets with that of the corresponding densities estimated from the trawl hauls will make it possible to get a better grip of how well the trawl hauls estimate the 'real' density of mackerel in the trawled volume.

These investigations should be combined with various studies involving echo sounders and sonars to study the mackerel's reactions both to the purse seine and the trawl catch

operation in order to evaluate how well the purse seine catches represent the true mackerel concentrations in an area as compared to the trawl hauls.

Based on the results from such a study it should be evaluated if purse seine will be a better method to obtain less biased estimates of mackerel density during the IESSNS survey for the future.

Because of the very direct way of estimating mackerel density using purse seines, it should also be evaluated if this method alone would be sufficient for obtaining spawning stock biomass indexes with an acceptable bias for the assessment of the mackerel stock in the future.

Analyses of the process error in the mackerel assessment

Thomas Brunel (Wageningen Marine Research)

Working document to the Interbenchmark Workshop on the assessment of northeast Atlantic mackerel

Introduction

Since the adoption of SAM as the method to assess the NEA mackerel stock, concern has been voiced about the process error (PE) in this assessment. The main point of concern was the existence of a structure (correlations in time and between ages) in this PE which is supposed to be independent and identically distributed in this assessment.

An illustration of this structure has been proposed in the form of a biomass accumulated PE (multiplying the deviations in abundance at age by corresponding weight at age and summing across ages for a given year). This biomass accumulated PE from the 2014 assessment (figure 1) showed an alternation of period of years where the PE was consistently positive, and periods of years where it was negative. In the latest years of this assessment, the magnitude of the positive deviations was larger, almost to the magnitude of the catches in weight. This has been interpreted as a sign that the model was artificially generating biomass in order to be able to fit the different observations.

In the 2018 assessment, the magnitude of the PE is about the same size as in the 2014 assessment (variance of 0.17, compared to 0.19 in 2014) but the autocorrelation pattern seems to have decreased in the recent years (figure 2).

In an attempt to understand the implication of this PE on the estimated stock trajectories, and to investigate whether the observed autocorrelation is due to any particular data source, two types of analyses were carried out :

- Run the assessment without PE. This was done by imposing a fixed value close to zero to the parameter corresponding to the magnitude of the PE ($\log(-5)$)
- Inspect the realisation of the PE for the leave of out runs of the 2018 assessment.

Results

- Model without PE

The assessment run without PE has a very similar SSB trend as the current assessment (figure 3). SSB is slightly higher in the recent years (since 2009) but the difference is small. The fishing mortality of the assessment without PE is slightly more variable than in the current assessment, and is lower in the recent years.

Model parameters (figure 4) show that the model without PE has slightly higher observation variances (and tag overdispersion), especially for the fit to catch-at-age matrix. Fishing mortality random walk deviations are larger (in agreement with the observation of a more variable F_{bar}) for the model without PE. Also in agreement with the high SSB since 2009, the model without PE has lower catchabilities for the IESSNS survey.

- PE in the leave one out runs

The estimated PE standard deviation is almost identical for all leave one out runs and for the current assessment (table 1). The PE in biomass is very similar al runs except for the one leaving out the RFID tags for which there is a persistence of positive deviations between 2008 and 2014, while for the other runs, deviations are distrubuted randomly after 2010.

Table 1 : estimated PE standard deviation for all leave one out runs and for the current assessment (with confidence intervals).

	value	CI lower	CI upper
WG2018	0.166	0.128	0.214
No egg	0.163	0.126	0.211
No Rind	0.170	0.132	0.219
IESSNS.5	0.170	0.134	0.216
RFID.13	0.160	0.123	0.207

Conclusions

The strong autocorrelations observed in the PE of the mackerel assessment in 2014 seem to have weakened in the 2018 assessment, with more random values observed in the recent years. This improvement was mainly achieved in the 2017 benchmark, when the RFID tags were included (as illustrated here by the run excluding the RFID tags).

Having a process error in the mackerel assessment only has a small influence on the estimated SSB trajectory, but slightly more influence on the fishing mortality. The process error contributes to achieve a better fit to the observations, especially for the catches-at-age. This feature has been observed for different assessments (figure 5, Brunel, personal communication): when running assessment for 5 stocks with different imposed values for the PE variance, all assessments showed a lower estimated observation variance for the catches when higher values of the PE were imposed. Other features from this study were that the fit to the surveys was not strongly modified when different values of the PE variance were imposed. Also, larger values of the PE variance was associated with wider confidence intervals on SSB and F_{bar} . Finally, autocorrelations in the PE was not associated with its magnitude.

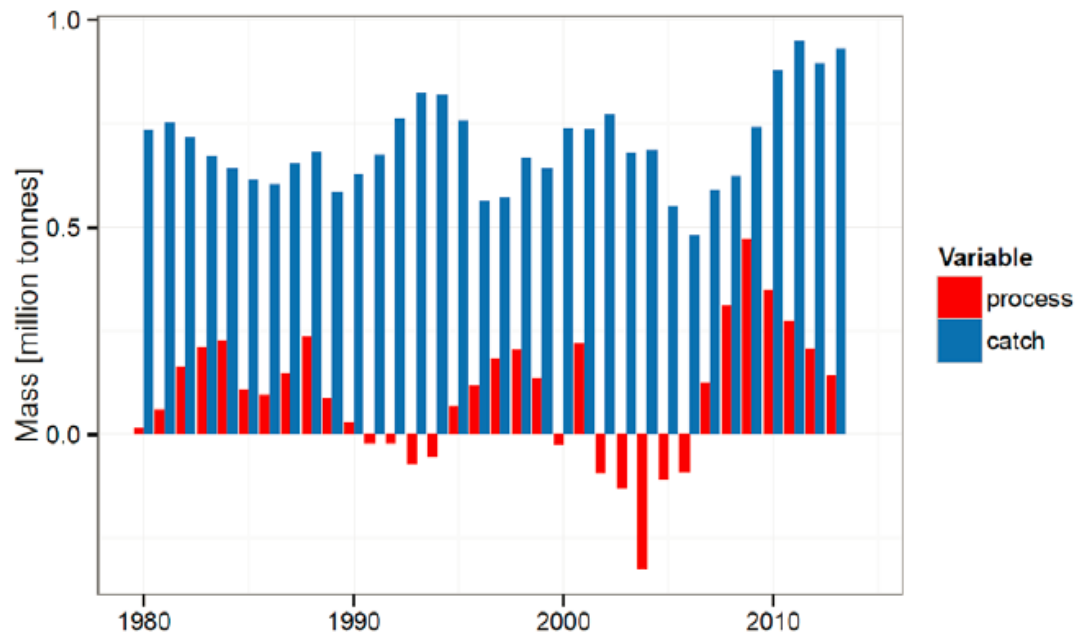


Figure 1 : biomass accumulated process error (annual deviations in abundance at age multiplied by weight at age and summed up across ages) compared to catches from the 2014 assessment (from ICES WKMACLTMP 2014).

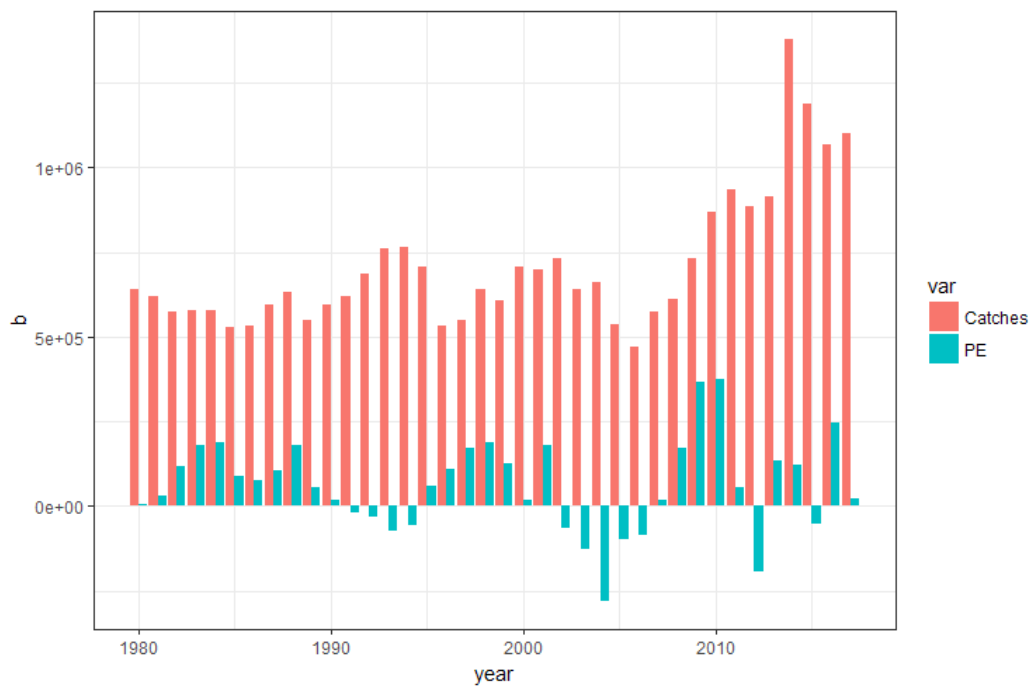


Figure 2 : biomass accumulated process error (annual deviations in abundance at age multiplied by weight at age and summed up across ages) compared to catches from the 2018 assessment.

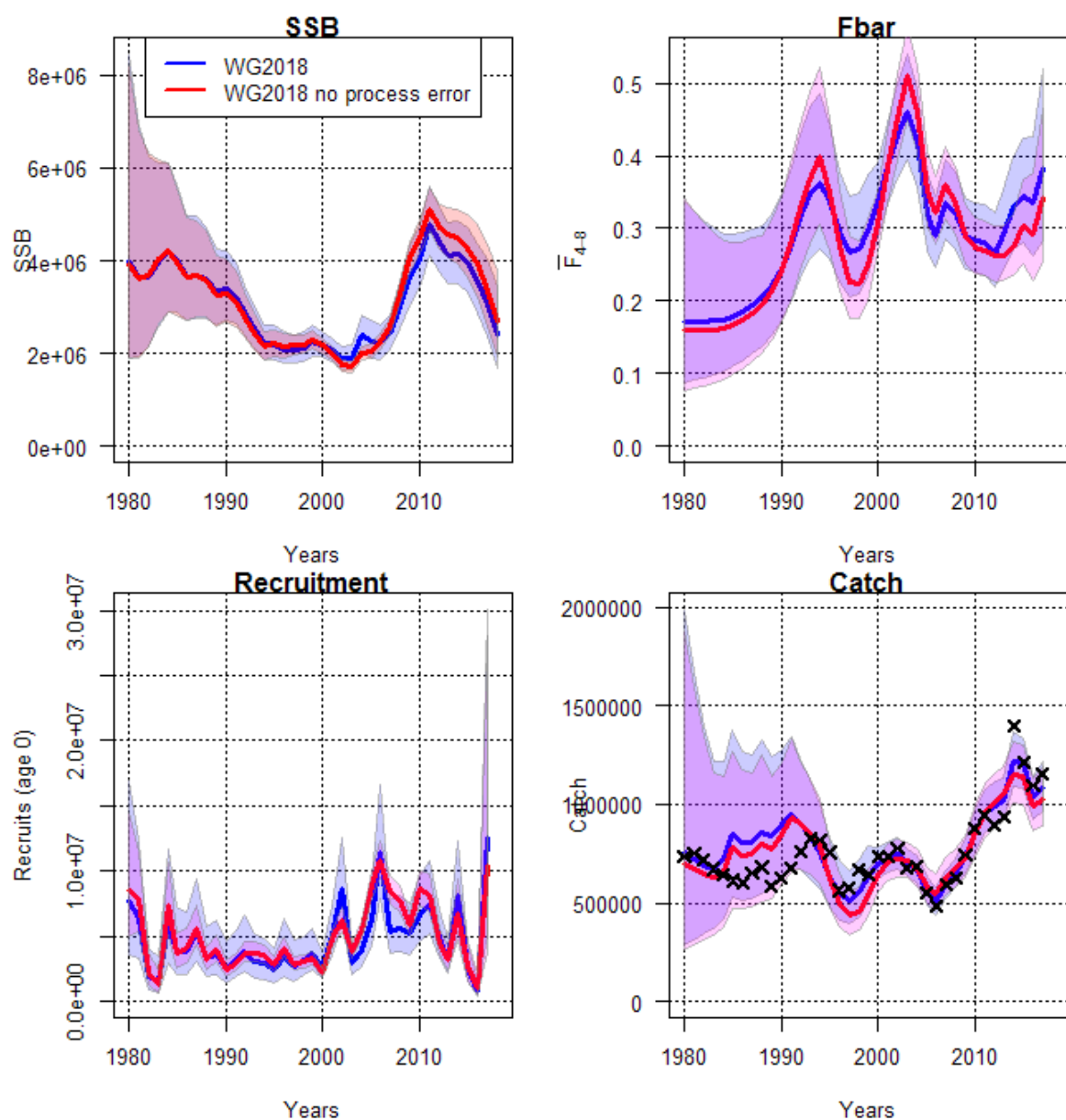


Figure 3 : comparison of the stock trajectories for the current mackerel assessment and the same assessment run fixing the process error variance to (almost) 0 ($\log(-5)$).

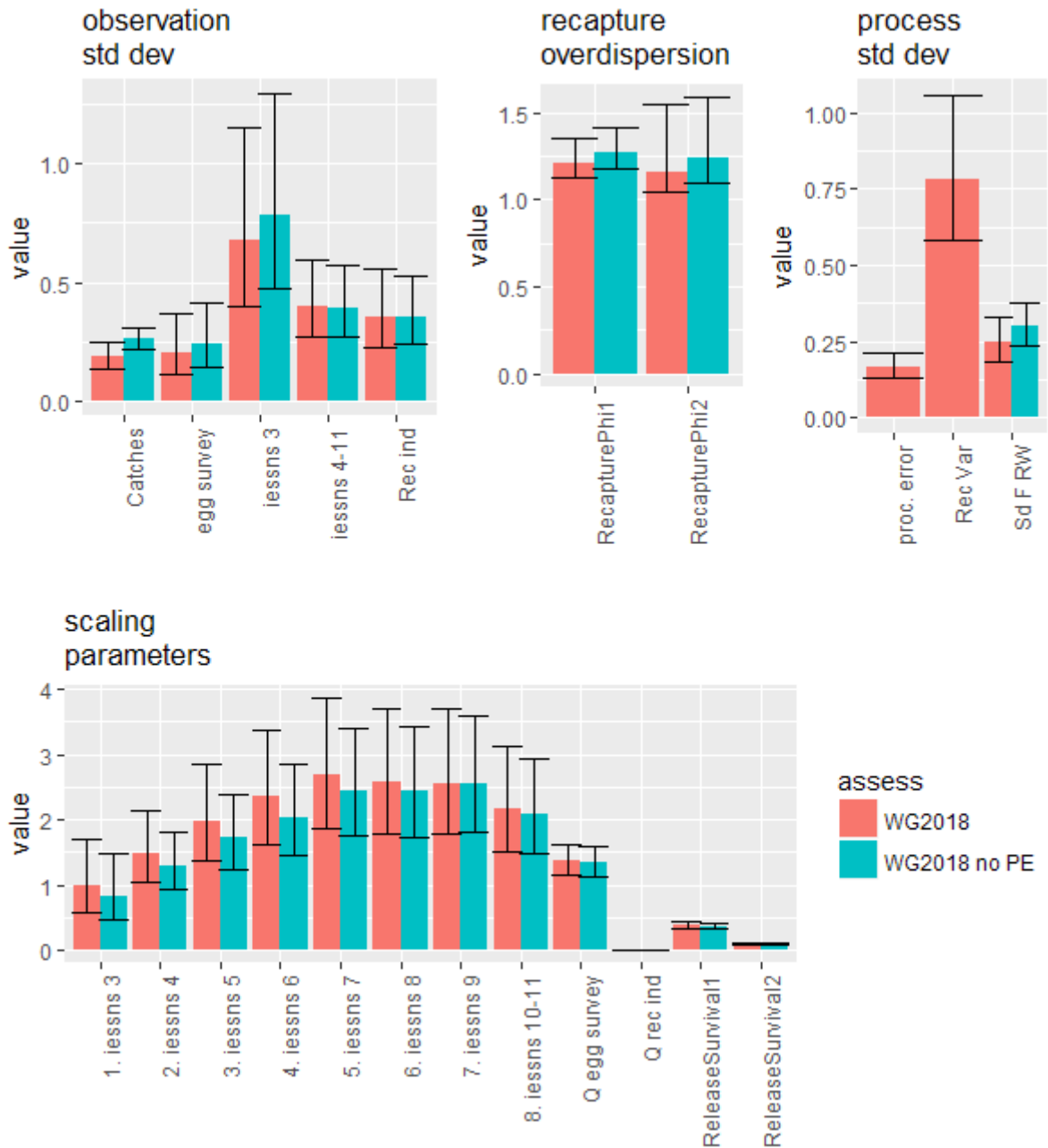


Figure 4 : comparison of the model parameters for the current mackerel assessment and the same assessment run fixing the process error variance to (almost) 0 (log(-5))

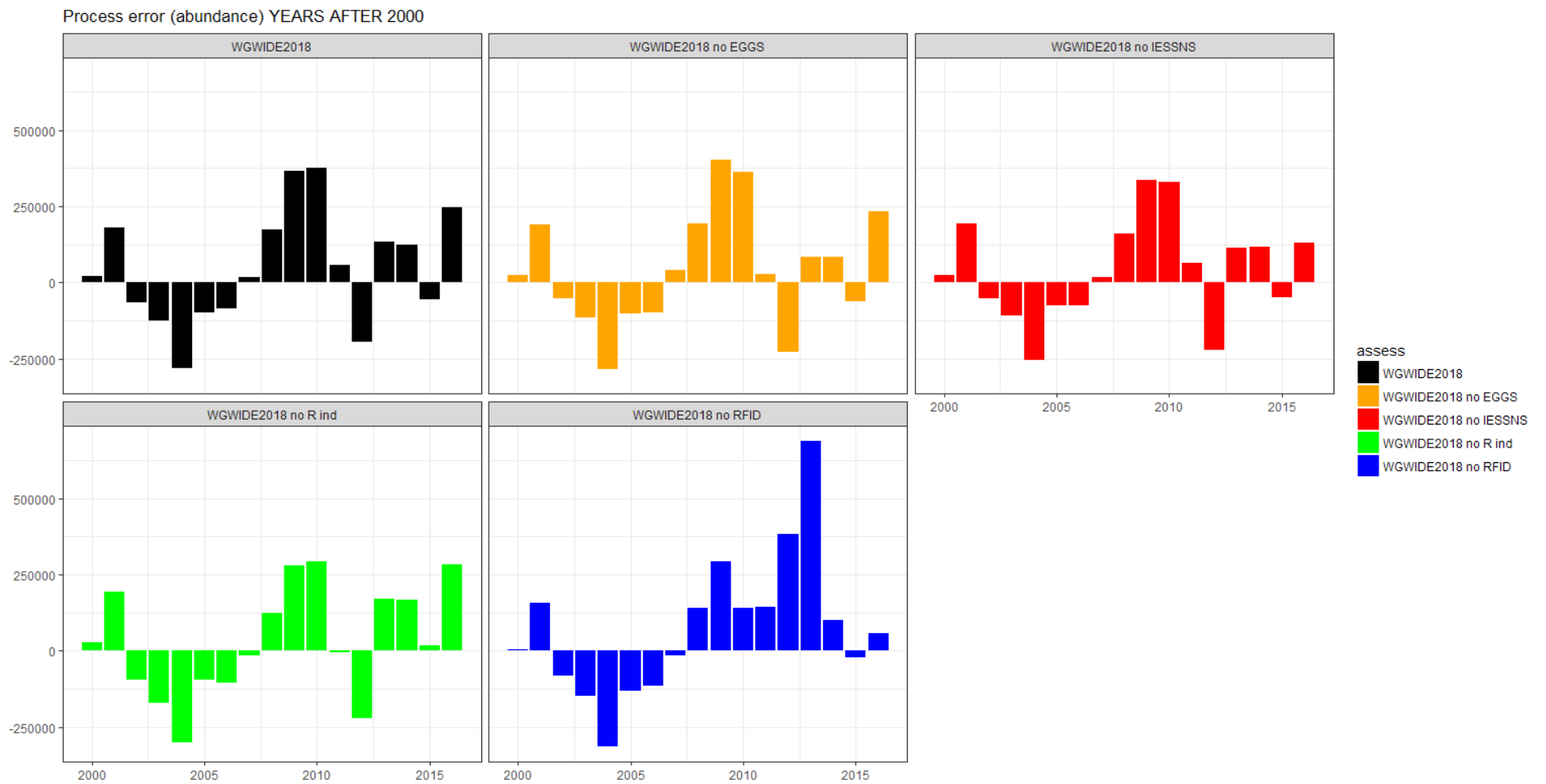


Figure 5 : comparison of process error realisation (biomass accumulated) for the current assessment and each of the leave one out runs.

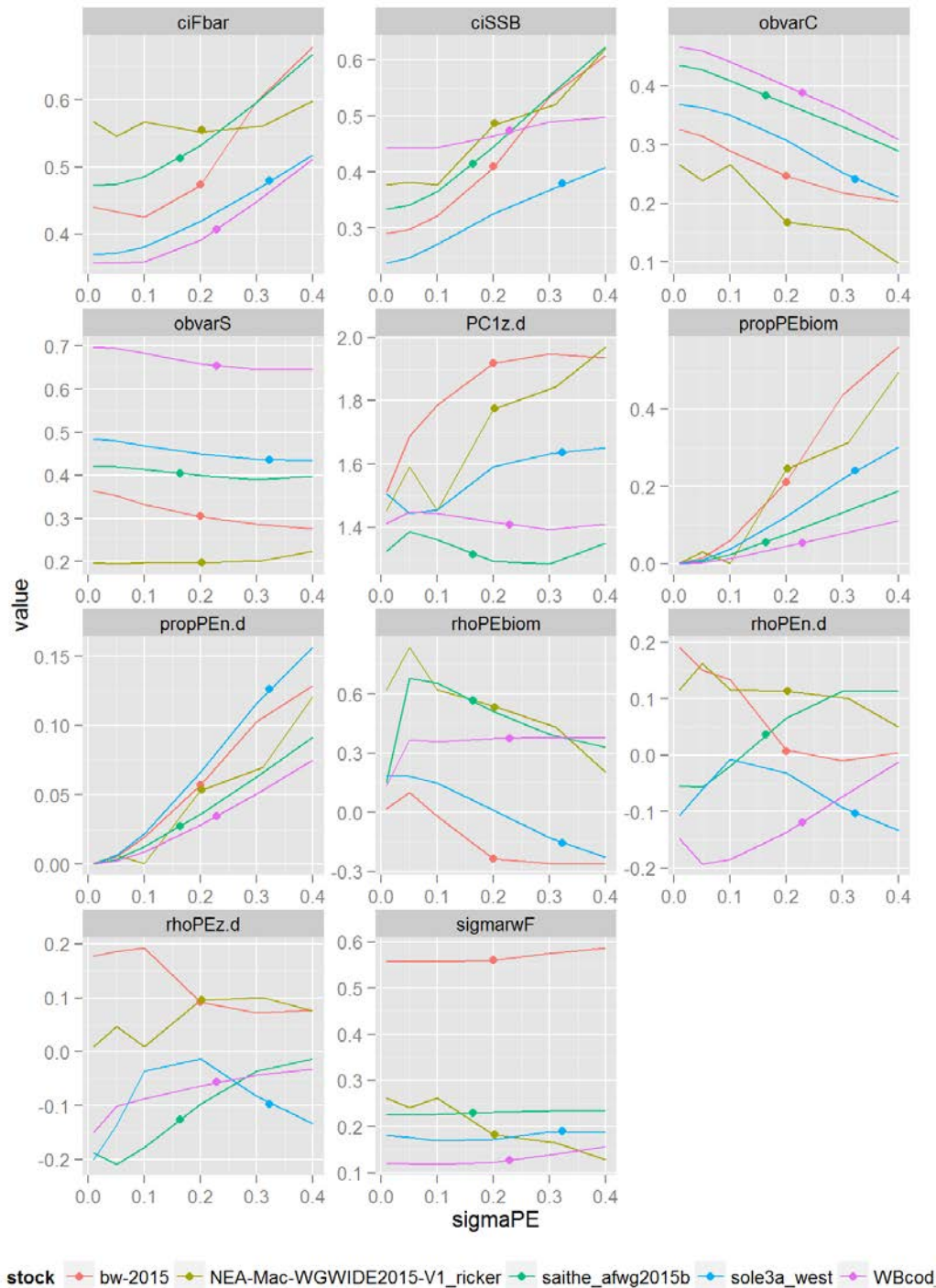


Figure 6 : example of 5 stock assessment run with fixed values of the PE variance. The assessments were run by fixing values of the PE variance in a range of values between 0 and 0.4 (actual parameter estimate in 2015 shown by the dot).

Different panels show different assessment characteristics :

- Magnitude of the PE:
 - sigmaPE : standard deviation of the deviations from the survival equation. This is a parameter estimated by SAM

- propPE_{n.d} : PE expressed as deviations in numbers at age, ${}_ND_{a,y}$, divided by the corresponding abundance at age, $N_{a,y}$, averaged over years and ages
- propPE_{biom} : biomass accumulated PE, ${}_BD_y$, divided by the corresponding catch biomass

- o correlations in the PE

- PC1z.d : degree of inter age correlation in the PE

Calculated as the percentage variance carried by the first component of a PCA performed on the process error expressed as mortality deviations, with ages used as variables and years used as individuals. The idea is that, if there are correlations among age-groups, these will result in a large percentage variance explained by the first component in the PCA.

- rhoPE_{z.d} and rhoPE_{n.d} : temporal autocorrelation of the process error

calculated as the mean across ages of the age specific autocorrelation (1 year lag) coefficient for the PE expressed as deviations in mortality and abundance at age, respectively.

- rhoPE_{biom} : temporal autocorrelation of the biomass converted process error

- o other descriptors of the SAM assessments

- sigmarwF : variance of the random walk of the fishing mortality calculated as the mean across ages of the variance parameter estimated by SAM
- obvarC : observation variance for the catches calculated as the mean across ages of the variance parameter estimated by SAM
- obvarS : observation variances for the surveys calculated as the mean across ages of the observation variance parameters estimated by SAM
- ciSSB and ciFbar : precision of the assessment with regard to the SSB and Fbar, calculated as the average over time of the width of the confidence interval, relative to the point estimate of SSB and Fbar

Investigating different model configurations to estimate post release survival in the mackerel assessment

Thomas Brunel (Wageningen Marine Research)

Working document to the Interbenchmark Workshop on the assessment of northeast Atlantic mackerel

Introduction

The current mackerel assessment uses 2 types of tagging data : the steel tags recaptured between 1986 and 2006 (from tagging experiments 1980 to 2004) and the RFID tags recaptured between 2012 and 2017 (tagged between 2011 and 2016). The Norwegian tagging program, however, continued from 2006 to 2009 using the steel tags and data from recapture years 2007 to 2010 are available.

At the previous benchmark, it was decided not to use this data because the tagging protocol had changed in 2006, getting towards an automation of the process, which might have influenced the survival rate of the fish at release (e.g. automatic jigging might be more harmful). For this reason, it was considered that the most recent steel tags could not be treated as the historical data, and they were not included in the assessment.

The current assessment indicates that post release survival of the new RFID tags is substantially lower than for the old steel tags (around 10% and 40% respectively). This difference could be explained by the difference in tagging practices, the new automated tagging being potentially more harmful for the fish. However, scientists involved in the tagging program doubt that the changes in practices could have reduce survival so much (based on observed vitality of the fish). The estimated parameter “survival” encompasses more than simply the survival rate of tagged fish after release. Factors such as scanning efficiency, errors in estimates of fish numbers scanned, loss of tags also contribute to the estimated value of the survival parameter.

In an attempt to understand why there is such a large difference in survival rate estimated by SAM for the steel tags and the RFID tags, the assessment was run with the most recent steel

tags (released after the change of tagging protocol) as a 3rd data series, with its own estimated survival rate.

Methods

Two new assessments were run and compared to the current assessment :

- Assessment using the steel tags from tagging experiments conducted from 2006 to 2009 (and recaptured between 2007 and 2010). This assessment estimates 3 set of parameters (survival rate and overdispersion) corresponding to 3 sets of tagging data :
 - o Steel tags from experiments carried out between 1980 and 2004
 - o Steel tags from experiments carried out between 2006 and 2009
 - o RFID tags from experiments carried out between 2011 and 2016

The first and third data sets correspond to the data currently used in the assessment

- Assessment also using all available tagging data, but with a further sub-setting of the steel tags data :
 - o Steel tags from experiments carried out between 1980 and 1995
 - o Steel tags from experiments carried out between 1996 and 2000
 - o Steel tags from experiments carried out between 2001 and 2004
 - o Steel tags from experiments carried out between 2006 and 2009
 - o RFID tags from experiments carried out between 2011 and 2016

The aim of this run is to see if there are some temporal variations in the estimated survival rate when we estimate it by groups of years.

Results

- Incorporation of steel tags from experiments carried out between 2006 and 2009

Incorporating the additional steel tag data had no visible influence on the assessment output (figure 1) with nearly identical SSB, F_{bar} , recruitment and estimated catches. Most model parameters have very close values (figure 2). The overdispersion parameter for the additional steel tags is higher (and more uncertain, steel 2 on figure 2) than for the older steel tags. This means that these additional tagging data have a low weight on the assessment which explains that no difference is found in estimated stock trajectories. The estimated survival rate for the steel tags released before 2006 and for the RFID are fairly similar between the two runs (table 1). The survival rate estimated for the latest steel tags is very close to the survival rate estimated for the RFID tags, around 10%.

Table 1 : estimated survival rates for the current assessment and the assessment including the latest steel tags

	Steel 1980-2004	Steel 2006-2009	RFID 2011-2016
Current assessment	0.385		0.099
Including latest steel tags	0.373	0.081	0.100

- Estimating survival rate by groups of years

The stock trajectories for the years for which catch data is used (after 2000, figure 3) were not substantially modified when the assessment was conducted with 4 groups of tagging data for the steel tags (release years 1980-1995, 1996-2000, 2001-2004 and 2006-2009). For the earlier years for which the assessment only has tagging data and egg survey indices, the assessment with survival estimated by groups of years gives a slightly higher SSB and estimated catches than the current assessment, but similar Fbar and recruitment.

Most model parameters are, again, similar between the two model (figure 4). Small non-significant differences are observed for the observation standard deviation for the egg survey and for the standard deviation of the process error.

The overdispersion of the steel tags for the release years 1980-1995 and 1996-2000 are similar to the value observed in the current assessment (slightly lower and slightly higher respectively). The over dispersion is higher for the last two groups of steel tags (2001-2004 and 2006-2009). The overdispersion of the RFID tags is identical in the two models.

The estimated survival is higher for the first 2 groups (releases from 1980-1995 and from 1996-2000) than in the current assessment (table 2 and figure 4 and 5). Estimated survival for the 3rd group (2001-2004) is slightly lower than the current assessment. Finally, survival for the last group (2006-2009) is markedly lower, similar to the values estimated in both assessment for the RFID tags.

Table 2 : estimated survival rates for the current assessment and the assessment with 4 survival rates estimated for the steel tags

	Steel 1980-2004	Steel 1980-1995	Steel 1996-2000	Steel 2001-2004	Steel 2006-2009	RFID 2011-2016
Current assessment	0.385					0.0998
4 groups steel tags		0.438	0.477	0.322	0.081	0.100

Conclusions

These analyses suggest that survival rate has decreased markedly after 2006 : the estimated survival rate for the fish tagged with steel tags after 2006 is similar to the estimate for the RFID tags. Survival rate estimated for the release years 2001-2004, although much higher than for the later period, is significantly lower than for the earlier periods. While the difference for the last period (2006-2009) can potentially be explained by the changes in the tagging practices, it is difficult to find an explanation for the value of releases 2001-2004 being lower than earlier data. Factors others that actual survival may have changed in time, and their effect could be encompassed in the parameter estimated as survival rate. There are for instance suspicions that numbers scanned may have been underestimated until the late 1990s. If one considers all things being equal in the model, a negative bias in the number scanned would result in a positive bias in the estimated survival rate ($N_{recap} = surv \times N_{scan} \times N_{tagged} / N_{pop}$). Following this reasoning, a smaller survival rate could be expected for earlier years in the steel tag data in case the number scanned were indeed underestimated.

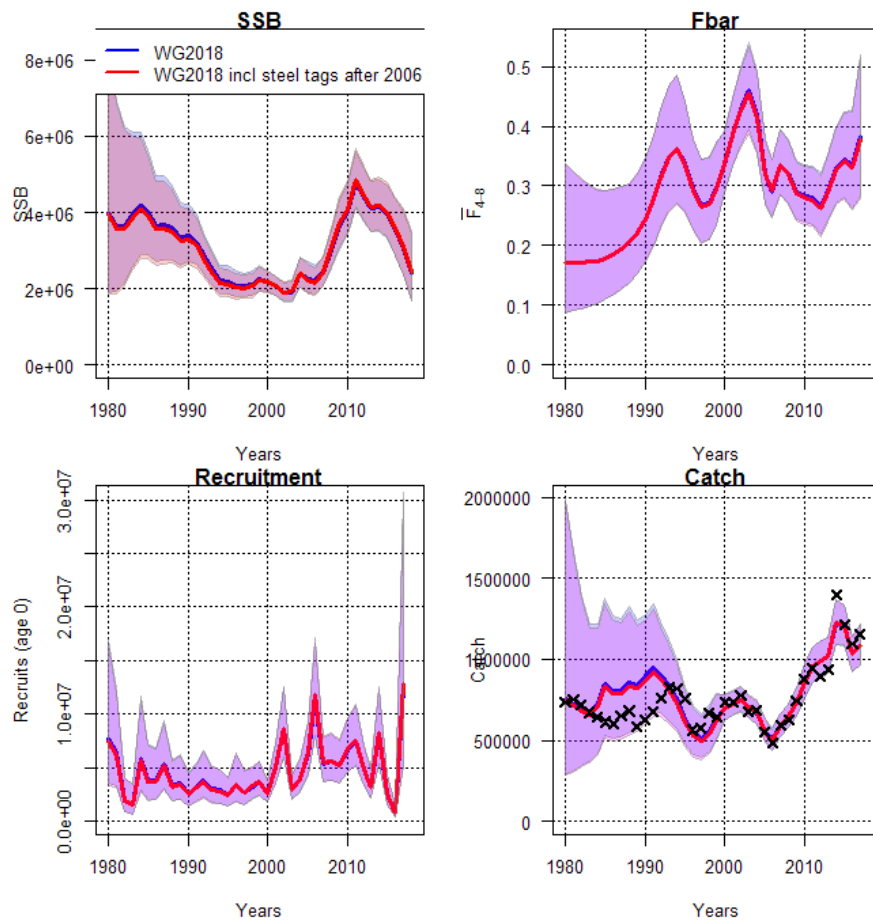


Figure 1 : comparison of stock trajectories estimates by the current assessment (WG2018) and the assessment using the additionnal steel tags (WG2018 incl steel tags after 2006).



Figure 2 : comparison of model parameter estimates between the current assessment (WG2018) and the assessment using the additional steel tags (WG2018 incl steel tags after 2006).

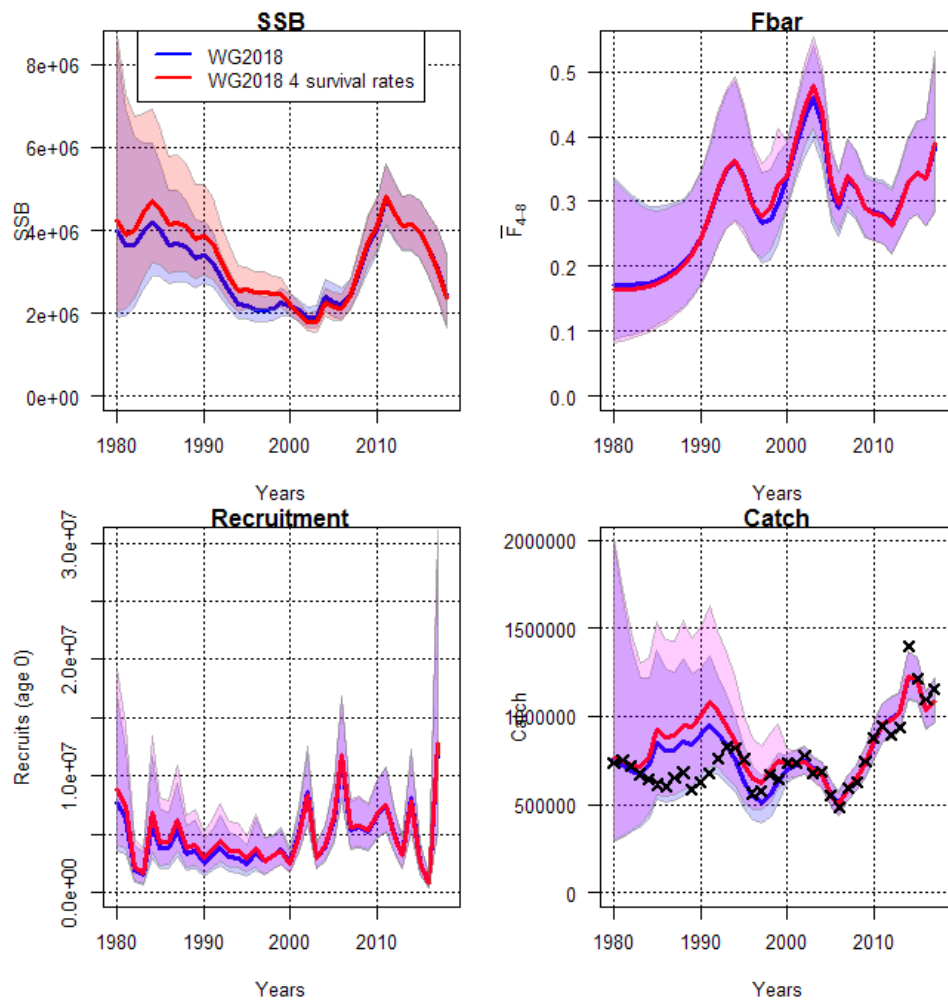


Figure 3 : comparison of stock trajectories estimated by the current assessment (WG2018) and the assessment with 4 subsets for the steel tags data (release years 1980-1995, 1996-2000, 2001-2004 and 2006-2009).

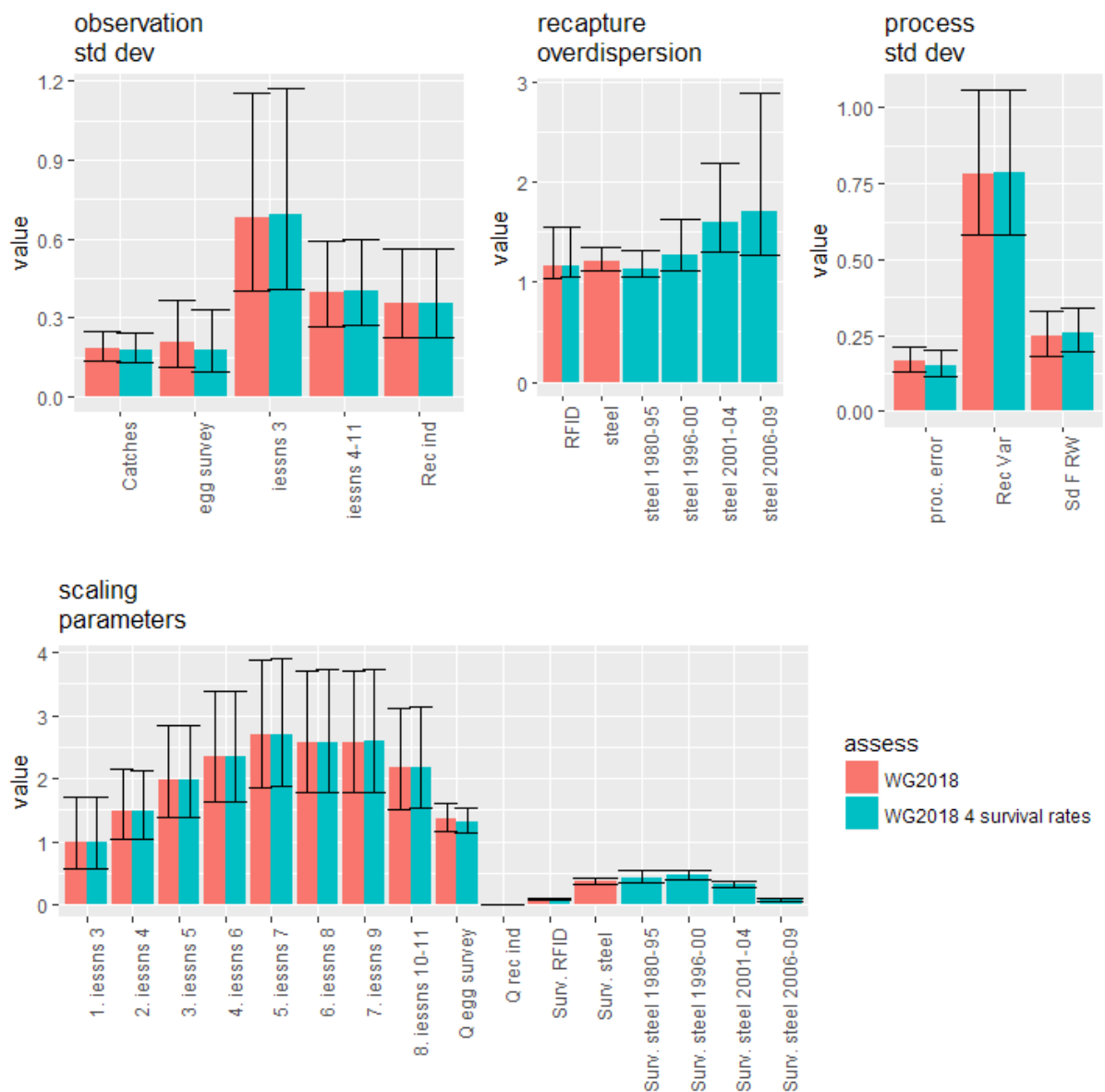


Figure 4 : comparison of model parameters estimates from the current assessment (WG2018) and the assessment with 4 subsets for the steel tags data (release years 1980-1995, 1996-2000, 2001-2004 and 2006-2009).

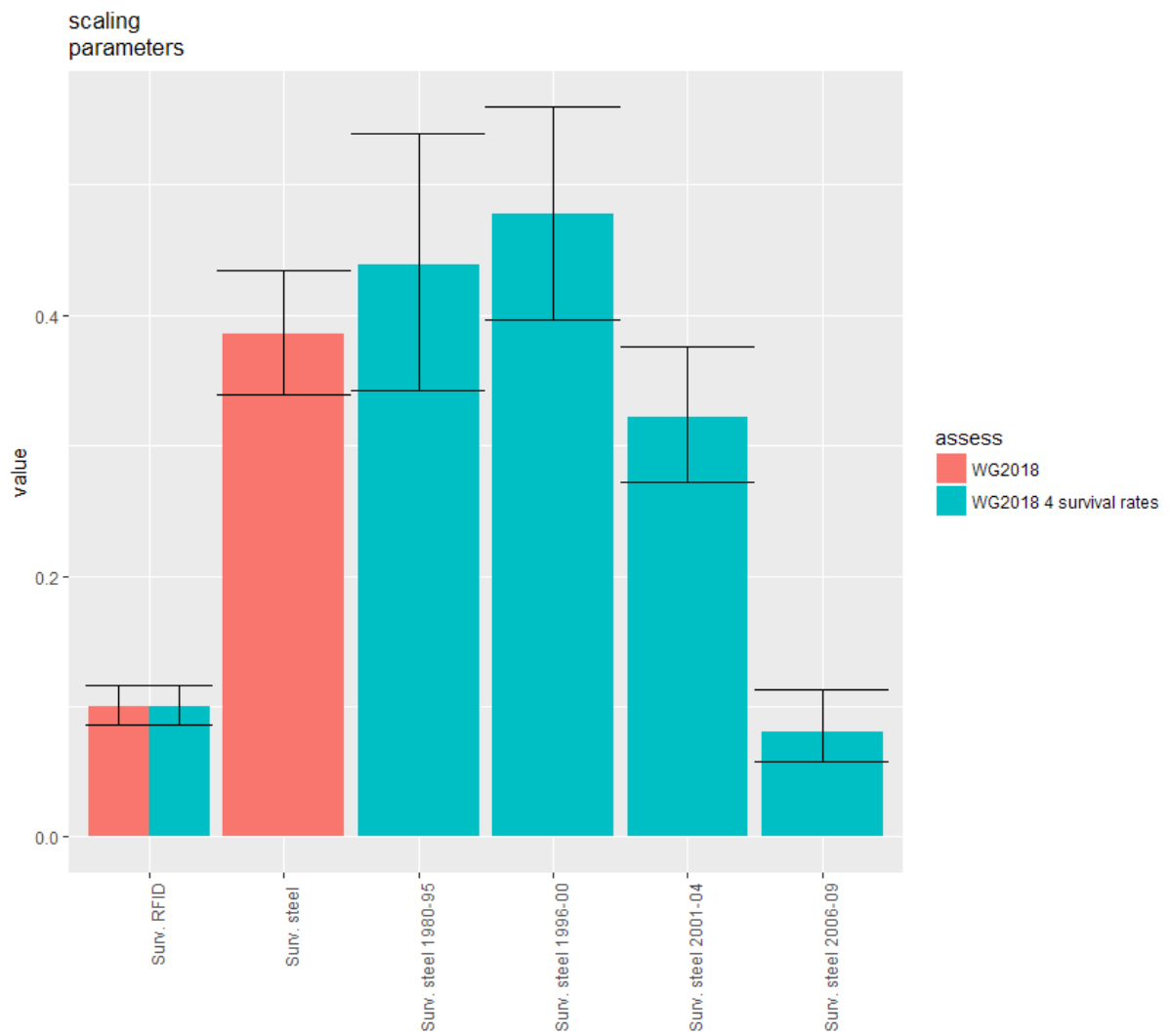


Figure 5 : detail of figure 4 focussing on estimated survival rates

North East Atlantic Mackerel assessment – sensitivity to error structure for the IESSNS

Working Document for the Benchmark Workshop on Widely Distributed Stocks (IBPNEAMac-2019). 11 -13th December 2018. Copenhagen (Denmark).

S. Sánchez¹ and T. Brunel².

¹ AZTI, Herrera Kaia - Portualdea z/g. E-20110 Pasaia – GIPUZKOA (Spain) | Mail: ssanchez@azti.es

² Wageningen Marine Research, Haringkade 1, 1976CP IJmuiden (The Netherlands) | Mail: thomas.brunel@wur.nl

1. Background information

North East Atlantic (NEA) mackerel assessment used was accepted at the Benchmark Workshop for Pelagic Stocks (WKWIDE) in 2017 (ICES, 2017a). The assessment model was the state-space assessment model (SAM). Currently SAM is provided as an R package, incorporating code for the inclusion of the RFID tags. For a more detailed description of the model see Nielsen and Berg (2014, 2016) and ICES (2017a).

As catches prior to 2000 were considered unreliable due to important underreporting, they were down weighted for this period by imposing a high observation variance for these catches in order to decrease their influence on the assessment. Additionally, model incorporates three survey indices (International Bottom-trawl Survey –IBTS- recruitment index, mackerel triennial egg survey SSB index and abundance indices from International Ecosystem Summer Survey in the Nordic Seas –IESSNS- survey) and tagging recapture data up to 2006. The age range in the model is 0-12, being age 12 a plus group and reference fishing mortality, F_{bar} , is calculated over ages 4 to 8. Compared to the previous assessment, the final accepted assessment in WKWIDE now uses, the IESSNS index expressed in abundance, for the ages 3 to 11, with an age varying estimated catchability, and a separate observation variance for age 3 and for ages 4 and older. The new assessment uses the new RFID tag/recapture data, parameterized with a survival rate and overdispersion parameters estimated separately from the historical steal tag data. Finally, the model uses an autoregressive (AR) autocorrelation structure for the observation for ages 3-11 for the IESSNS. No other change to the data used or the model configuration was made compared to the previous assessment.

In present work, the sensitivity of the assessment to the error structure for the IESSNS was analysed.

2. Material and methods

In order to test the sensitivity of the assessment to different error structures for the IESSNS, the assessment data and configuration used in the last ICES advice for mackerel; i.e. the one

agreed in WGWIDE 2018 (ICES, 2018), was taken as the base case. For the alternative case, the correlated error structure on the IESSNS is not included, so errors are assumed to be independent.

3. Results

The estimated AR error correlation matrix (Figure 3.1) shows a high correlation between neighboring age-classes (of 0.82) and these correlations decrease as age groups are more distant.

Figure 3.2 to Figure 3.10 show the results for the sensitivity analysis of the assessment model to the IESSNS error structure (i.e. comparing the one used in the last assessment in WGWIDE 2018 -base case-, which assumed a correlated error structure, relative to one which assumes independent error for the IESSNS among ages). All runs are available at the ICES working group SharePoint (IBPNEAMac 2019), in the “07. Software” folder ([link](#)).

When IESSNS errors are assumed to be independent, the historical stock development estimated by the model was very similar for the initial years, but increasing differences appear since 2011 (Figure 3.2). The assessed stock trends showed an upward revision of the SSB and downward revision for the fishing mortalities, relative to when correlated error structure is assumed (WGWIDE 2018). However, estimates are within the confidence intervals. The uncertainty in both SSB and F estimates was slightly lower in the case where IESSNS errors are assumed to be independent (Figure 3.3).

Regarding model parameter estimates, when IESSNS errors are assumed to be independent, the observation variances for the catches, the IESSNS and the recruitment decreases, whereas increased the observations variances for the egg survey (also recruitment variance and standard deviation for the F random walks, but to a lesser extent) and process error kept invariant (Figure 3.4). So, the weight given by the assessment to the egg survey increases, whereas it decreases for the rest of the indices and the catches. Both, the estimate of over dispersion of the RFID tag recaptures and the estimated post tagging survival rates slightly increased. However, in the case of the steel tags they kept almost invariant. Finally, uncertainty decreases for all the scaling parameters.

The standard deviations of the parameter estimates (Figure 3.5) were lower for most of the parameters, except for the observation variance of the catches and the scaling parameter for the egg survey index when assuming independent IESSNS errors.

Regarding the retrospective patterns (Figure 3.6), less marked retrospective patterns appear when assuming independent IESSNS errors (with SSB and F values within the confidence interval), whereas the year effect of including a new egg survey estimate appears one year before and the revision is lower.

Residual plots for the survey indices (Figure 3.7 and Figure 3.8) improve when assuming correlated errors for the IESSNS survey. However, tagging residuals (Figure 3.9 and Figure 3.10) show some clear trends.

Comparing the likelihoods of the models with and without correlation structure (log likelihood increasing from -2604.799 to -2785.824 for one additional parameter) shows that the assumption of independent observation errors for the IESSNS index is rejected ($p = 7.25 \text{ e-}10$).

Table 0.1. Comparison of the models with and without correlation structure in the IESSNS residuals (log-likelihood, number of parameters, p-value and AIC by model).

model	Log lik	N pars	P val	AIC
No correlation	-2804.799	22		5654
AR correlation	-2785.824	23	7.254674e-10	5618

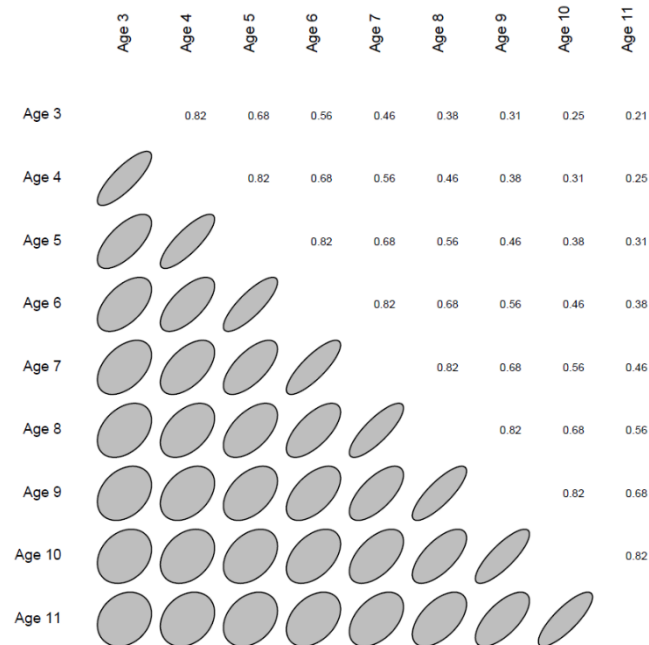


Figure 3.1. Correlation structure estimated in the WGWIDE 2018 assessment for the IESSNS residuals.

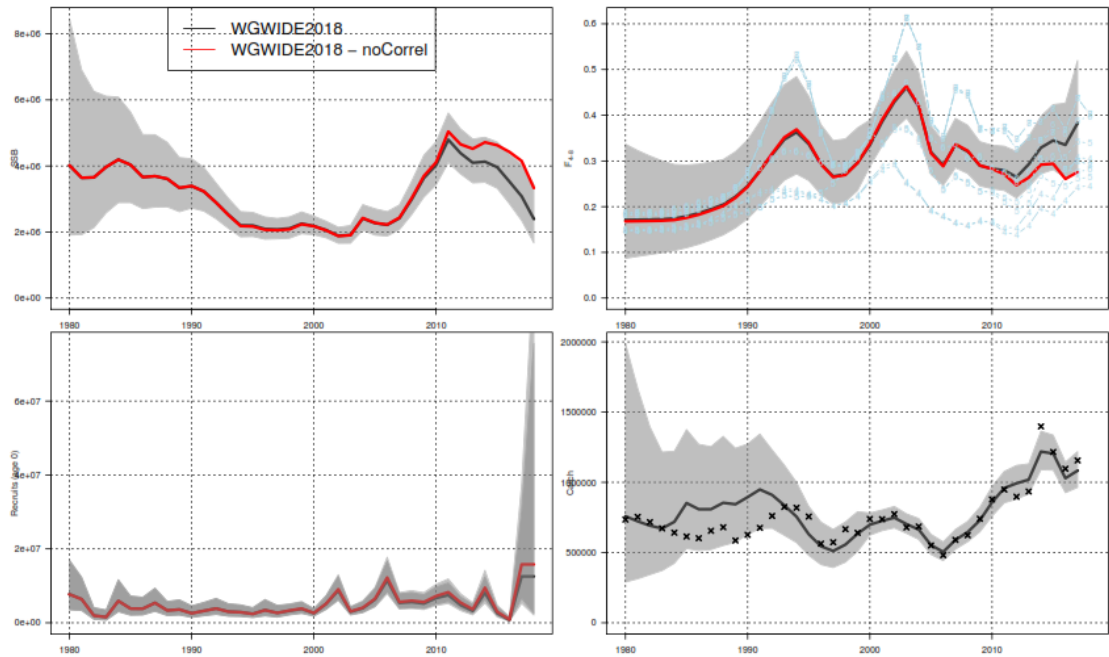


Figure 3.2. Summary of stock assessment, comparison between WGWIDE 2018 assessment (in black with confidence interval as dashed grey area) with the alternative cases excluding RFID tagging information (in RED). From left to right and top to bottom: SSB (in tonnes), F (F_{bar} for ages 4-8), recruits (millions) and catches (in tonnes).

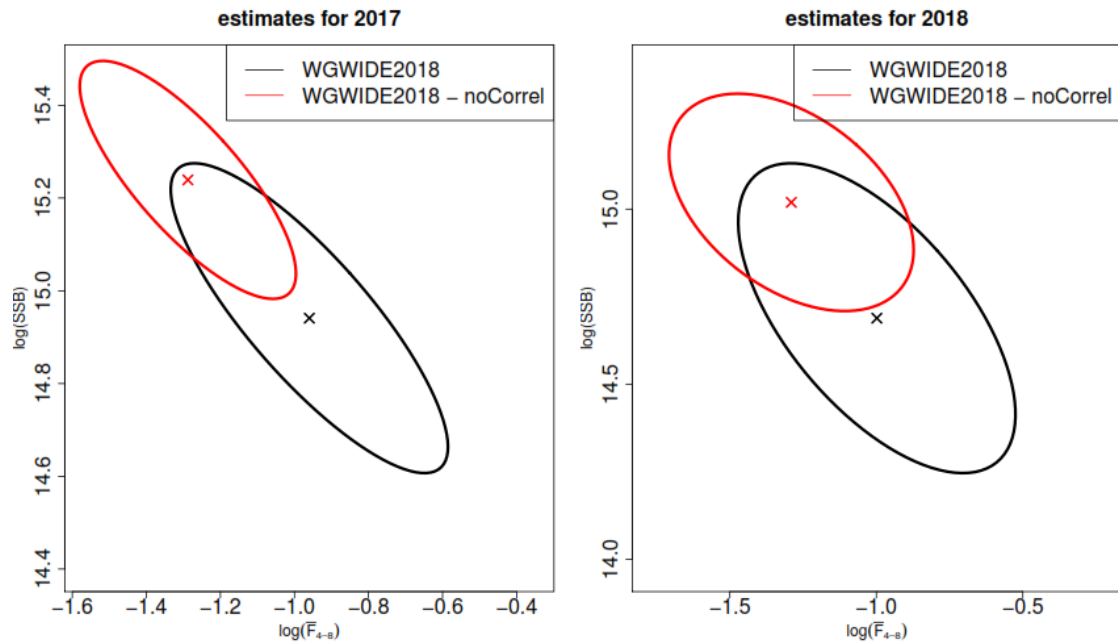


Figure 3.3. Joint distribution of the (log) estimates of SSB and F_{bar} for the last two years in the assessment, for WGWIDE 2018 assessment (in black) and the alternative case which assumes independent errors for the IESSNS (in red). The point depicts the point estimate and the ellipsis represents the probability intervals at different confidence levels.

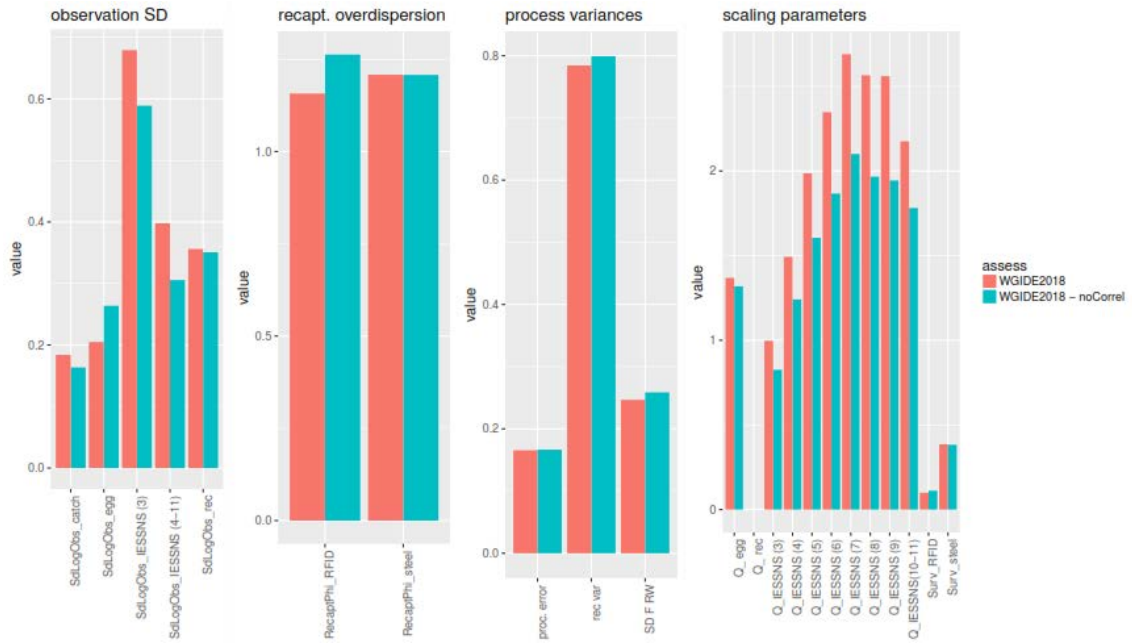


Figure 3.4. Parameter estimates for two alternative SAM runs: base case (in red) and the alternative case which assumes independent errors for the IESSNS (in green). From left to right: i) observation standard deviation for the catches and the surveys; ii) tagging recapture overdispersion; iii) process variances: process error, recruitment variance and standard deviation for the F random walk; and iv) scaling parameters: catchability for the different surveys and tagging survival rate.

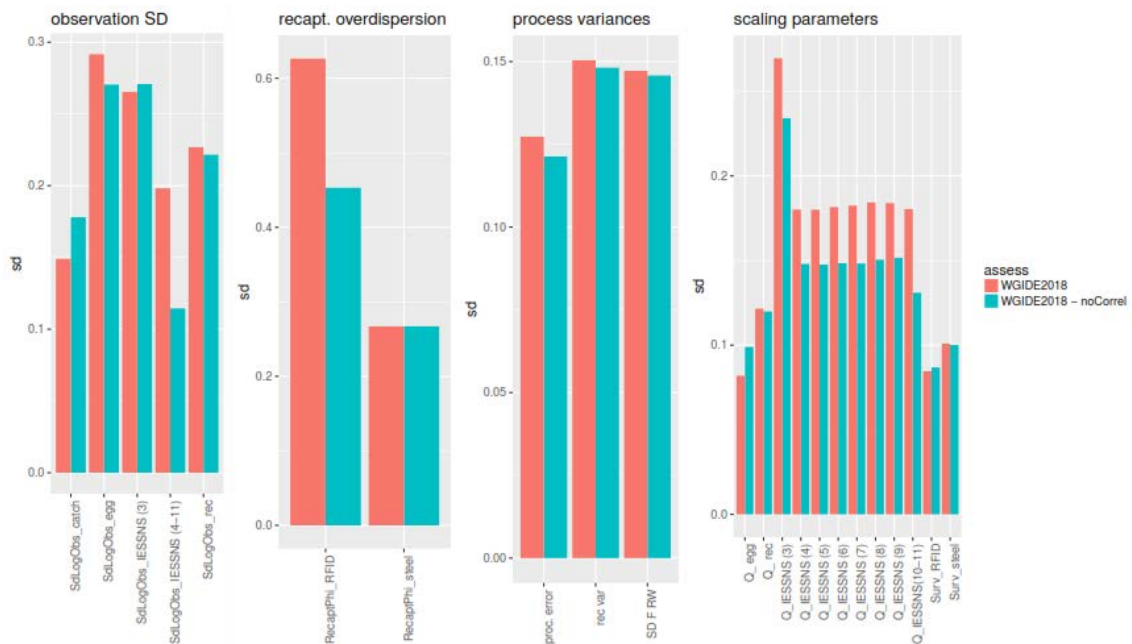


Figure 3.5. Standard deviation of the model parameter estimates for alternative SAM runs: base case (in red) and the alternative case which assumes independent errors for the IESSNS (in green). From left to right: i) observation standard deviation for the catches and the surveys; ii) tagging recapture overdispersion; iii) process variances: process error, recruitment variance and standard deviation for the F random walk; and iv) scaling parameters: catchability for the different surveys and tagging survival rate.

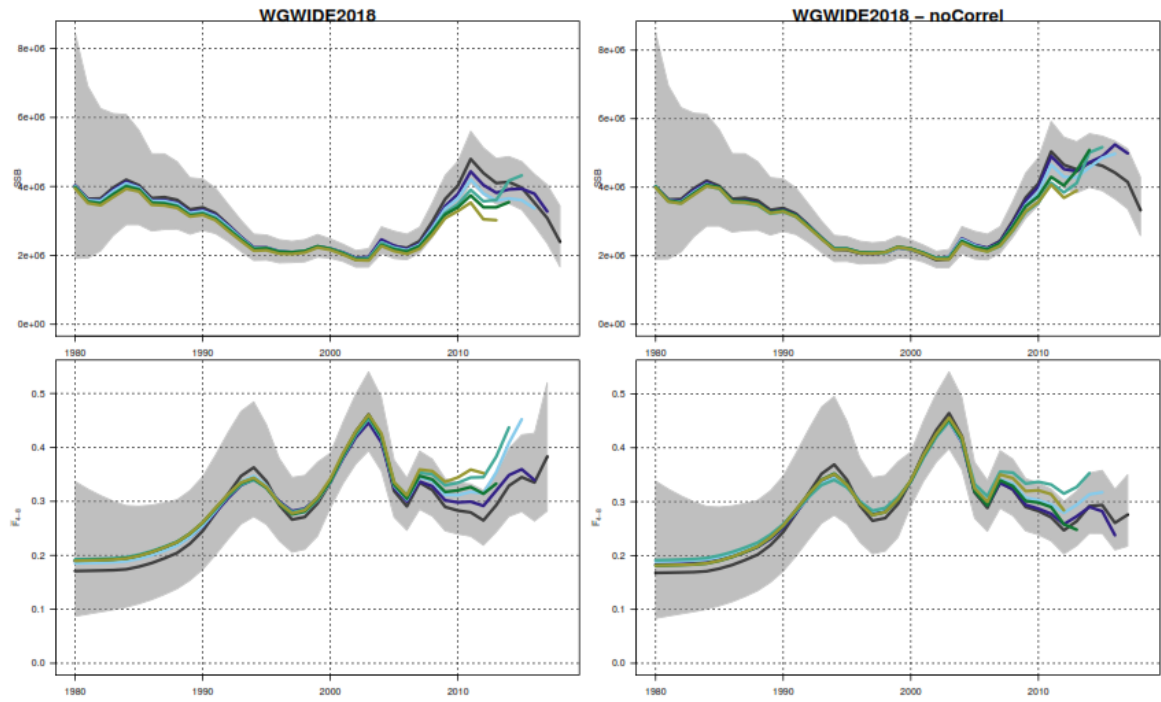


Figure 3.6. Retrospective analysis of SSB (in tonnes) (see top plots) and F (F_{bar} for ages 4-8) (see bottom plots) for alternative SAM runs: base case (on the left) and the alternative case which assumes independent errors for the IESSNS (on the right) The 95% confidence interval is shown for the base case.

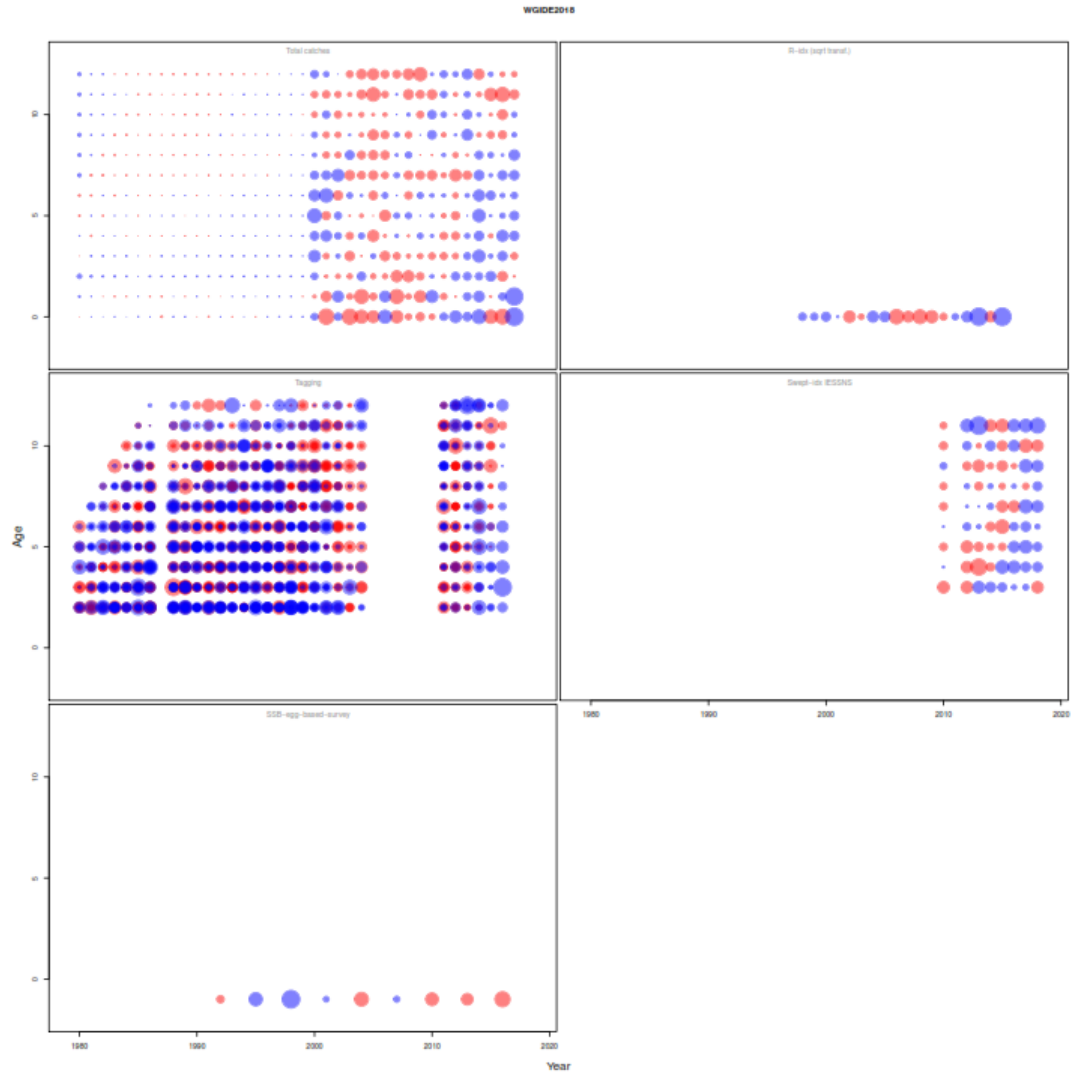


Figure 3.7. OSA (One Step Ahead) residuals (see Berg and Nielsen, 2016) for the WGWIDE 2018 assessment from top to bottom and left to right: catch at age, recruitment index, tagging, IESSN survey and egg survey. Blue circles indicate positive residuals (observation larger than predicted) and filled red circles indicate negative residuals.

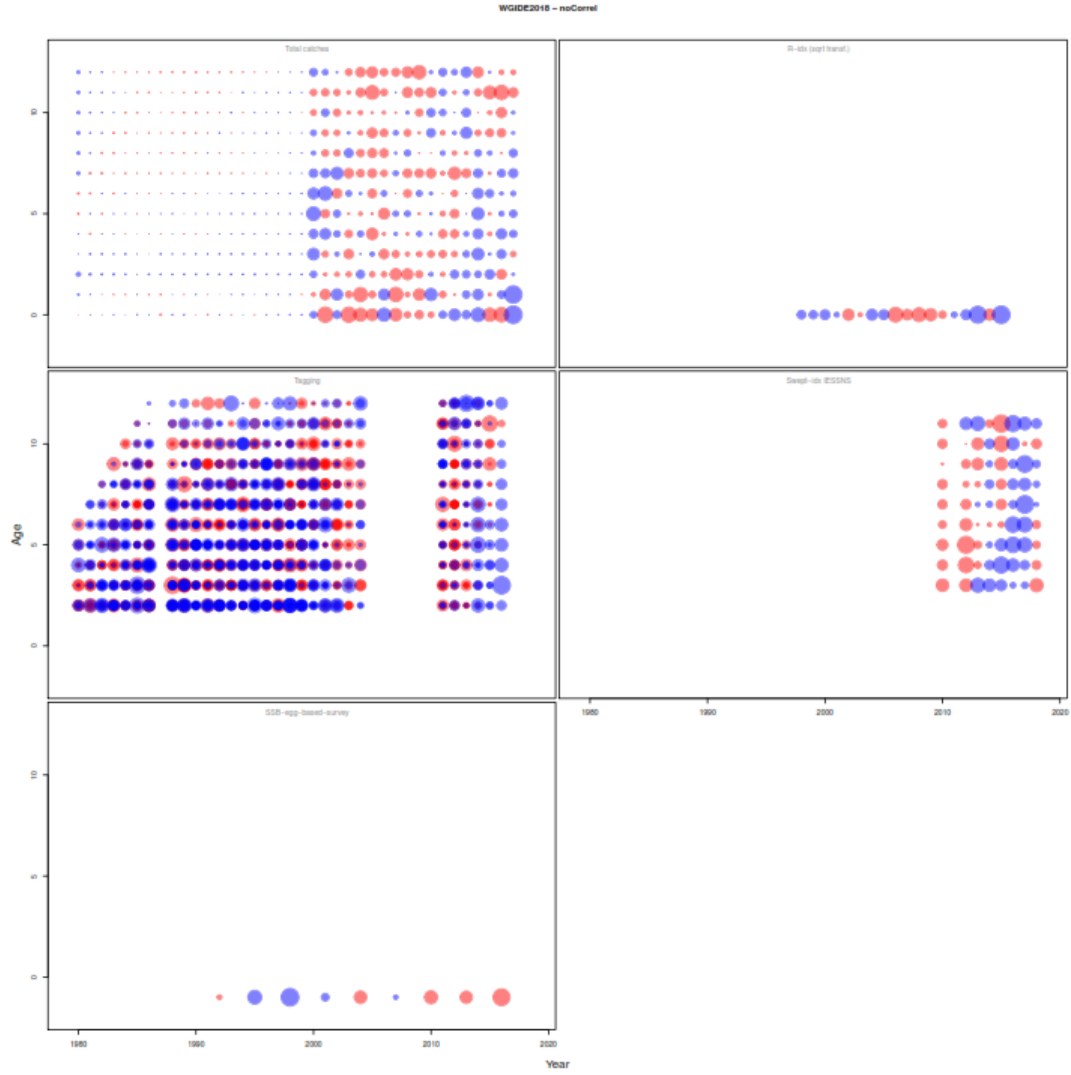


Figure 3.8. OSA (One Step Ahead) residuals (see Berg and Nielsen, 2016) for the alternative case in which independent errors for the IESSNS are assumed, from top to bottom and left to right: catch at age, recruitment index, tagging, IESSN survey and egg survey. Blue circles indicate positive residuals (observation larger than predicted) and filled red circles indicate negative residuals.

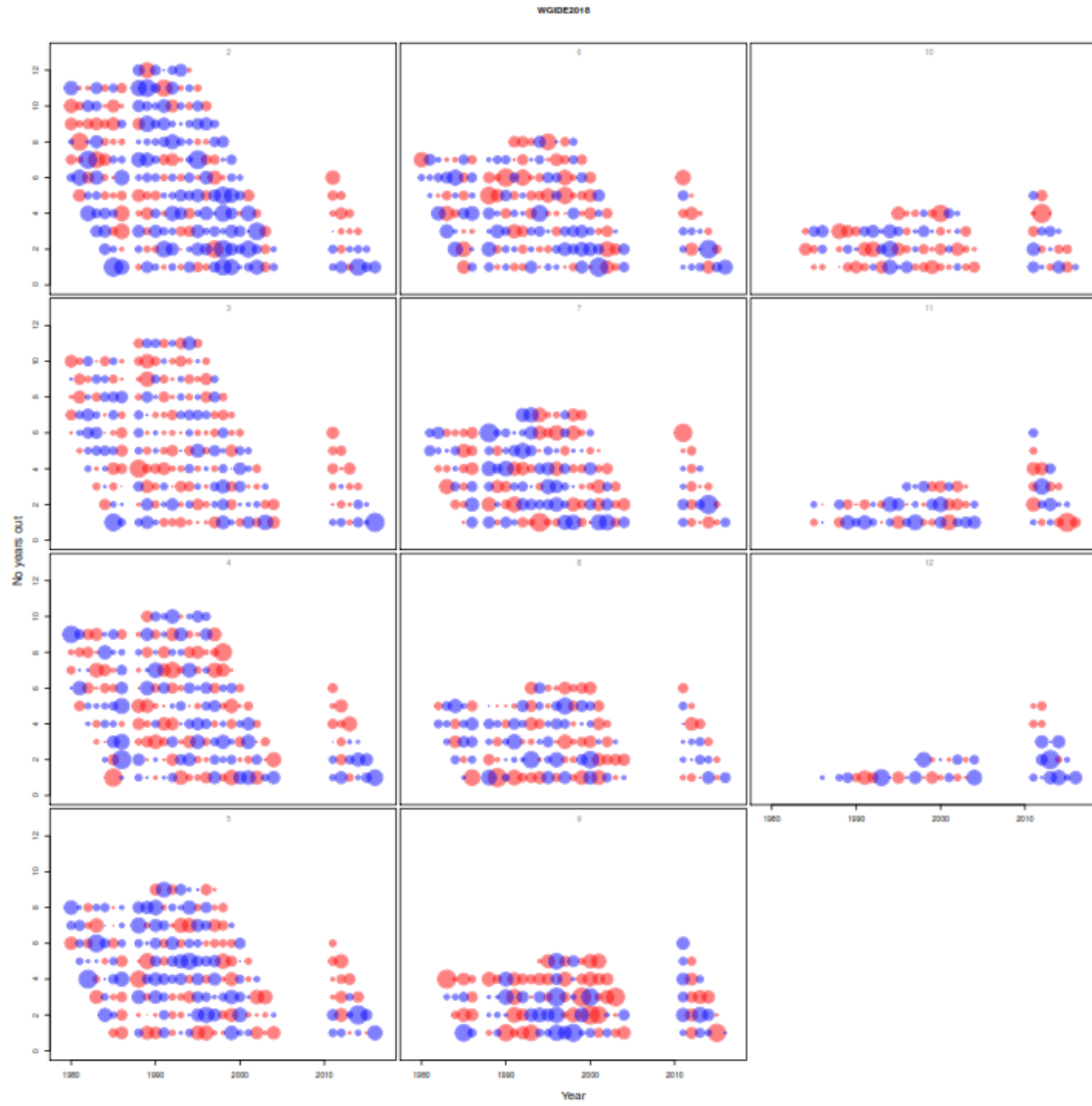


Figure 3.9. OSA (One Step Ahead) residuals (see Berg and Nielsen, 2016) from tagging information for the WGWIDE 2018 assessment. Blue circles indicate positive residuals (observation larger than predicted) and filled red circles indicate negative residuals.

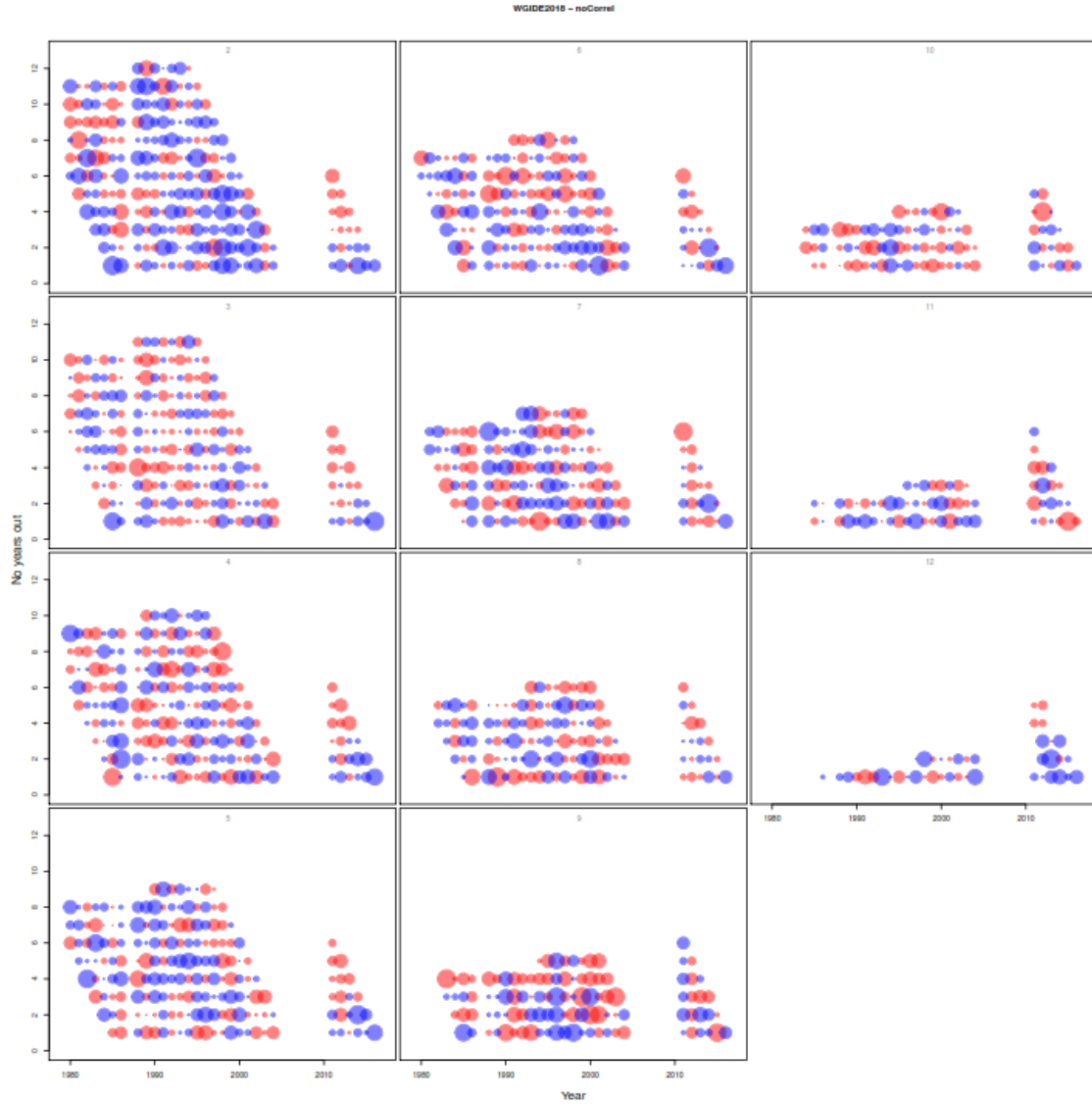


Figure 3.10. OSA (One Step Ahead) residuals (see Berg and Nielsen, 2016) from tagging information for the alternative case in which independent errors for the IESSNS are assumed. Blue circles indicate positive residuals (observation larger than predicted) and filled red circles indicate negative residuals.

4. Conclusion

- There is some improvement when IESSNS errors are assumed to be independent (more accurate parameter estimates, less marked retrospective patterns and lower revision when new egg index is included). However, tag residuals show some clear trends.
- On the other hand, patterns are observed in the tagging residuals.
- A selection based on a statistical test opts for a model with correlation.

5. References

- Berg, C. W. and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES Journal of Marine Science: Journal du Conseil* 73: 1788-1797.
- ICES. 2017a. Report of the Benchmark Workshop on Pelagic Stocks (WKWIDE 2017). 17-21 February 2014; Copenhagen, Denmark. ICES CM 2014/ACOM:43. 339 pp.
- ICES. 2017b. Report of the Working Group on Widely Distributed Stocks (WGWIDE). 30 August-5 September 2017, ICES HQ, Copenhagen, Denmark. ICES CM 2017/ACOM:23. 1095 pp.
- ICES. 2018. Report of the Working Group on Widely Distributed Stocks (WGWIDE). 28 August-3 September 2018; Torshavn, Faroe Islands. ICES CM 2018/ACOM:23. 488 pp.
- Nielsen, A. and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research* 158: 96-101.

North East Atlantic Mackerel assessment – sensitivity to RFID tags

Working Document for the Benchmark Workshop on Widely Distributed Stocks (IBPNEAMac-2019). 11 -13th December 2018. Copenhagen (Denmark).

S. Sánchez¹ and T. Brunel².

¹ AZTI, Herrera Kaia - Portualdea z/g. E-20110 Pasaia – GIPUZKOA (Spain) | Mail: ssanchez@azti.es

² Wageningen Marine Research, Haringkade 1, 1976CP IJmuiden (The Netherlands) | Mail: thomas.brunel@wur.nl

1. Background information

North East Atlantic (NEA) mackerel assessment used was accepted at the Benchmark Workshop for Pelagic Stocks (WKWIDE) in 2017 (ICES, 2017a). The assessment model was the state-space assessment model (SAM). Currently SAM is provided as an R package, incorporating code for the inclusion of the RFID tags. For a more detailed description of the model see Nielsen and Berg (2014, 2016) and ICES (2017a).

As catches prior to 2000 were considered unreliable due to important underreporting, they were down weighted for this period by imposing a high observation variance for these catches in order to decrease their influence on the assessment. Additionally, model incorporates three survey indices (International Bottom-trawl Survey –IBTS- recruitment index, mackerel triennial egg survey SSB index and abundance indices from International Ecosystem Summer Survey in the Nordic Seas –IESSNS- survey) and tagging recapture data up to 2006. The age range in the model is 0-12, being age 12 a plus group and reference fishing mortality, F_{bar} , is calculated over ages 4 to 8. Compared to the previous assessment, the final accepted assessment in WKWIDE now uses, the IESSNS index expressed in abundance, for the ages 3 to 11, with an age varying estimated catchability, and a separate observation variance for age 3 and for ages 4 and older. The new assessment uses the new RFID tag/recapture data, parameterized with a survival rate and overdispersion parameters estimated separately from the historical steal tag data. Finally, the model uses an autoregressive (AR) autocorrelation structure for the observation for ages 3-11 for the IESSNS. No other change to the data used or the model configuration was made compared to the previous assessment.

In present work, the sensitivity of the assessment to the omission of the RFID data was analysed.

2. Material and methods

In order to test the sensitivity of the assessment to the use of the RFID data, the assessment data and configuration used in the last ICES advice for mackerel; i.e. the one agreed in WGWIDE 2018 (ICES, 2018), was taken as the base case. For the alternative case, the model

was fit using the same data and configuration as in the base case, but RFID tags were excluded from the assessment data.

3. Results

Figure 3.11 to Figure 3.20 show the results for the sensitivity analysis of the assessment model to the omission of the RFID tagging data relative to the last assessment in WGWIDE 2018. All runs are available at the ICES working group SharePoint (IBPNEAMac 2019), in the “07. Software” folder ([link](#)).

When removing the RFID tags, the historical stock development estimated by the model was very similar for the initial years, but marked differences appear mainly in the last years (Figure 3.11). The assessed stock trends showed an upward revision of the SSB, with increasing SSB until 2015, relative to the base case (WGWIDE2018 assessment) where SSB had a maximum in 2011 and downward revision for the fishing mortalities. The uncertainty both in SSB and F estimates was very similar for both cases (Figure 3.13).

Repeating this exercise based on the 2017 assessment yields different results (Figure 3.12), as in this case the sensitivity to the RFID tags is also large, but less than in the 2018 assessment. This suggests that some of the new data incorporated in the 2018 changed drastically the behavior of the assessment (as there has been no revision in the survey indices).

Regarding model parameter estimates, when omitting the RFID tagging data, the observation variances for the IESSNS and the recruitment and process error, recruitment variance and standard deviation for the F random walks decreased, whereas observation variances for the egg survey index increased and the variances for the catches marginally changes (Figure 3.14). That is the weight given by the assessment to the egg survey increases, whereas it decreases for the rest of the indices. Both, the estimate of over dispersion of the steel tag recaptures and the estimated post tagging survival rates remained unchanged.

The standard deviations of the parameter estimates (Figure 3.15) were higher for most of the parameters, except for the scaling parameter where they were lower for all the indices except the egg survey one, when removing RFID information.

Regarding the retrospective patterns (Figure 3.16), more marked retrospective patterns appear when not including the RFID tagging data (with SSB and F values out of the confidence interval with less years of data removed), whereas the year effect of including a new egg survey estimate were less marked in this case.

Residual plots for the survey indices (Figure 3.17 and Figure 3.17) and for the tagging (Figure 3.18 and Figure 3.18) were very similar when using or not the in-year indices.

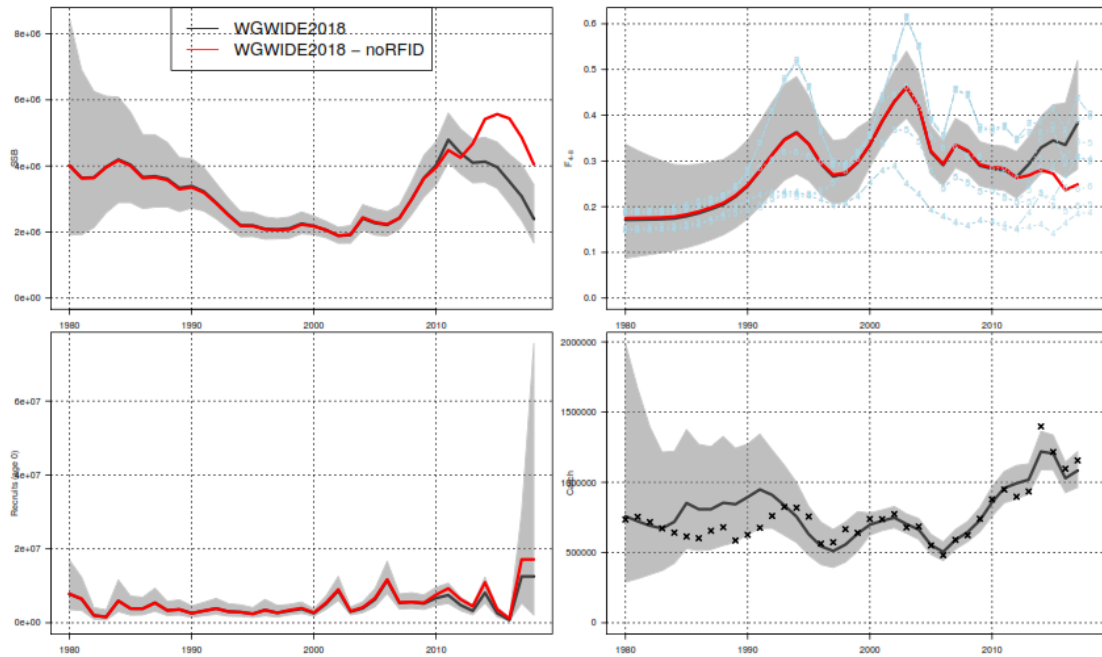


Figure 3.1. Summary of stock assessment, comparison between WGWIDE 2018 assessment (in black with confidence interval as dashed grey area) with the alternative cases excluding RFID tagging information (in RED). From left to right and top to bottom: SSB (in tonnes), F_{4-8} (F_{bar} for ages 4-8), recruits (millions) and catches (in tonnes).

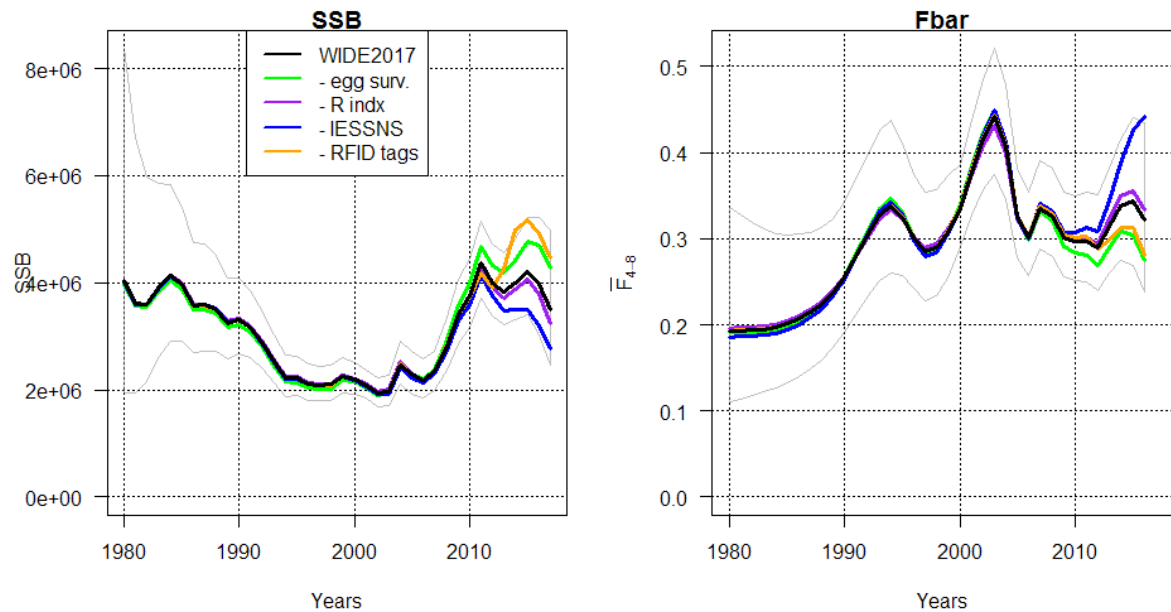


Figure 3.2. Leave one out run based on the WGWIDE 2017 assessment (ICES, 2017b).

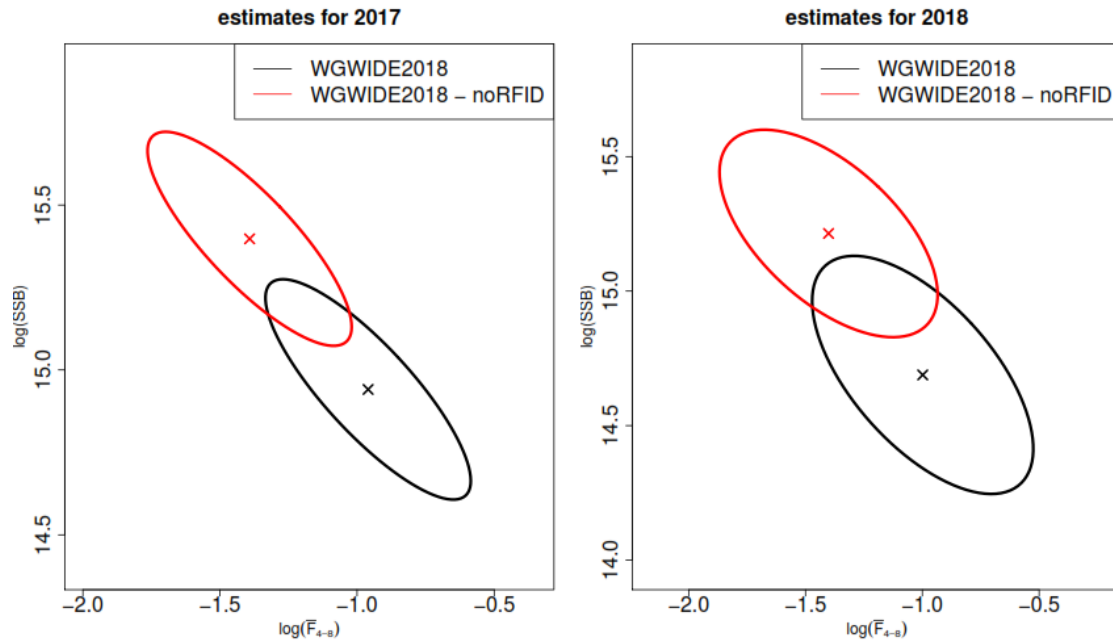


Figure 3.3. Joint distribution of the (log) estimates of SSB and F_{bar} for the last two years in the assessment, for WGWIDE 2018 assessment (in black) and the alternative case excluding RFID tagging information (in red). The point depicts the point estimate and the ellipsis represents the probability intervals at different confidence levels.

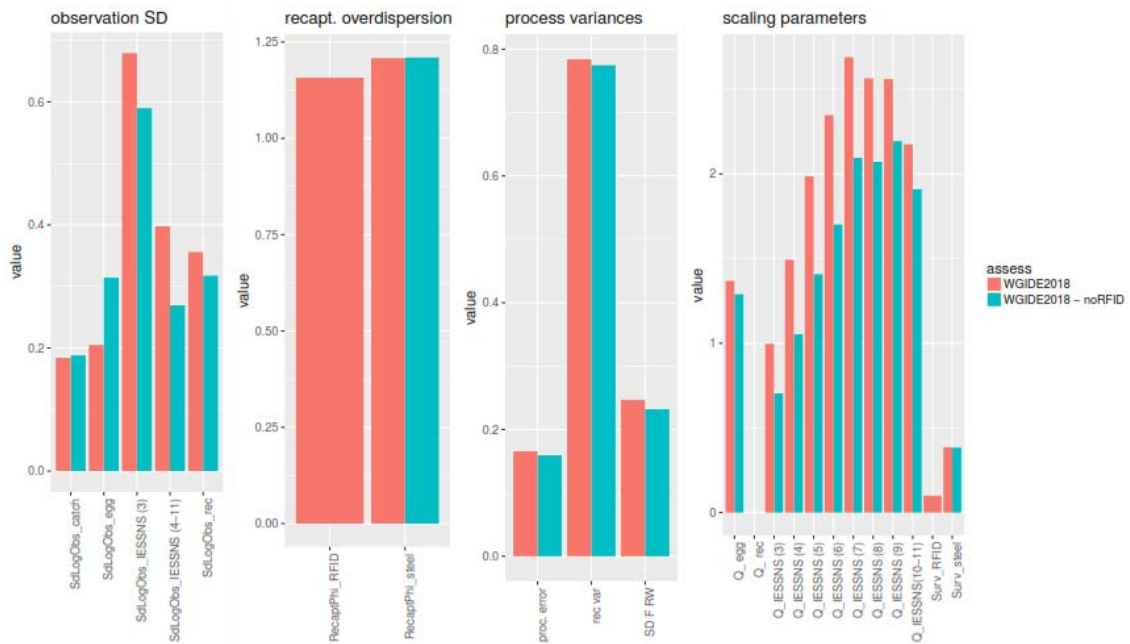


Figure 3.4. Parameter estimates for two alternative SAM runs: base case (in red) and the alternative case excluding RFID tagging information (in green). From left to right: i) observation standard deviation for the catches and the surveys; ii) tagging recapture overdispersion; iii) process variances: process error, recruitment variance and standard deviation for the F random walk; and iv) scaling parameters: catchability for the different surveys and tagging survival rate.

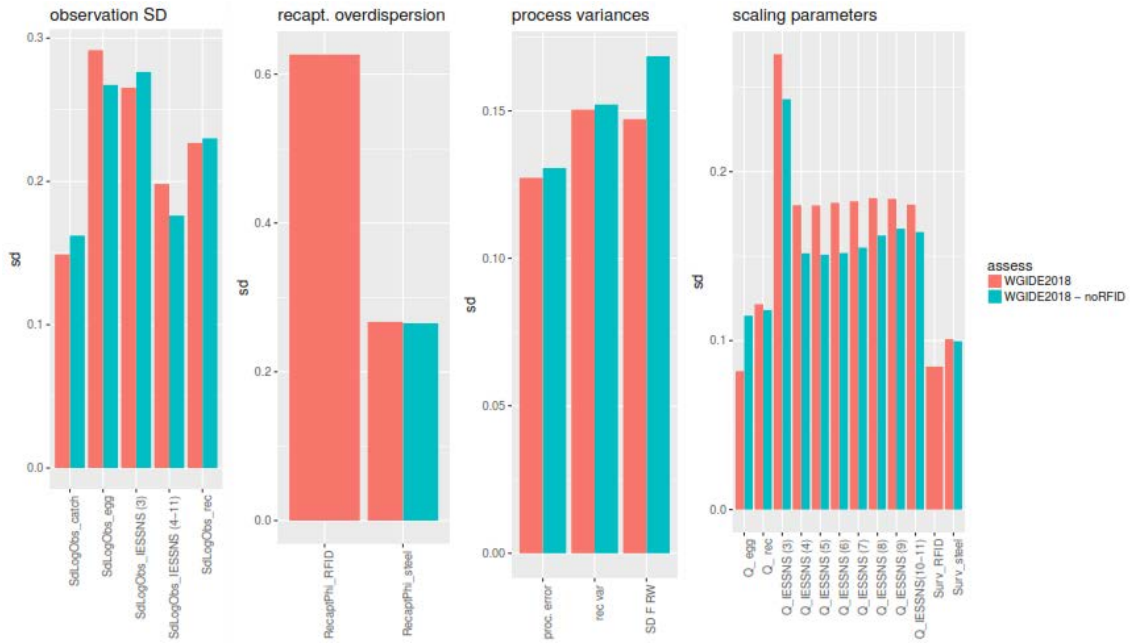


Figure 3.5. Standard deviation of the model parameter estimates for alternative SAM runs: base case (in red) and the alternative case excluding RFID tagging information (in green). From left to right: i) observation standard deviation for the catches and the surveys; ii) tagging recapture overdispersion; iii) process variances: process error, recruitment variance and standard deviation for the F random walk; and iv) scaling parameters: catchability for the different surveys and tagging survival rate.

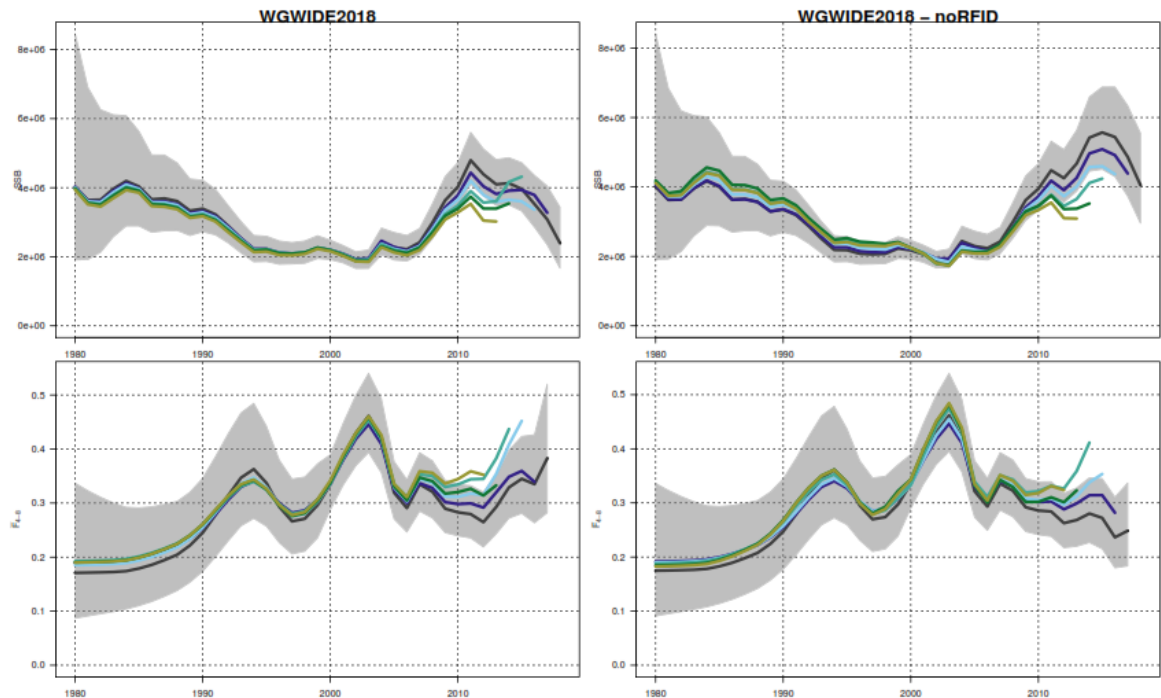


Figure 3.6. Retrospective analysis of SSB (in tonnes) (see top plots) and F (F_{bar} for ages 4-8) (see bottom plots) for alternative SAM runs: base case (on the left) and the alternative case excluding RFID tagging information (on the right) The 95% confidence interval is shown for the base case.

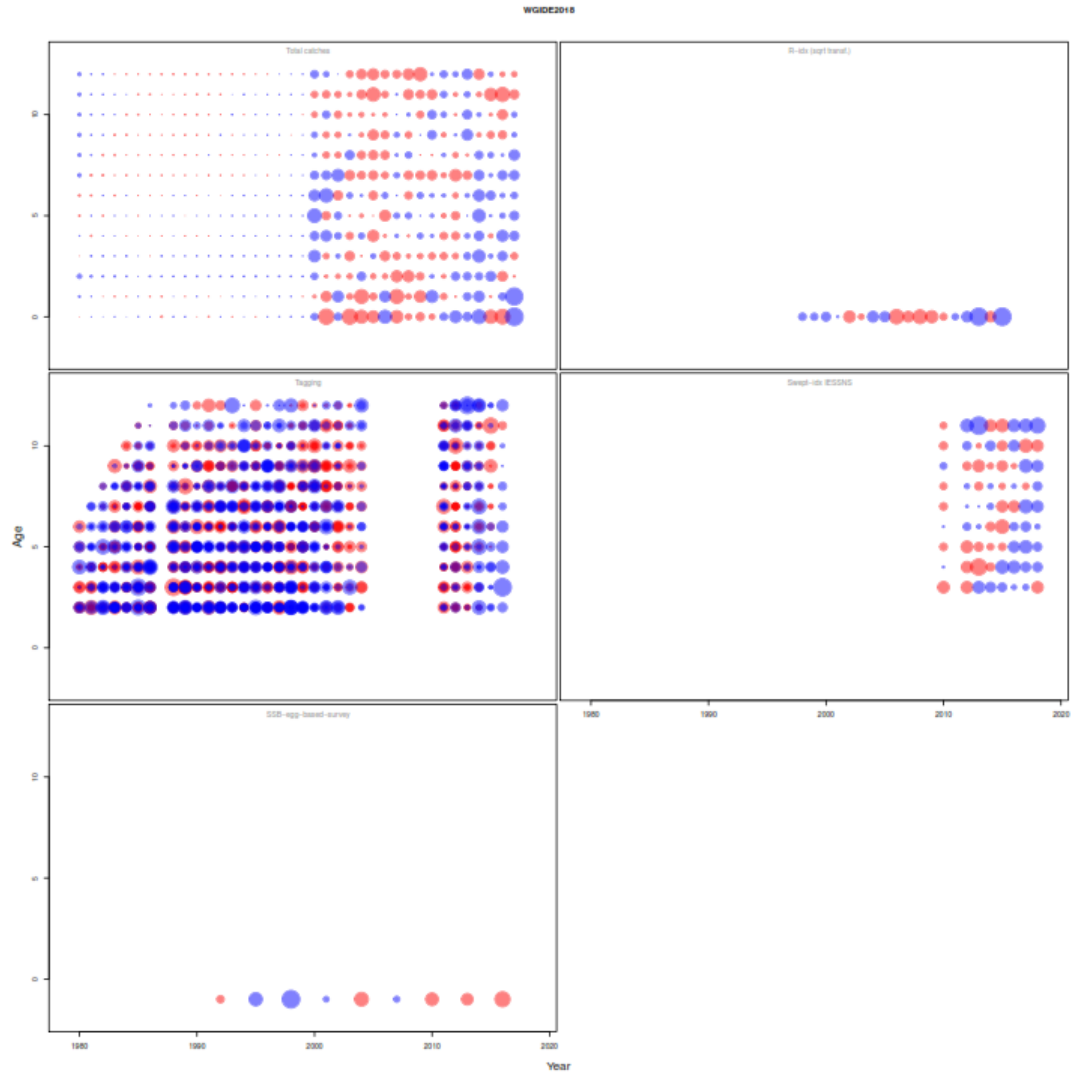


Figure 3.7. OSA (One Step Ahead) residuals (see Berg and Nielsen, 2016) for the WGWIDE 2018 assessment from top to bottom and left to right: catch at age, recruitment index, tagging, IESSN survey and egg survey. Blue circles indicate positive residuals (observation larger than predicted) and filled red circles indicate negative residuals.

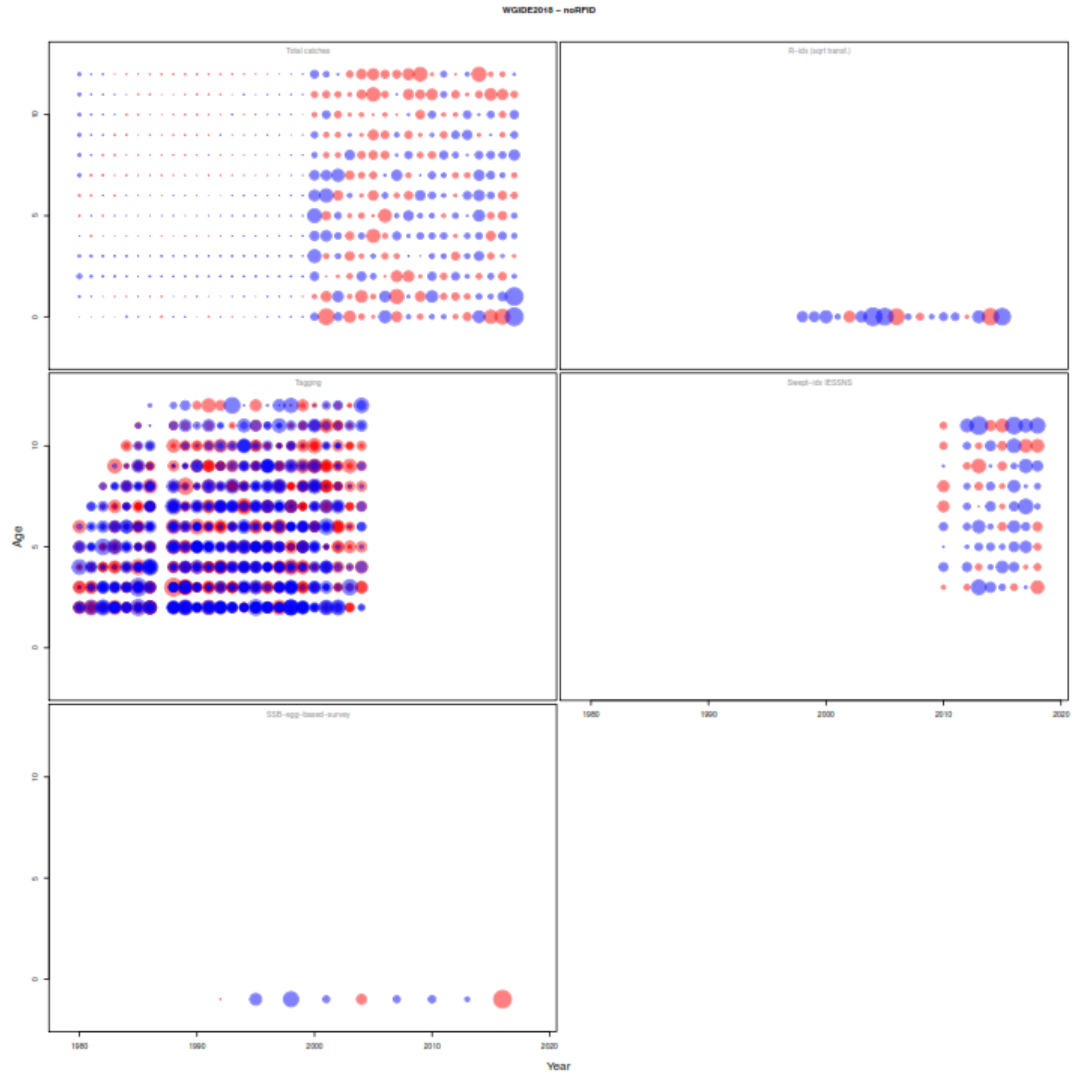


Figure 3.8. OSA (One Step Ahead) residuals (see Berg and Nielsen, 2016) for the alternative case in which RFID tagging information has been excluded, from top to bottom and left to right: catch at age, recruitment index, tagging, IESSN survey and egg survey. Blue circles indicate positive residuals (observation larger than predicted) and filled red circles indicate negative residuals.

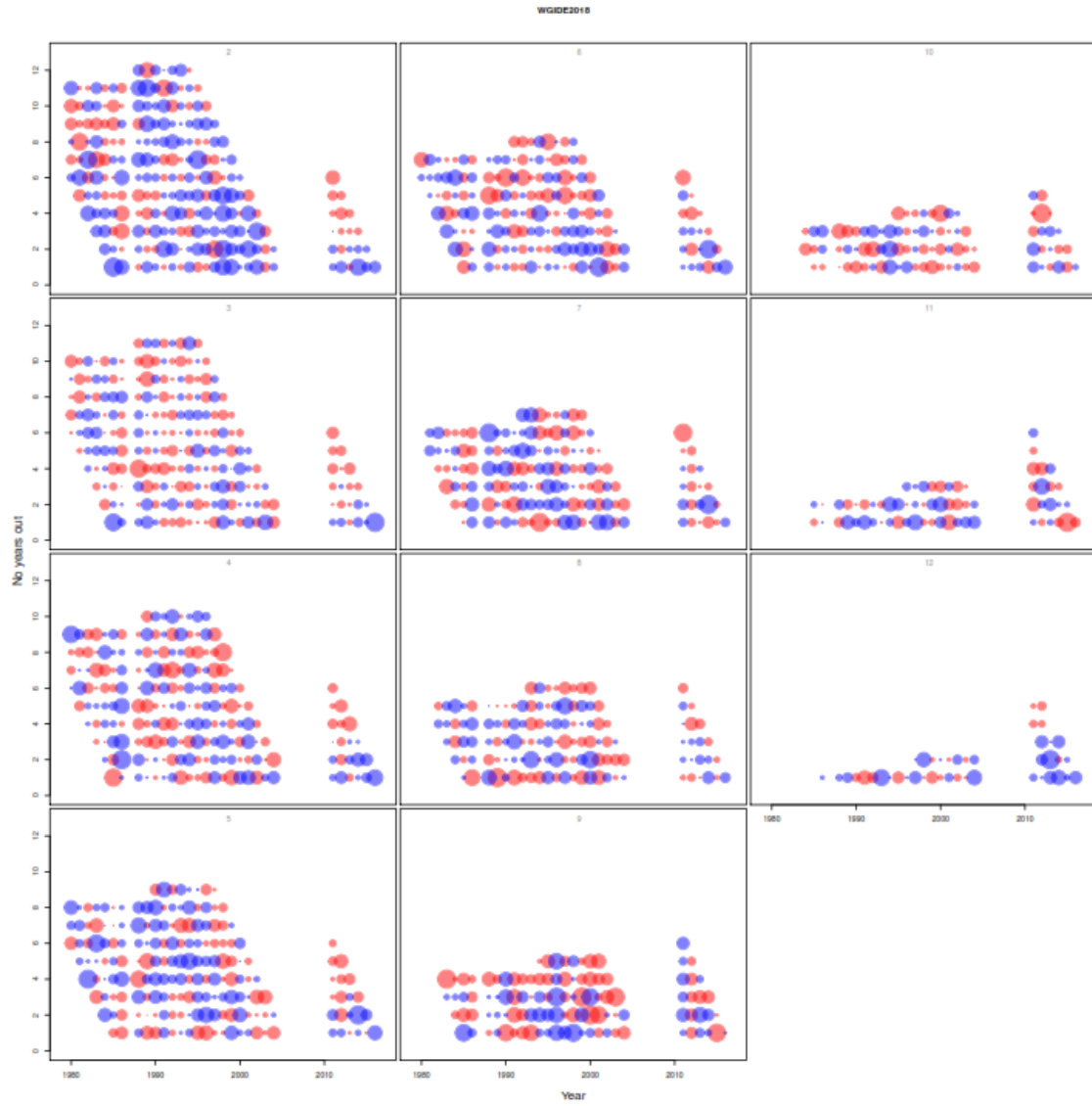


Figure 3.9. OSA (One Step Ahead) residuals (see Berg and Nielsen, 2016) from tagging information for the WGWIDE 2018 assessment. Blue circles indicate positive residuals (observation larger than predicted) and filled red circles indicate negative residuals.

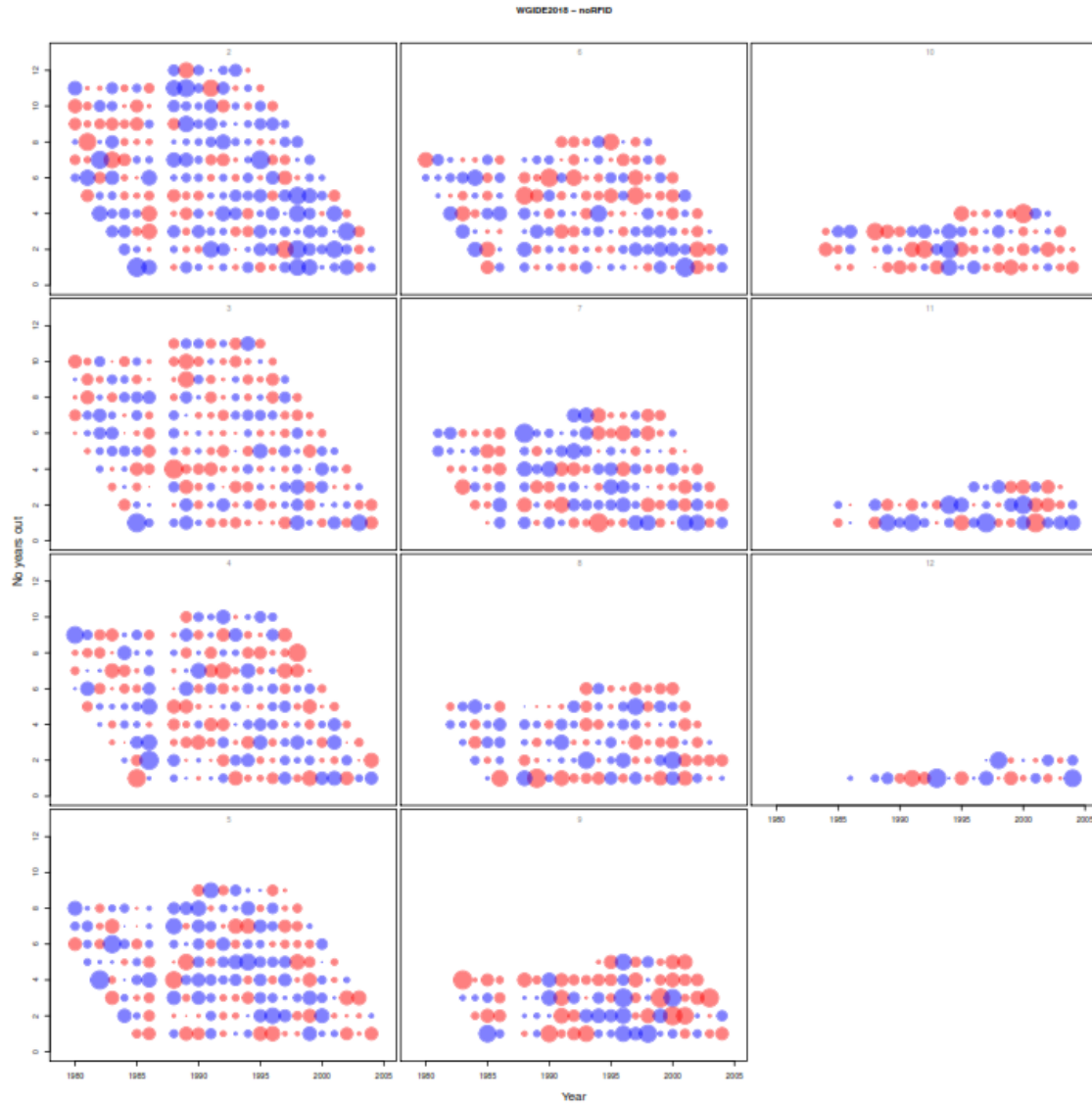


Figure 3.10. OSA (One Step Ahead) residuals (see Berg and Nielsen, 2016) from tagging information for the alternative case in which RFID tagging information has been excluded. Blue circles indicate positive residuals (observation larger than predicted) and filled red circles indicate negative residuals.

4. Conclusion

4.1. Omission of RFID tags

- There is no marked improvement when RFID tagging data are not used. Moreover, although the year effect of including a new egg survey estimate seems to less clear, the retrospective patterns are more marked in this case. However, there is still unknown if these patterns are due to correlation (due to increasing mixing as year after tagging increase) or of tag loss.
- RFID tagging data is an important source of information, but as it is a short series, the impact of it in the assessment still could change from one year to the following until a longer series is available, as we observed when including the 2017 tag recoveries.

5. References

- Berg, C. W. and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES Journal of Marine Science: Journal du Conseil* 73: 1788-1797.
- ICES. 2017a. Report of the Benchmark Workshop on Pelagic Stocks (WKWIDE 2017). 17-21 February 2014; Copenhagen, Denmark. ICES CM 2014/ACOM:43. 339 pp.
- ICES. 2017b. Report of the Working Group on Widely Distributed Stocks (WGWIDE). 30 August-5 September 2017, ICES HQ, Copenhagen, Denmark. ICES CM 2017/ACOM:23. 1095 pp.
- ICES. 2018. Report of the Working Group on Widely Distributed Stocks (WGWIDE). 28 August-3 September 2018; Torshavn, Faroe Islands. ICES CM 2018/ACOM:23. 488 pp.
- Nielsen, A. and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research* 158: 96-101.

Sensitivity of the mackerel assessment to the input data, with particular emphasis on the RFID tagging data

Thomas Brunel (Wageningen Marine Research)

Working document to the Interbenchmark Workshop on the assessment of northeast Atlantic mackerel

Introduction

During the 2018 WGWIDE, analyses were carried out that suggested that the tagging data had become the most influential data source in the mackerel assessment. The aim of the present working document is to conduct an extensive sensitivity analysis of the assessment output to the different data sources used in the assessment to get a better understanding of the respective weight of these data sources.

Methods

- Leave one out runs

Leave one out (LOO) runs consist in running the assessment excluding one by one each of the data sources (except the catches). In previous years, WGWIDE conducted LOO runs using the available function in the stockassessment library. In the case of the run without tagging data, the function removed the whole data set (i.e. steel tags and RFID tags). In this specific run, the only information available for the 20 years prior to 2000 consists of 3 egg survey points and 2 data point for the recruitment index. The SAM model for this run estimated a F random walk variance almost equal to zero (i.e. constant Fbar), indicating that such an assessment was clearly not comparable with the current assessment. In order to have an assessment which is consistent with the current one, but still estimating the influence of the tagging data, the LOO run for the tagging data was conducted by removing only the RFID tags.

- Sensitivity to single data year

In order to understand the patterns observed in the LOO runs, and check for the influence of any specific data year, the model was run removing data for each year one at the time for each data source.

Results

- Leave one out runs

The LOO run clearly show the strong influence of the RFID tags on the mackerel assessment (figure 1). As observed in previous years, removing the recruitment index has little influence on the SSB and Fbar trajectories. Removing the egg survey and the IESSNS had effects in the opposite directions : upwards change in SSB when the egg survey is removed and the opposite for the IESSNS, as observed since the 2014 benchmark. These three LOO runs give broadly similar trends in the recent years for the SSB.

The run removing the RFID tags is markedly different, with an increasing SSB until 2015, while the other runs (including the current assessment) have a maximum SSB in 2011. The estimated Fbar is also much lower in the recent years.

Repeating this exercise on the basis of the 2017 assessment yields different results : the sensitivity to the egg survey and to the IESSNS is higher (larger differences between the 2017 assessment and the run without any of the 2 surveys). The sensitivity to the RFID tags is also large, but less than in the 2018 assessment. This suggests that some of the new data incorporated in the 2018 changed drastically the behavior of the assessment (as there has been no revision in the survey indices).

- Influence of specific data years

The different runs each removing one specific year of recaptures for the RFID tagging data show that the latest recaptures (2017) have a strong effect on the assessment (figure 3). Removing this data year (run highlighted in green on figure 3) results in a much higher SSB and lower Fbar in the recent years. In comparison, removing the earlier years of recapture have less effect. The effect actually becomes smaller when older recapture years are removed (because tags recovered come from less release years).

Removing the SSB indices from the egg survey one by one did not result in any substantial difference in the stock trajectories (figure 4).

Removing IESSNS index years did not affect substantially the stock trajectories either, except for the most recent data year (2018, in green on figure 5), which had a small effect.

- Influence on the 2017 recaptures

The model run without the 2017 recaptures has slightly different parameters than the current assessment (figure 6) : observation standard deviation for the egg survey is higher, indicating a poorer fit, while it is higher for the IESSNS. The overdispersion for the RFID tags is a little lower when the recapture year 2017 is excluded.

The differences in SSB observed on figure 3 are due to higher stock numbers at age (figure 7). Differences in stock numbers for age 6 and older are large for the last 3 years (2016-2018). For younger ages, large differences are also found further back in time, up to 2012 for age 0. Differences in fishing mortality are large for all ages since 2014 (figure 8).

The residuals (one step ahead) from the current assessment for the recaptures of the RFID tags show a clear pattern for the recaptures from 2017 (figure 9) : while recaptures from earlier releases (before 2014) all tend to have negative residuals, recapture from more recent experiments (2015 and 2016) have positive residuals. Negative residuals for the earlier years mean that the observed recaptures are lower than expected by the model, given the estimated abundance at age for these years and estimated survival rate (and the opposite for the last two years). This means that the 2017 recapture data provide information that the abundance-at-age for the earlier years (2011-2014) are higher than expected by the model, and recent abundances at age are lower.

The residuals plot also illustrate how the number of observations increase for each new year of recapture, which helps to understand why recent recapture years (2016 and 2017) have more influence than older years (e.g. 2012), as illustrated on figure 3.

In order to visualize the information on abundance at age provided by each observation in the RFID tagging data, the equation used to obtain expected recaptures in the model was reversed to obtain an abundance:

$$N = survival \times N_{release} \times N_{scan} / N_{recap},$$

Where *survival* is the estimated survival rate (around 10%), *N_{release}* is the number of fish tagged (of a particular year class in a particular year), *N* is the stock abundance (of the same year class in the same year), *N_{scan}* and *N_{recap}* are the number (of that same year class) scanned and recaptured in a specific year.

The estimates of abundance at age back-calculated from the tag recapture for a given (release) year are compared for the successive recapture years (figure 10). Given that the RFID time

series is still short, the comparison can only be made for a limited number of years, but it appears that the back calculated abundance at age from recent recapture years (2016-2017) indicate higher abundances at age for the years 2011 and 2012 than earlier recapture years and than the model estimates. The opposite is observed for the recent years (2015 and 2016) for which the recent recaptures indicate lower abundances than the model estimate.

Broadly speaking, the recaptures from 2017 (and to a lesser extend 2016) give a perception of abundances-at-age decreasing since 2011 at a stronger rate that indicated by earlier recaptures and estimated by the SAM assessment, as illustrated on figure 11.

Conclusions

The relative influence of the different data sources on the assessment has changed between the 2017 and the 2018 assessments. In the 2017 assessment, the egg survey and the IESSNS both have a strong influence on the assessment, even if the RFID tags have an even larger influence. In the 2018 assessment, the influence of the two surveys is much smaller, and the influence of the RFID increased considerably.

This change in the relative weight of the data sources is related to the inclusion of the recapture data from 2017. The inclusion of this data is responsible for most of the change in perception of the SSB after 2013 between the current and the previous assessment. This data provides information that stock numbers have been going down continuously since 2011, while other data from other recapture years and the surveys indicate a later decrease.

This perception of a data source indicating a faster decline than other data sources is coherent with the pattern observed in the residuals for the recaptures from 2017, with less recaptures than expected (i.e. providing the information that the stock is larger than expected) for the earlier release years (2011-2012) and the opposite for the latest release years. To some extent, the same pattern is visible for the recapture years 2016 and 2015. One possible explanation, already discussed at WGWIDE, could be that the mortality induced by tagging does not only operate immediately after release, but continues through the years, or that a proportion of the tagged fish loose their tags each year. The longer the fish remain at sea after tagging, the lower the concentration of tags in the population.

In addition to this potential bias, the increasing weight of the RFID tagging data can be explained by the fact that the number of data points contained in each recapture year is

increasing as more recent recapture years contain data from more release years : 1 times 12 for the recaptures from 2012, 6 times 12 for the recaptures from 2017.

Potential changes in assessment procedure

- Data selection

The two problems identified (potential bias due to continuous mortality after tagging and increasing number of data point available) could maybe be circumvent by operating a selection within the data available. For each recapture year, the data for only 2 or 3 release years could be used, starting from the first recapture year for which this number of release years is available. This would ensure that the same number of data points is added at every update of the assessment. This would limit the number of years spent at sea before recapture and thereby minimize the effect of the potential bias due to continued tagging mortality. However, potentially informative data would be lost (recaptures from old releases).

- Modelling the tag recaptures as sum across recapture years

Alternatively, a way to use all the tagging recapture data could consist in modelling not the number of fish recaptured each year from a given tagging experiment, but the sum of the successive recaptures on a given cohort.

Currently in the model each recapture event is an observation. A recapture event is fish of a given cohort from a given tagging experiment recaptured in a given year.

$$NRecap_{recY,relY,relA} = S \times NScanned_{recY,relY,relA} \times \frac{NRel_{relY,relA}}{N_{relY,relA}}$$

with :

$NRecap_{recY,relY,relA}$	number recaptured in the year $recY$ from fish released in year $relY$ at age $relA$
$NScanned_{recY,relY,relA}$	number fish scanned in the year $recY$ from fish released in year $relY$ at age $relA$
$NRel_{relY,relA}$	number of fish tagged at age $relA$ in release year $relY$
$N_{relY,relA}$	stock abundance at age $relA$ in release year $relY$
S	survival rate

This means that a given abundance at age $N_{relY,relA}$ in the model is influenced by multiple $NRecap_{recY,relY,relA}$.

As the series is still young, the number of recapture years influencing each $N_{relY,relA}$ (even since the first RFID tagging experiment in 2011) increases with each new year of data (as is visible on the figure 10). If indeed, there is tag loss or continued post tagging mortality, the new recaptures (e.g 2017) will give a perception of the stock abundance of old releases (e.g. 2011) which is higher than older recaptures which would induce an upward revision of these $N_{relY,relA}$ at each new assessment.

In addition, the number of data points contained in each new recapture year is higher than the previous recapture year (until the series is long enough).

In order to avoid these two potential sources of problem, the assessment could model the total number of recaptures from a given experiment and age. This way, there would be only one observation corresponding to each $N_{relY,relA}$, which would be the sum of the recaptures on this specific release year and age across successive recapture years :

$$\begin{aligned} NRecap_{relY,relA} &= \sum_{recY} S \times NScanned_{recY,relY,relA} \times \frac{NRel_{relY,relA}}{N_{relY,relA}} \\ &= S \times \frac{NRel_{relY,relA}}{N_{relY,relA}} \times \sum_{recY} NScanned_{recY,relY,relA} \end{aligned}$$

With each new year of data a term $NScanned_{recY,relY,relA}$ would be added to the expected recaptures.

Fish recaptured after many years (hence recaptured as old age group) would have low number scanned, and hence a small contribution to the $NRecap_{relY,relA}$, which would minimise the effect of any bias due to potential tag loss.

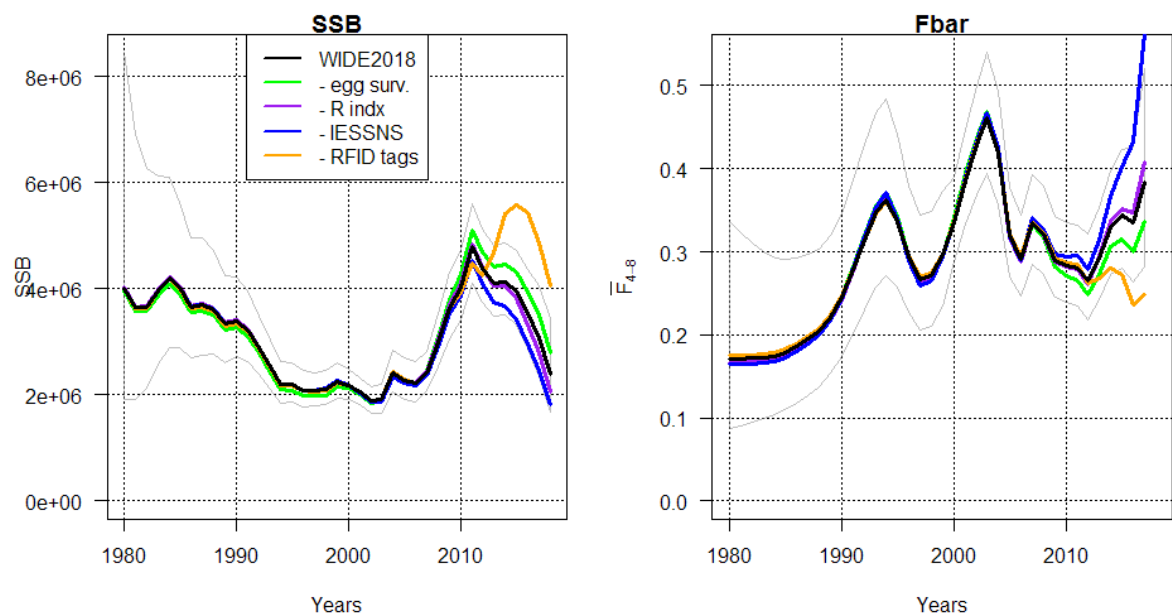


Figure 1 : leave one out runs based on the 2018 assessment

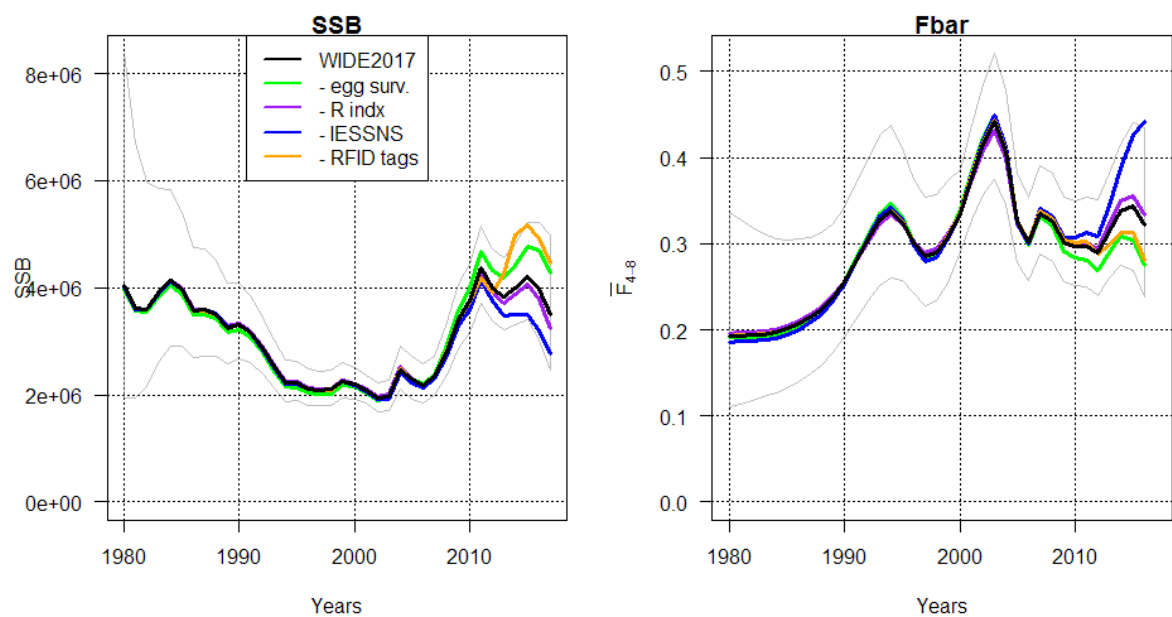


Figure 2 : leave one out run based on the 2017 assessment

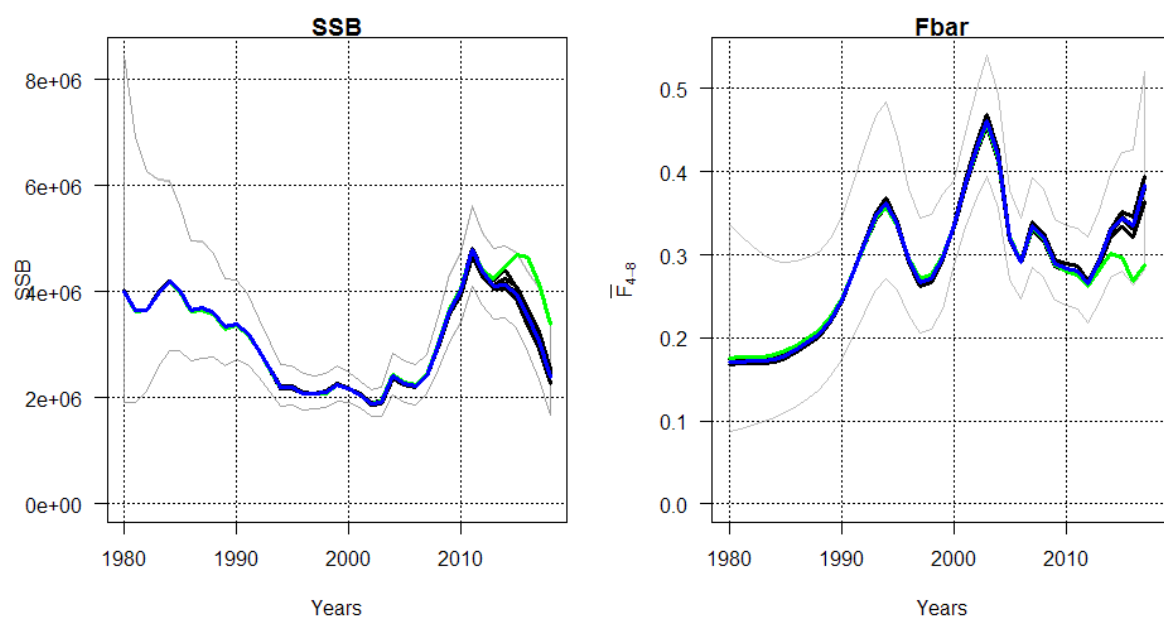


Figure 3 : sensitivity to each of the recapture years for the RFID tags (in blue is the current assessment, and in green removing the 2018 index, highlighted because it had the largest effect)

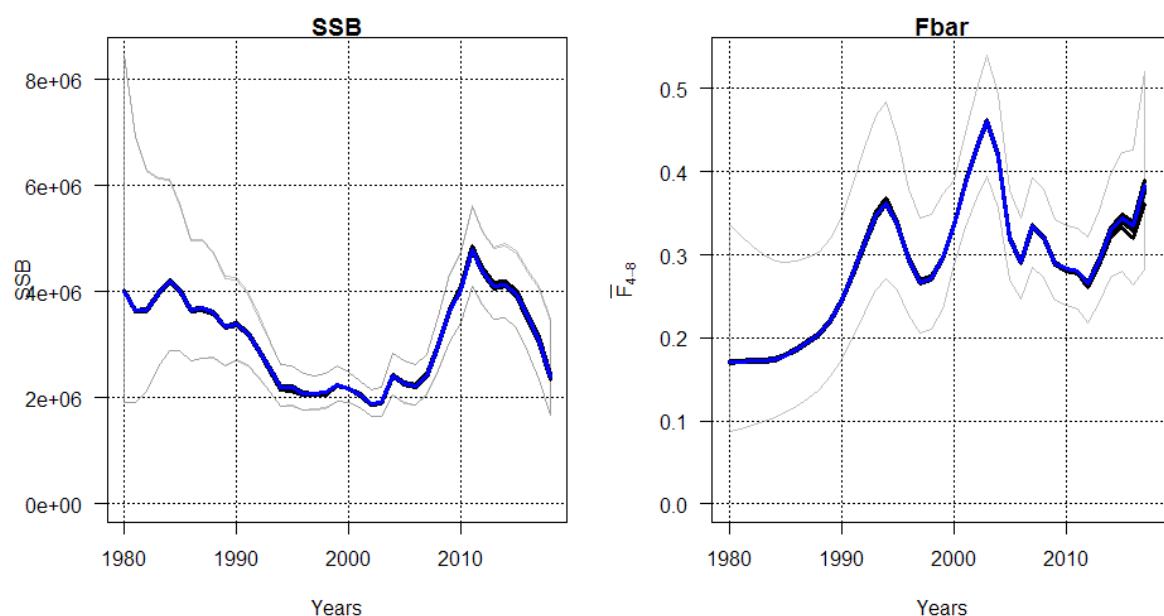


Figure 4. sensitivity to each of the years of the egg survey index (in blue is the current assessment)

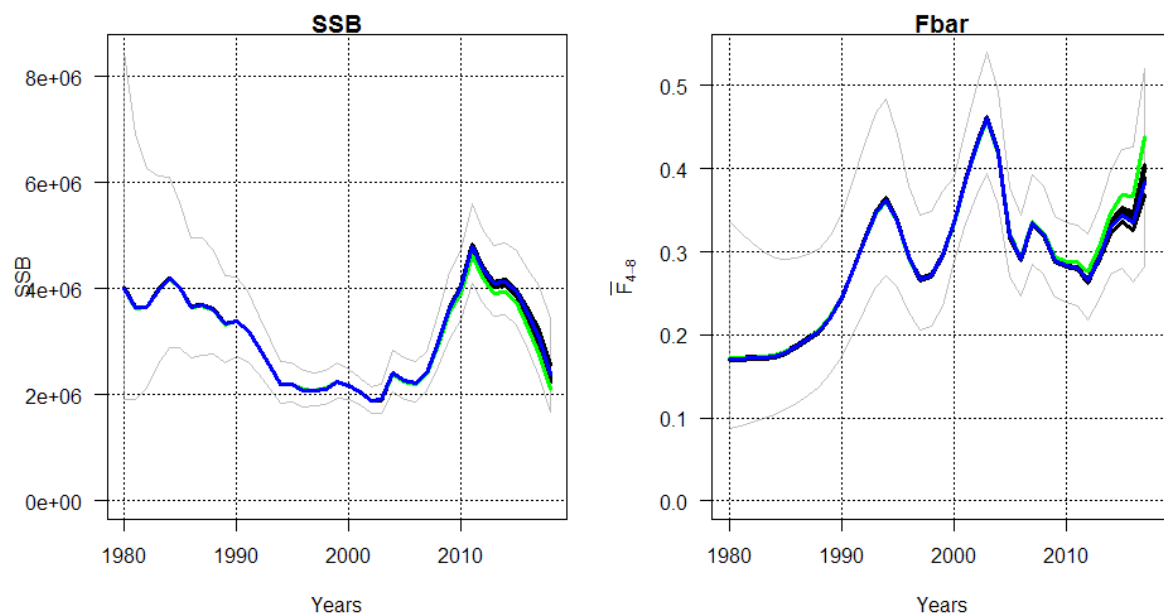


Figure 5. sensitivity to each of the years of the IESSNS index (in blue is the current assessment)

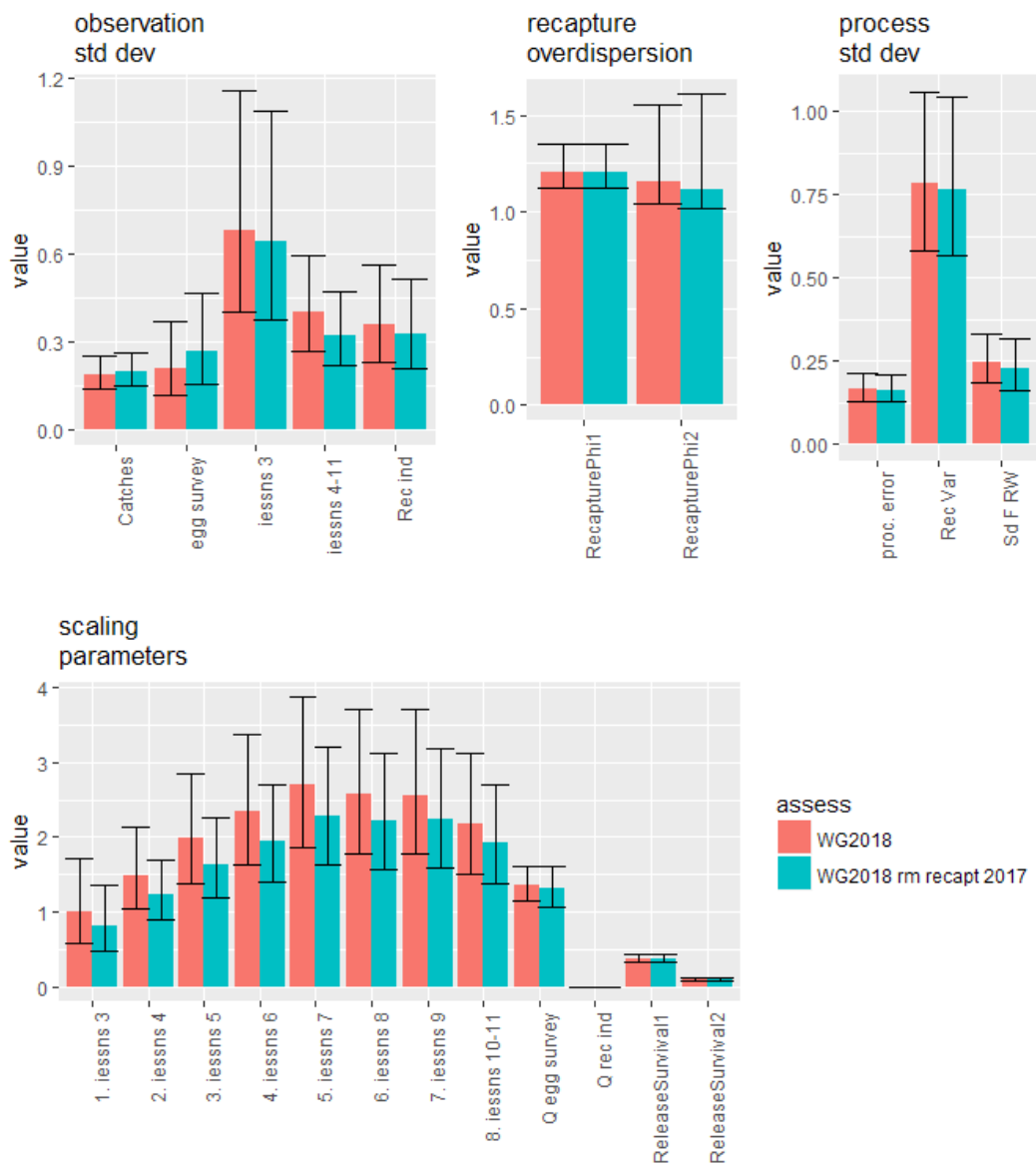


Figure 6. Comparison of model parameters between the WGWIDE 2018 assessment and the same assessment run without the recaptures from 2017.

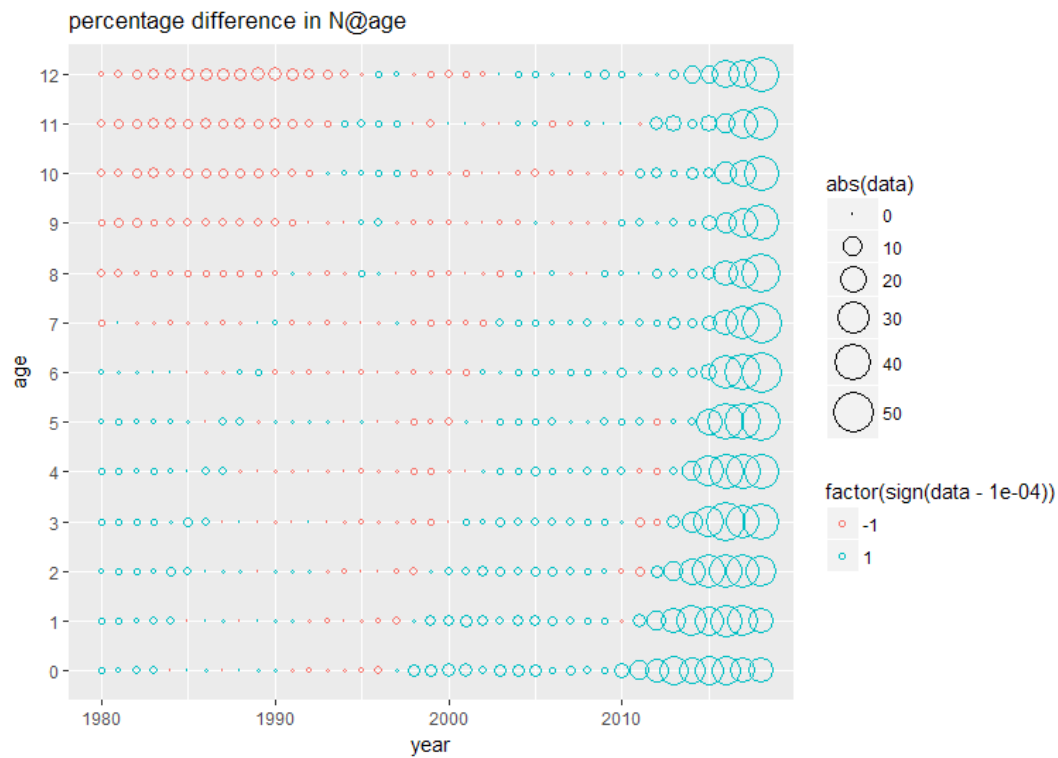


Figure 7. percentage difference in abundance at age between the model excluding the 2017 recaptures and the current assessment

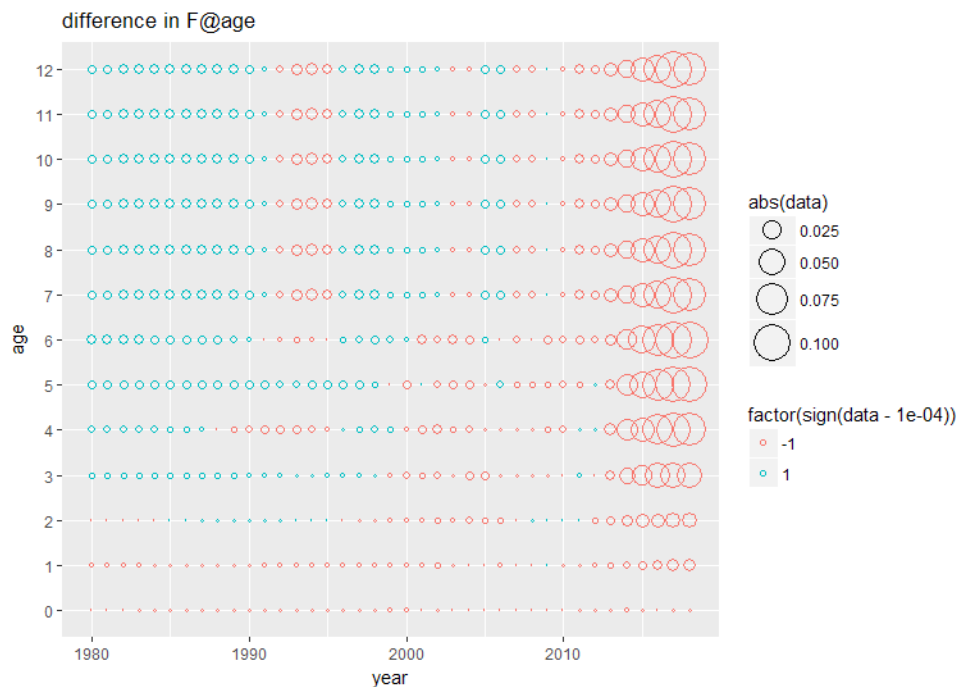


Figure 8 Difference in fishing mortality-at-age between the model excluding the 2017 recaptures and the current assessment (expressed in absolute value)

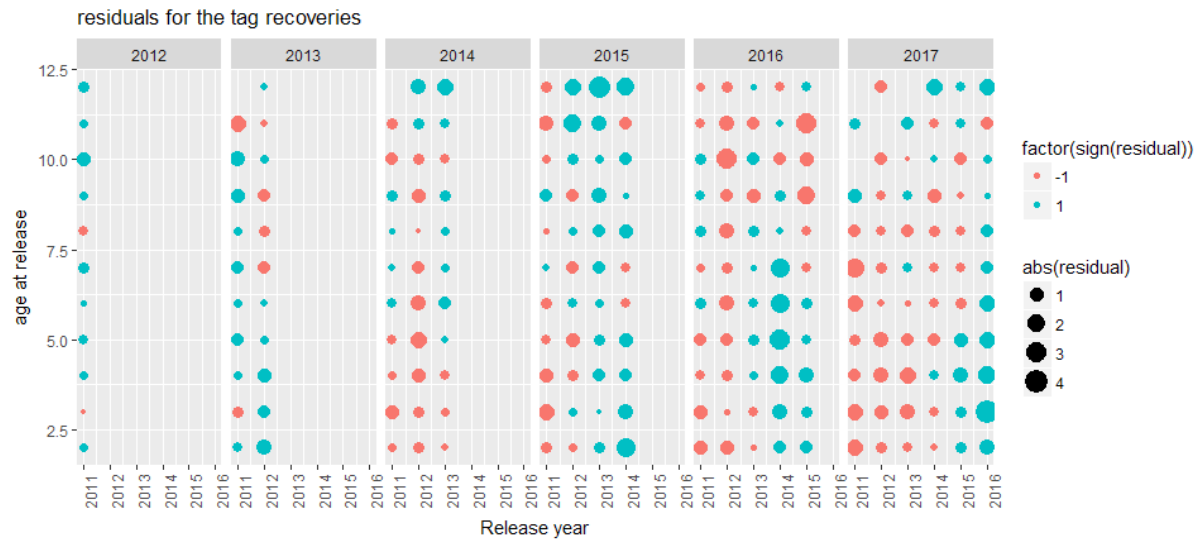


Figure 9 : residuals (one step ahead) for the RFID tag recaptures grouped by recapture years.

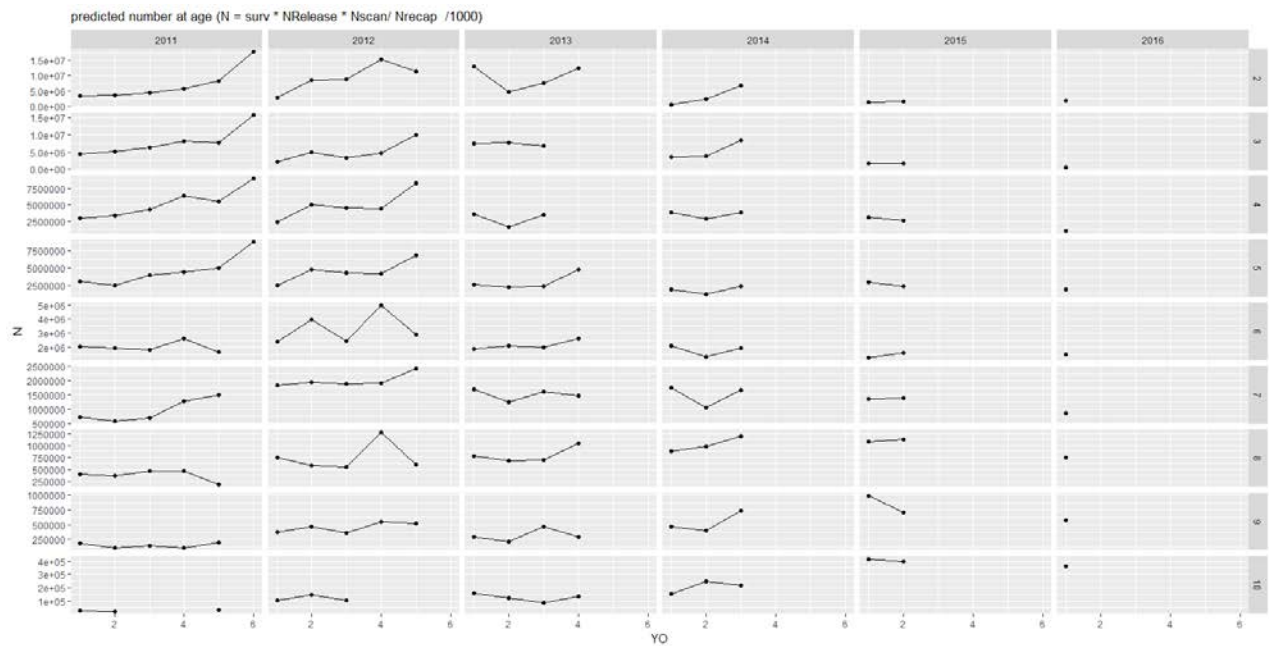


Figure 10. Abundance-at-age (in row) per year (in columns) estimated from the recapture data as a function of the number of years between tagging and recapture.

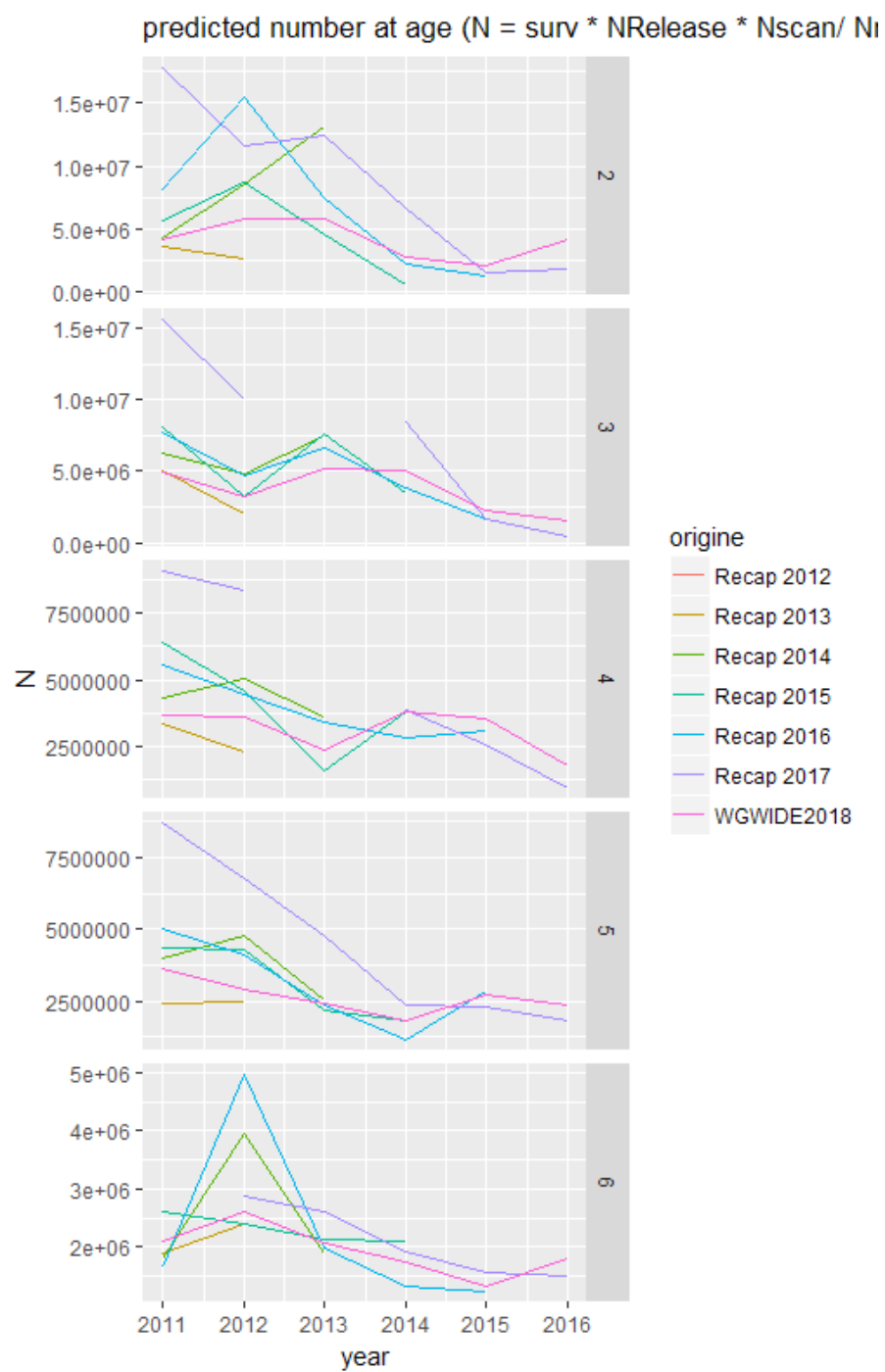


Figure 11. Abundance-at-age over time for a selection of age-groups, same data as figure 10.

Developing abundance at age indices from tag-recapture data

Sindre Vatnehol and Aril Slotte (22.01.2019)

Purpose:

The mark-recapture data is complex, and adds new information of the historical stock size for each new year with recapture data. It is thus difficult to evaluate how informative the data is, the quality, or if any biases have been introduced. We have investigated the possibility to transform this data into indices for abundance at age, similar as the IESSNS survey data, with the purpose to explore trends in the cohorts and in the stock.

Data

We have been using the tag-recapture data found in WGWISE2018.V2 in stockassessment.org, and the code for reading this information. When developing the indices, $r_{a,n}$ has not been rounded to nearest integer as done in the SAM model.

Method:

The indices, and its standard deviation, were made by using a variation of the Bayesian Lincoln-Petersen's mark-recapture equation, where the cohort size at a tag-release year was found by summing the number of tagged fish and the number of scanned fishes for all recapture years;

$$K_{a,y} = \frac{(R_{a,y} - 1)(\sum_{n=1}^N k_{a,n} - 1)}{(\sum_{n=1}^N r_{a,n} - 2)}$$

is the equation for computing the indices, and

$$sd_{K_{a,y}} = \frac{(\sum_{n=1}^N k_{a,n} - \sum_{n=1}^N r_{a,n} + 1)(R_{a,y} - \sum_{n=1}^N r_{a,n} + 1)}{(\sum_{n=1}^N r_{a,n} - 2)(\sum_{n=1}^N r_{a,n} - 3)}$$

is the equation for computing the standard deviation. Here, the subscript a indicate the age/year class, y is the year the fish was tagged, and n is number of years a tagged fish has been out before recaptured (Years out). Also, R is the number of fishes tagged, r is the number of fishes with tags recaptured from the total k fish. K is then the stock number.

The indices and the standard deviation are shown in table 1-2 when $N = 1$ (i.e. when the tags was recaptured one year after the release), and in table 3-4 when $N \rightarrow \infty$. From these tables, the cohorts' trend, ranged from year-classes 1981 to 2014 are seen in figures 1-6, and the consistency of the age data is seen in figure 7. The indices have not been scaled with any tagging mortality or tag-loss.

Total stock biomass index

In figure 8, the total stock biomass index was computed for each survey year. This was computed via,

$$K_y = \frac{(\sum_{a_{min}}^{a_{max}} R_{a,y} - 1)(\sum_{a_{min}}^{a_{max}} \sum_{n=1}^N k_{a,n} - 1)}{(\sum_{a_{min}}^{a_{max}} \sum_{n=1}^N r_{a,n} - 2)}$$

when all the recapture years (Years out) were aggregated into one index, or

$$K_{y,n} = \frac{(\sum_{a_{min}}^{a_{max}} R_{a,y} - 1)(\sum_{a_{min}}^{a_{max}} k_{a,n} - 1)}{(\sum_{a_{min}}^{a_{max}} r_{a,n} - 2)}$$

when computing the index for different recapture years. Here a_{max} and a_{min} are the maximum and minimum age of the fish respectively.

Summary:

The steel-tag series (1983-2004) show rather noisy cohort trends, i.e. year class 1985 figure 1. The RFID tag series (2010-2017), however, have nicer cohort trends. Including all tag data appear to upscale the cohort size in some of the year classes (i.e. Year class 2010 figure 5). Only using data where the tag was recovered 1 year after the release do appeared to have a higher internal consistency between the cohorts than when using all the tag data, figure 7.

In figure 8, the total stock biomass indices are seen for both the RFID tag and the steel tag series for different age intervals and different combination of the Years out. The general trend in the RFID series is that the stock biomass index increases the longer the tags have been out at sea, especially for the younger ages. This trend was however not seen in the steel tag series.

Table 1. Index of abundance at age where only tags being out less than or equal to 1 year have been included. The numbers are in millions of individuals and -1 indicate non valid data.

Year/Age	2	3	4	5	6	7	8	9	10	11
1985	-1	223.24	-1	-1	-1	-1	-1	-1	-1	-1
1986	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1987	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1988	-1	-1	23802.76	-1	-1	-1	-1	-1	-1	-1
1989	9322.7	3670.66	-1	-1	-1	-1	-1	-1	-1	-1
1990	-1	12456.54	-1	-1	-1	-1	-1	-1	-1	-1
1991	-1	-1	25878.3	-1	-1	-1	-1	-1	-1	-1
1992	-1	-1	9511.67	16743.16	-1	-1	-1	-1	-1	-1
1993	20635.32	-1	32837.76	3836.72	2544.18	6179.8	-1	-1	-1	-1
1994	12499.32	9486.99	6365.48	4194.18	5368.21	-1	-1	-1	-1	-1
1995	-1	8196.4	7822.53	-1	7180.4	-1	610.97	-1	-1	-1
1996	-1	8122.35	10091.58	17075.08	2181.91	-1	-1	384.41	-1	-1
1997	4177.31	10368	3898.66	3587.28	2623.51	627.9	712.79	405.25	817.61	57.95
1998	-1	1739.18	6720.56	4759.48	3656.87	839.16	2402.78	330.08	-1	-1
1999	-1	10238.91	-1	2440.99	2078.03	-1	-1	-1	-1	-1
2000	-1	-1	4021.9	4932.87	7268.18	-1	990.86	-1	-1	-1
2001	-1	3272.17	2239.61	3657.2	725.05	985.34	-1	-1	-1	-1
2002	2823.01	8849.95	8916.04	8493.22	5250.31	549.43	1195.52	822.87	-1	298.1
2003	26058.14	622.22	4418.41	5194.31	-1	-1	-1	-1	-1	-1
2004	25661.07	73432.35	2089.17	-1	1626.94	-1	1146.75	-1	-1	-1
2005	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2006	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2007	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2008	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2009	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2010	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2011	75547.94	54177.34	45305.6	37924.18	28335.38	12148.55	7137	2473.58	-1	-1
2012	51027	30610.6	25582.16	27056.2	25722.67	20578.31	8095.48	4990.26	1556.86	846.32
2013	777462.9	-1	79124.13	27481.56	20257.43	18437.43	8056.81	3179.19	1819.65	789.29
2014	9178.7	41687.88	39591.99	18388.8	20580.4	17161.52	8627.57	4856.79	1612.54	1012.64
2015	40127.01	20338.74	31191.86	28348.88	12287.57	13569.97	10716.74	9787.25	4230.15	3779.56
2016	25532.74	4432.58	9511.1	17872.4	14875.87	8536.64	7331.73	5720.74	3589.84	2021.15

Table 2. Standard deviation of the index of abundance at age where only tags being out less than or equal to 1 year have been included. The numbers are in millions of individuals. The numbers are in millions of individuals and -1 indicate non valid data.

Year/Age	2	3	4	5	6	7	8	9	10	11
1985	-1	223.02	-1	-1	-1	-1	-1	-1	-1	-1
1986	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1987	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1988	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1989	4657.37	1222.5	-1	-1	-1	-1	-1	-1	-1	-1
1990	-1	4150.53	-1	-1	-1	-1	-1	-1	-1	-1
1991	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1992	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1993	-1	-1	-1	1277.3	423.21	-1	-1	-1	-1	-1
1994	-1	4740.5	6361.57	1047.6	-1	-1	-1	-1	-1	-1
1995	-1	1364.67	3908.66	-1	-1	-1	304.91	-1	-1	-1
1996	-1	2027.72	3360.76	-1	1089.37	-1	-1	-1	-1	-1
1997	2085.63	2070.9	133.99	238.61	523.91	89.26	118.4	202.21	-1	19.02
1998	-1	346.24	1342.26	593.91	730.29	75.99	-1	329.26	-1	-1
1999	-1	10225.8	-1	347.59	345.18	-1	-1	-1	-1	-1
2000	-1	-1	2005.4	-1	-1	-1	-1	-1	-1	-1
2001	-1	465.94	171.69	456.16	65.51	98.06	-1	-1	-1	-1
2002	-1	802.79	2226.14	8484.06	1747.46	136.19	1191.65	409.8	-1	-1
2003	8677.27	205.9	550.68	5187.41	-1	-1	-1	-1	-1	-1
2004	25632.84	-1	-1	-1	1624.63	-1	-1	-1	-1	-1
2005	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2006	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2007	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2008	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2009	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2010	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2011	-1	10150.61	16454.3	4495.47	8850.06	4944.18	6439.17	1255.58	-1	-1
2012	24513.91	19436.71	1773.57	1801.62	1503.95	1417.49	842.21	1461.41	963.58	567.86
2013	-1	-1	-1	2772.71	897.16	901.38	188.81	157.22	164.77	204.78
2014	2237.27	3441.2	1298.5	293.38	167.91	139.28	84.08	113.99	46.01	102.14
2015	-1	5455.58	1198.08	565.8	236.67	157.94	190.32	162.02	93.88	475.58
2016	8320.76	162.6	542.71	268.02	168.94	171.85	83.95	67.31	48.7	25.22

Table 3. Index of abundance at age where all tag data have been included. The numbers are in millions of individuals and -1 indicate non valid data.

Year/age	2	3	4	5	6	7	8	9	10	11
1983	10199.3	-1	-1	-1	-1	-1	-1	-1	-1	-1
1984	-1	40526.22	-1	-1	-1	-1	-1	-1	-1	-1
1985	-1	1750.99	18398.98	-1	-1	-1	-1	-1	-1	-1
1986	22184.93	-1	852.15	17563.12	-1	-1	-1	-1	-1	-1
1987	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1988	6893.93	6249.9	13857.9	-1	1862.19	2617.97	-1	-1	-1	-1
1989	9743.69	7144.04	9736.52	54696.7	-1	-1	-1	-1	-1	-1
1990	5387.52	14148.92	10320.67	1672.69	-1	-1	-1	-1	-1	-1
1991	5884.4	8423.86	12411.31	2704.66	10877.46	12343.76	667.35	-1	-1	-1
1992	3572.65	10613.74	5355.69	6791.05	16953.37	1100.8	3881.71	-1	-1	-1
1993	7119.6	6736.29	6498.45	3137.46	3348.87	1774.49	1720.45	4245.28	-1	-1
1994	9668.52	5709.11	2932.03	3500.71	1490.13	3374.2	3007.81	598.2	206.24	-1
1995	3742.69	6284.51	3616.57	1415.13	3681.88	801.09	637.18	697.91	1145.75	-1
1996	4661.1	6792.03	5097.95	2275.99	1508.27	1565.24	471.7	138.57	795.72	363.22
1997	5142.96	5173.32	4375.93	2950.69	2090.93	659.48	952.05	589.7	388.27	66.93
1998	982.54	1805.81	4218.5	4950.64	2164.74	1301.97	1761.91	640.81	1380.27	-1
1999	1118.36	14295.42	4543.9	3669.67	1914.35	4050.78	1597.58	2322.43	-1	-1
2000	-1	4925.84	5915.09	1720.83	2711.15	2123.64	619.14	734.25	-1	-1
2001	3554.19	3859.09	2480.33	3750.6	1010.06	1749.54	3478.61	5911.97	583.63	-1
2002	1770.11	7262.76	4787.55	4733.18	7245.1	521.75	2098.73	886.55	-1	422.4
2003	23332.59	1403.95	4726.92	5380.54	-1	-1	-1	-1	-1	-1
2004	20580.21	32925	4868.02	6587.03	1074.3	-1	2007.81	-1	-1	-1
2005	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2006	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2007	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2008	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2009	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2010	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2011	58374.08	74065.21	54770.84	40497.96	23543.22	8380.53	4648.18	2211.25	742.67	-1
2012	87515.92	41154.21	43249.33	40816.23	30770.55	18942.65	6882.45	4053.04	1775.63	593.59
2013	80240.14	147962.42	30355.55	25686.14	20562.71	15021.6	7381.52	3017.83	1374.72	584.21
2014	23979.33	45294.97	33696.11	16180.65	17025.82	14156.2	9434.23	4897.87	2003.14	921.54
2015	16675.13	16640.01	27937.23	25338.18	13782.12	13484.28	10865.28	8501.9	4104.75	2439.57
2016	25532.74	4432.58	9511.1	17872.4	14875.87	8536.64	7331.73	5720.74	3589.84	2021.15

Table 3. Standard deviation of the index of abundance at age where all tags have been included. The numbers are in millions of individuals. The numbers are in millions of individuals and -1 indicate non valid data.

Year/age	2	3	4	5	6	7	8	9	10	11
1983	10186.98	-1	-1	-1	-1	-1	-1	-1	-1	-1
1984	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1985	-1	349.2	3677.33	-1	-1	-1	-1	-1	-1	-1
1986	-1	-1	282.47	5851.14	-1	-1	-1	-1	-1	-1
1987	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1988	2291.24	779.6	1977.02	-1	1858.62	522.7	-1	-1	-1	-1
1989	882.77	647.78	3241.65	-1	-1	-1	-1	-1	-1	-1
1990	335.16	830.8	5152.68	554.48	-1	-1	-1	-1	-1	-1
1991	344.04	840.18	1548.87	384.54	-1	-1	-1	-1	-1	-1
1992	1180.44	1763.75	1067.5	1129.24	-1	156.4	1938	-1	-1	-1
1993	590.08	745.74	462.55	260.39	333.87	294.68	858.36	-1	-1	-1
1994	1202.54	236.63	153.37	151.55	114.01	561.49	3004.58	198.8	68.3	-1
1995	147.14	141.89	123.88	66.72	458.82	61.17	79.2	231.8	-1	-1
1996	209.33	259.24	187.59	94.05	99.77	194.55	77.91	19.45	-1	-1
1997	299.89	119.16	52.87	52.27	86.57	32.55	72.75	97.82	128.89	16.41
1998	154.73	65.45	116.18	175.83	76.76	61.56	585.49	159.22	-1	-1
1999	540.12	3562.42	410.29	214.3	94.76	1346.89	795.8	-1	-1	-1
2000	-1	1631.27	1176.49	283.29	448.73	703.8	153.18	730.02	-1	-1
2001	437.68	191.32	66.42	162.13	43.37	124.14	867.56	-1	290.43	-1
2002	292.27	240.73	280.32	943.57	1807.86	64.22	2091.94	220.17	-1	-1
2003	2909.94	347.8	361.95	2684.92	-1	-1	-1	-1	-1	-1
2004	6844.98	10966.09	-1	3289.21	267.62	-1	-1	-1	-1	-1
2005	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2006	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2007	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2008	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2009	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2010	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2011	2139.99	1296.68	2106.26	870.98	1045.12	444.32	510.74	291.87	415.68	-1
2012	5980.16	1973.83	593.81	578.65	370.16	291.33	152.42	173.69	151.4	53.69
2013	6580.63	59244.25	3404.02	628.13	266.23	207.99	50.3	40.05	31.39	29.24
2014	2035.64	1278.22	356.58	87.74	50.53	45.31	41.82	53.43	27.09	45.74
2015	7280.41	1548.07	509.34	253.95	140	84.9	105.73	76.16	49.35	140.75
2016	8320.76	162.6	542.71	268.02	168.94	171.85	83.95	67.31	48.7	25.22

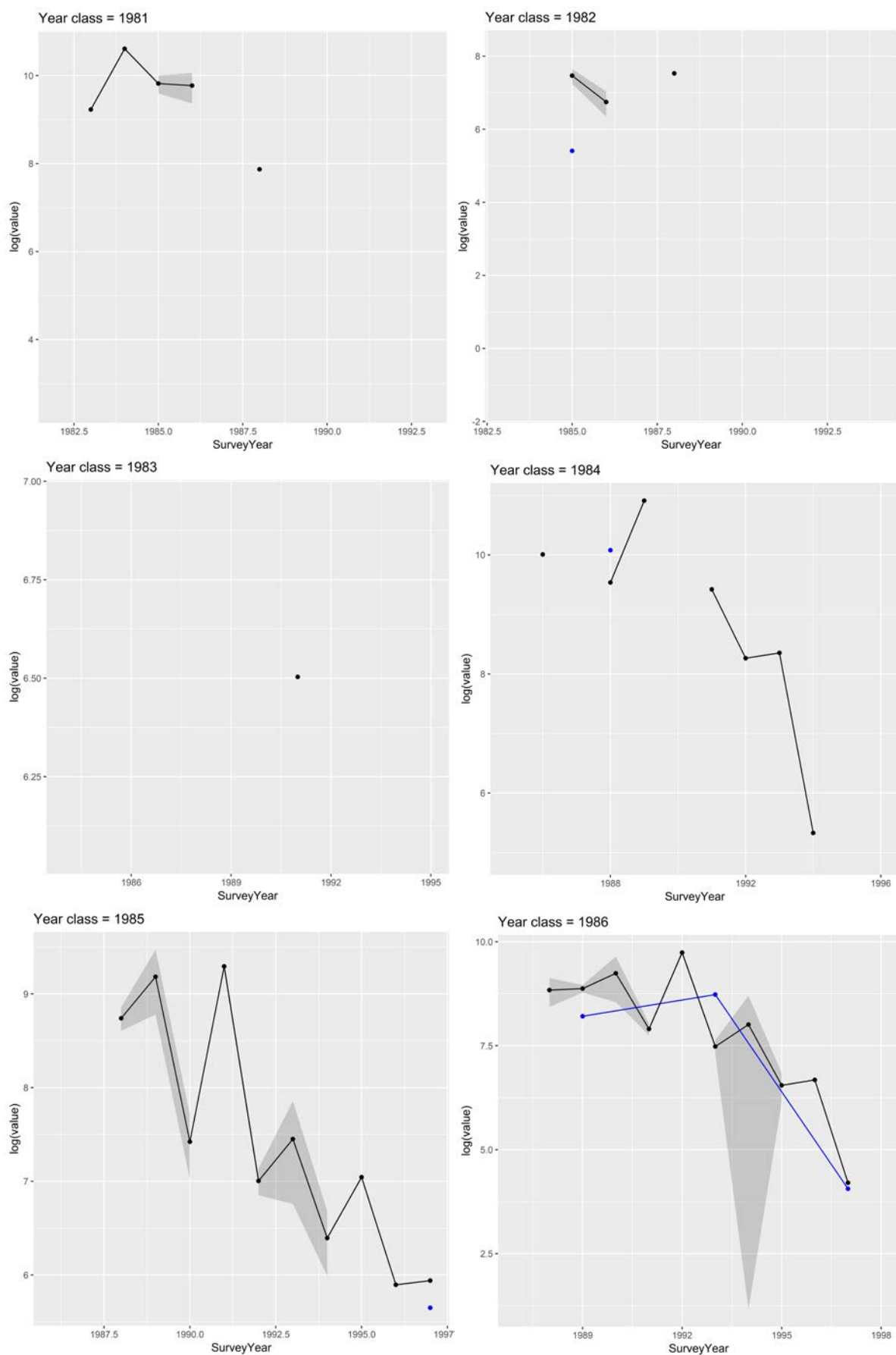


Figure 1: Cohort lines for year classes 1981-1986, when using all tagging data (black line) or when only using those recapture less than 1 year after release.

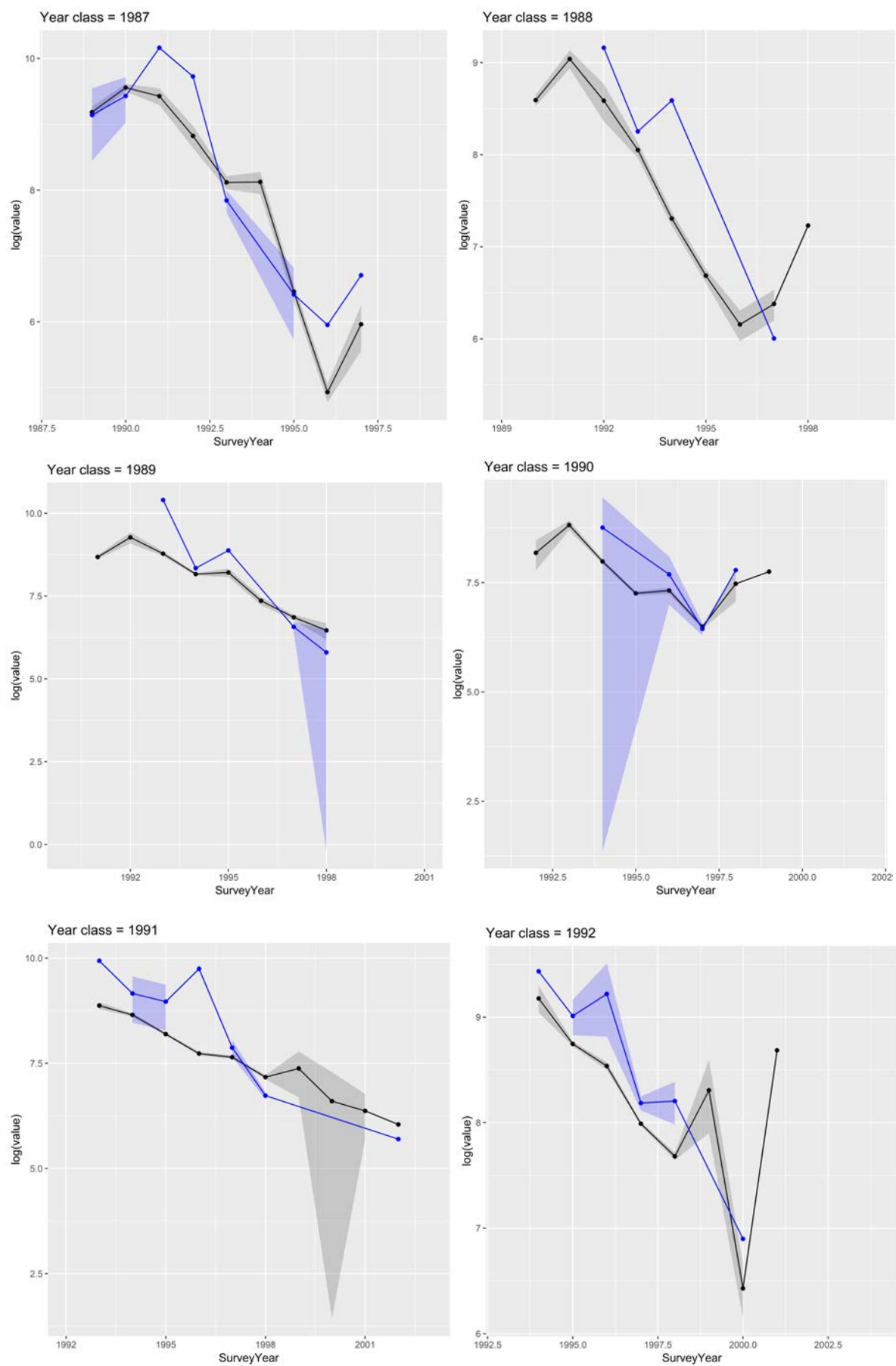


Figure 2: Cohort lines for year classes 1987-1992, when using all tagging data (black line) or when only using those recaptures less than 1 year after release.

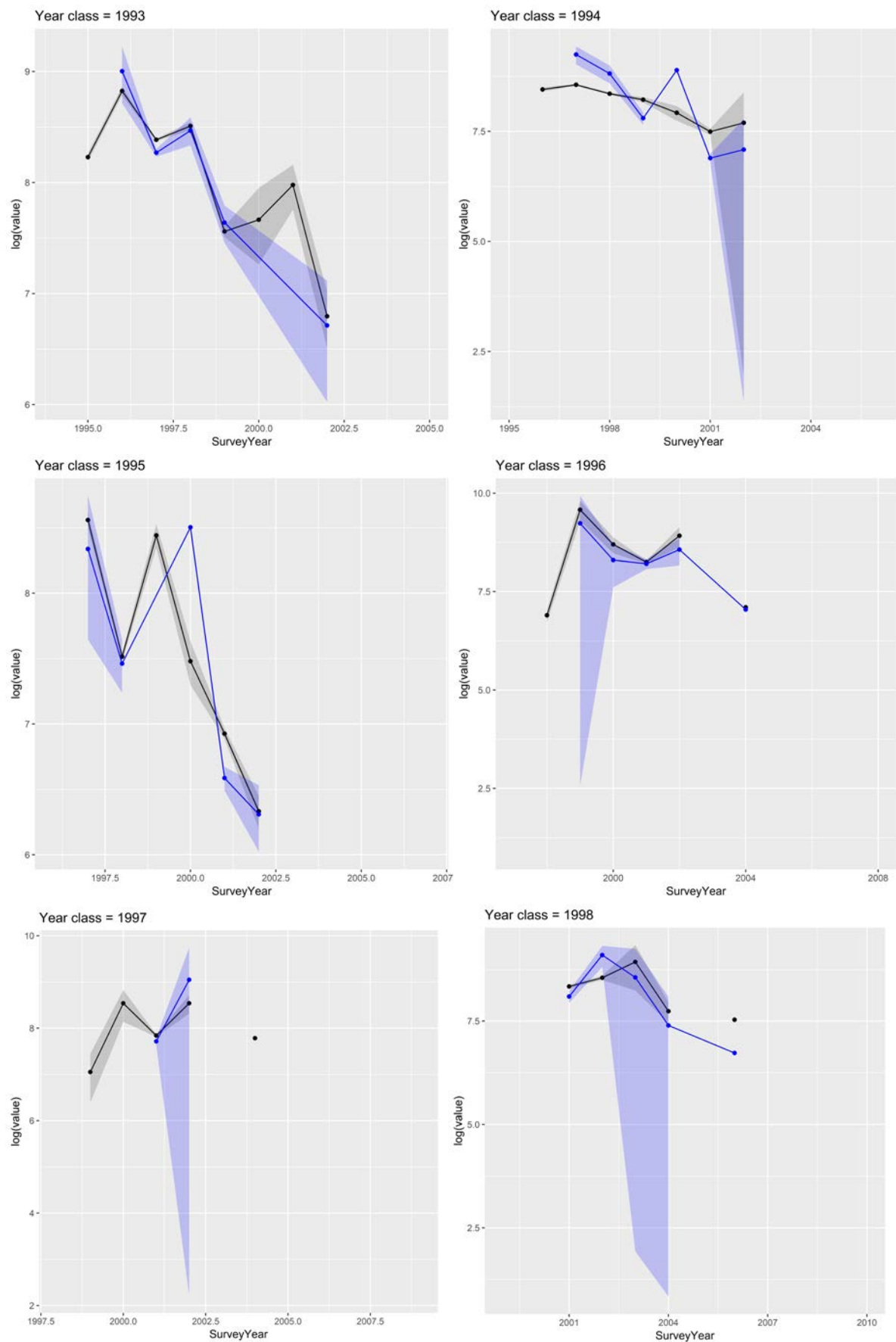


Figure 3: Cohort lines for year classes 1993-1998, when using all tagging data (black line) or when only using those recaptures less than 1 year after release (blue).

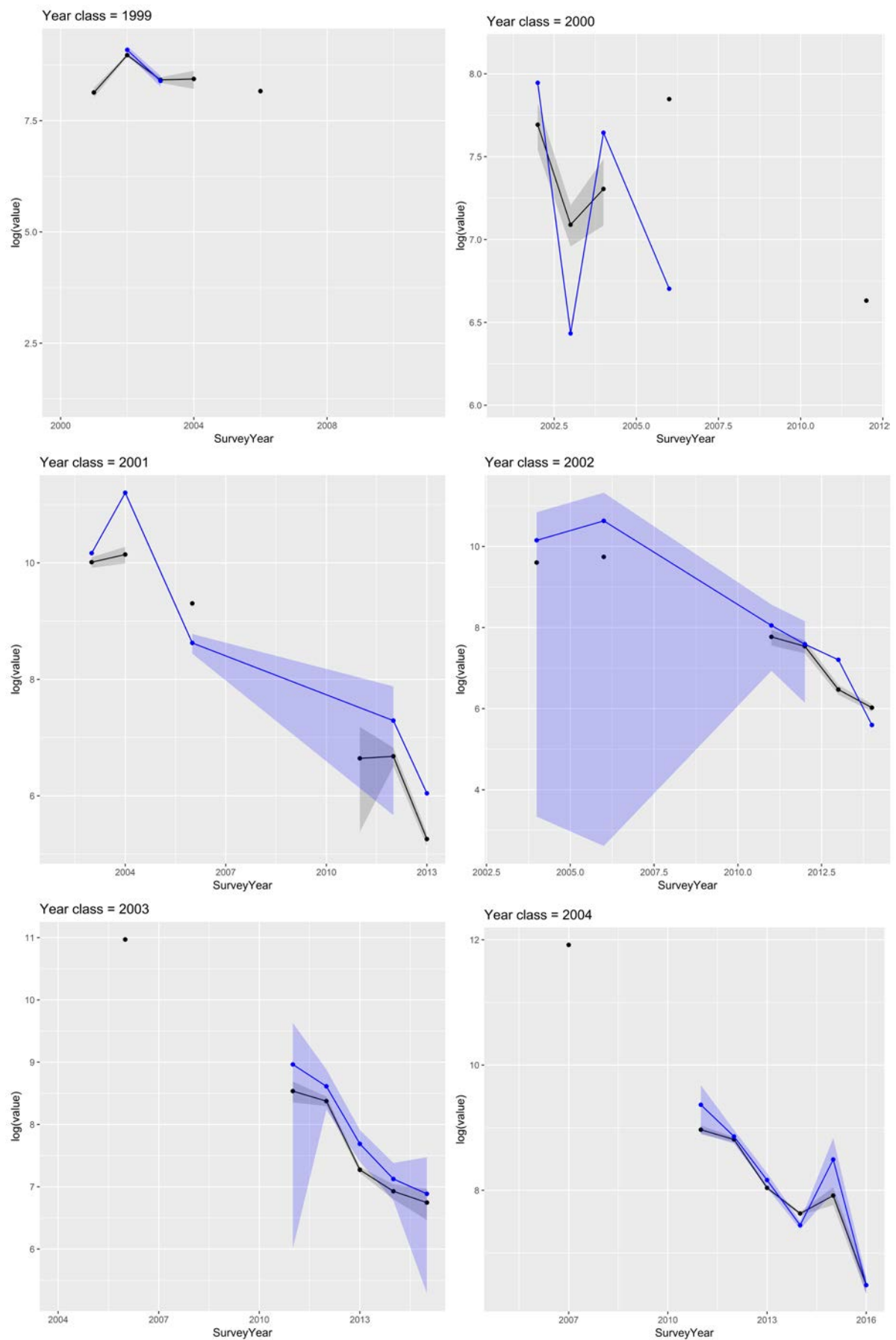


Figure 4: Cohort lines for year classes 1999-2004, when using all tagging data (black line) or when only using those recaptures less than 1 year after release (blue).

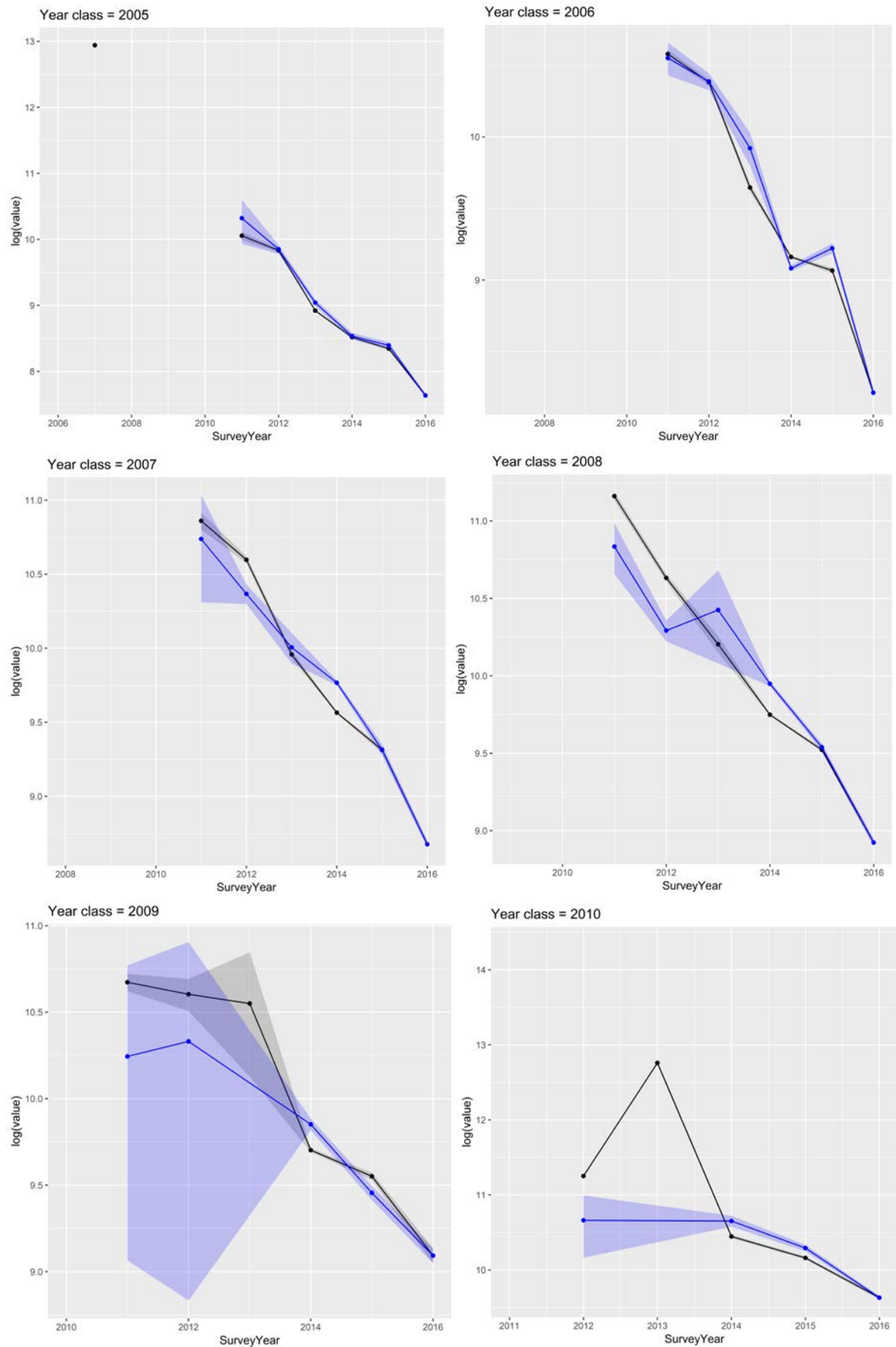


Figure 5: Cohort lines for year classes 2005-2010, when using all tagging data (black line) or when only using those recaptures less than 1 year after release (blue).

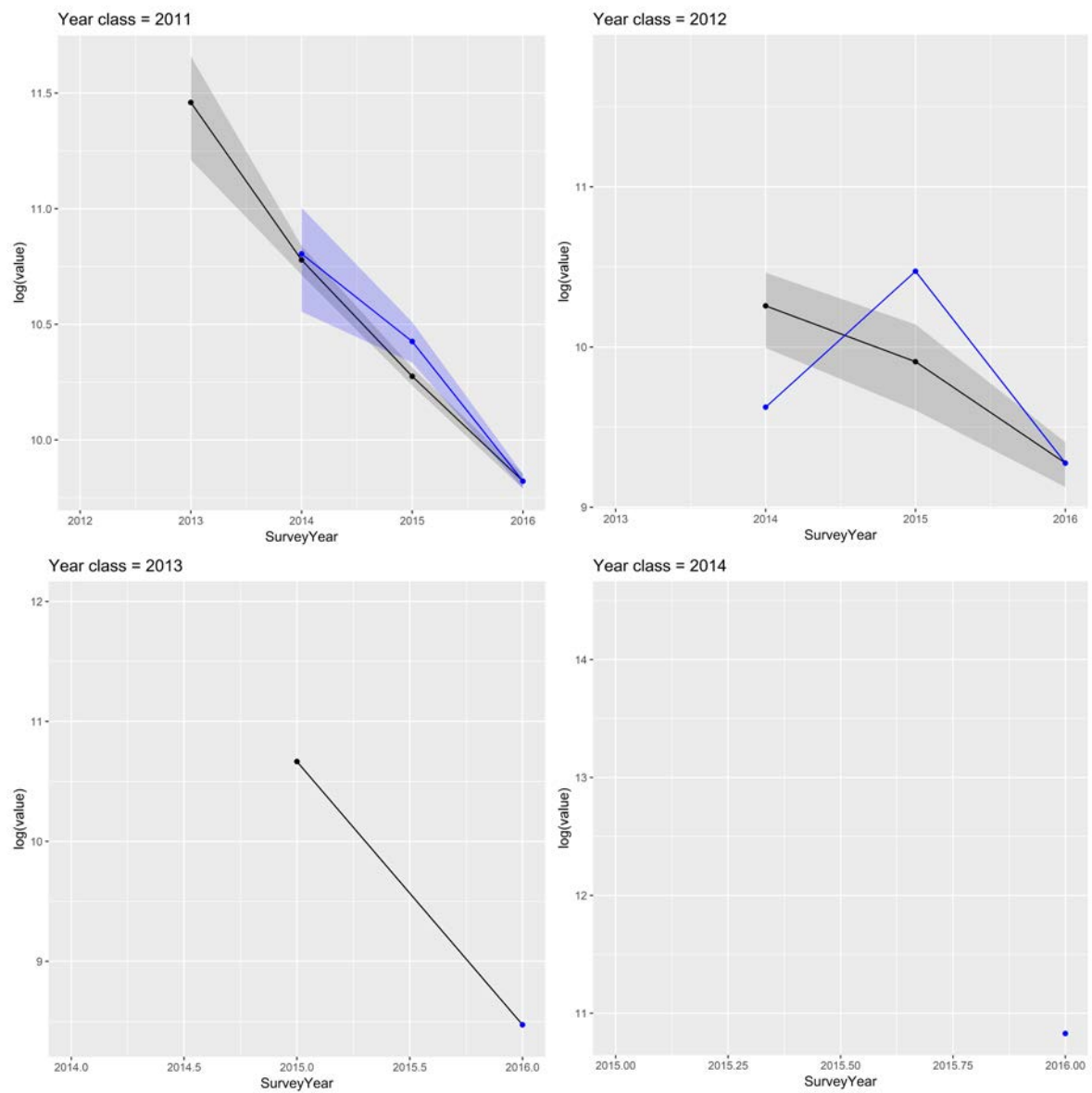


Figure 6: Cohort lines for year classes 2011-2014, when using all tagging data (black line) or when only using those recaptures less than 1 year after release (blue).

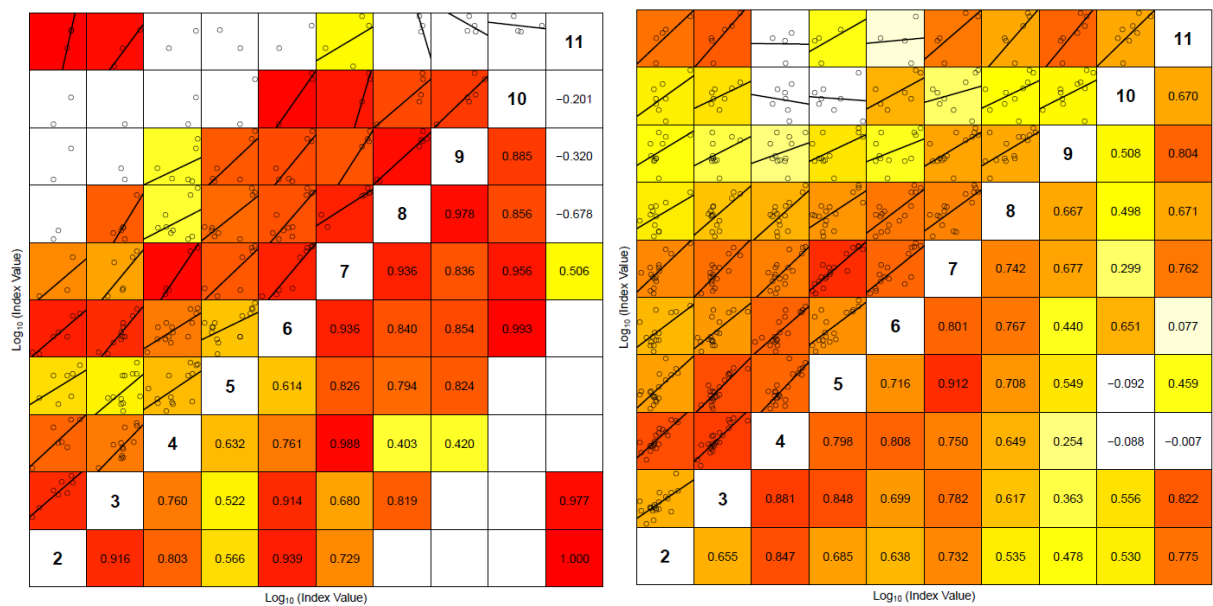


Figure 7: The internal consistency of between the cohorts using only tags that was recaptured one year after release (left) and when using all tags (right)

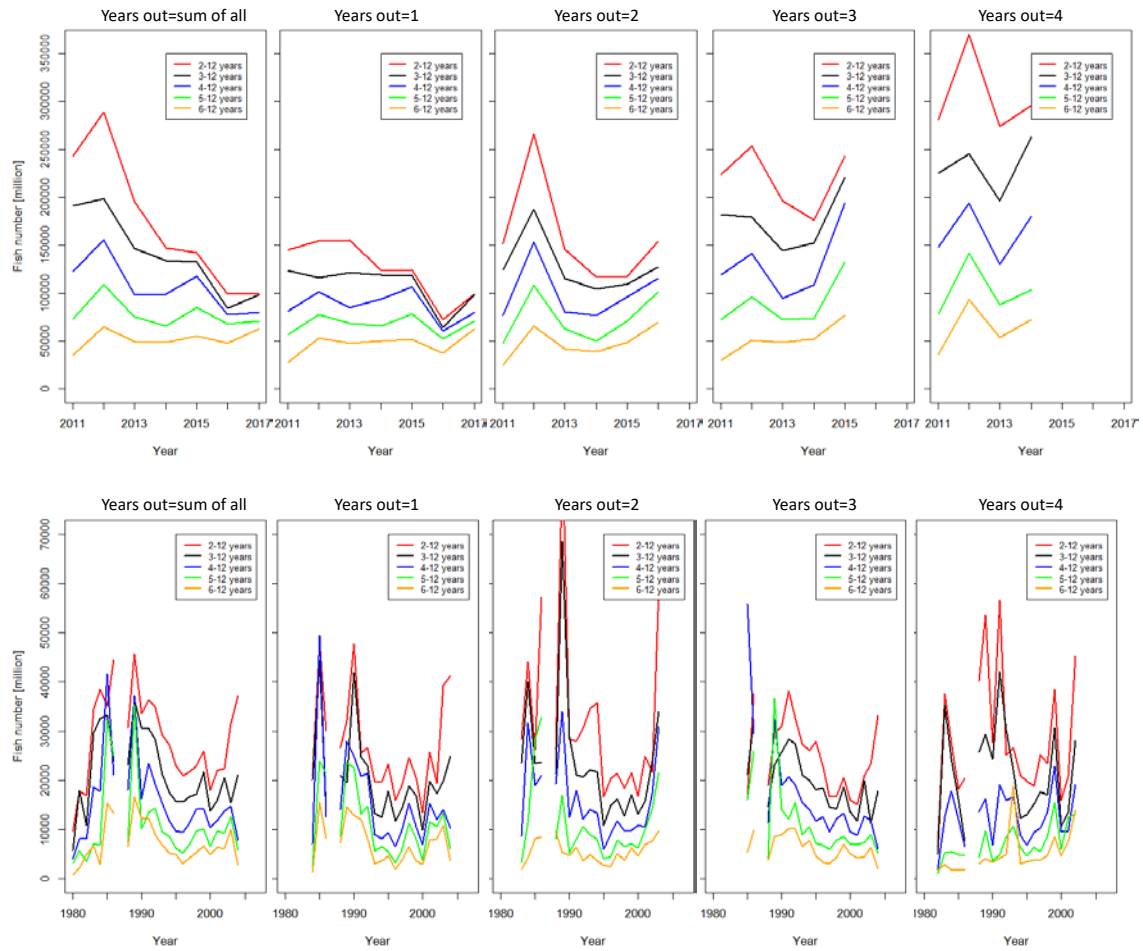


Figure 8: Total stock biomass index using the RFID series (upper) and the Steel series (lower). The stock biomass index where computed using the ages 2-12 (red), 3-12 (black), 4-12 (blue), 5-12 (green) and 6-12 (orange). For the left figures, data from all recapture years are used, while data only from years out equal 1, 2, 3 and 4 are seen for the subsequent figures.

Truncating the catch time series for Mackerel

Claus R. Sparrevohn

Background and purpose

At the WKNAMMM meeting in 2013 (ICES, 2013a) historical unaccounted landings was suggested as one of the potential sources leading to an unstable and possible biased assessment. On basis of that the North East Atlantic Fisheries Commission (NEAFC) send a request in 2013, asking ICES to evaluate to what extend the stock size could be underestimated as a result of unaccounted catches. WGWIDE (2013b and c) did respond to that and did find that unallocated catches and specially varying proportion of unallocated catches during different periods, have the potential to create a bias in both the stock size level and stock size trajectory. WGWIDE did also conclude that the basis for approximating the quantity of unallocated catches was lacking, and therefore did not find any possibility to alter the timeseries of historical ICES catches.

Based on this it was decided in the 2014 benchmark to truncate the catch timeseries. This was done by down-weighting catches prior to 2000. It was also decided in the Long-Term Management Plan Evaluation in 2015 to only use recruitment pairs from 1990 and onwards, a decision which has been follow in subsequent evaluations.

The purpose of this WD is to revisit the decision of choosing 2000 as the breakpoint, and to run sensitivity test comparing the estimated SSB, R and F between runs with different years truncated (down weighted, DW). All runs are based on assessment used to provide catch advice for 2019 (WGWIDE2018.V2) and catches was truncated by down weighting.

Results

It was possible to make the assessment converge also when all catches prior to 2017 was down-weighted, however the uncertainty in the assessment did increase considerable which can be seen on figure 1, where the baserun is compared to a run where catches prior to 2017 is down weighted.

On figure 2, 3 and 4 the effect of truncating the timeseries on the SSB, F and R is shown. The overall trend is that there is minor effect on including catch data from 1990 or 1995 and onwards up until 2000. I.e. including older data might not matter that much, However, excluding additional years from 2000 does create a trend in both SSB (figure 2), fishing mortality F (figure 3) and recruitment R (figure 4).

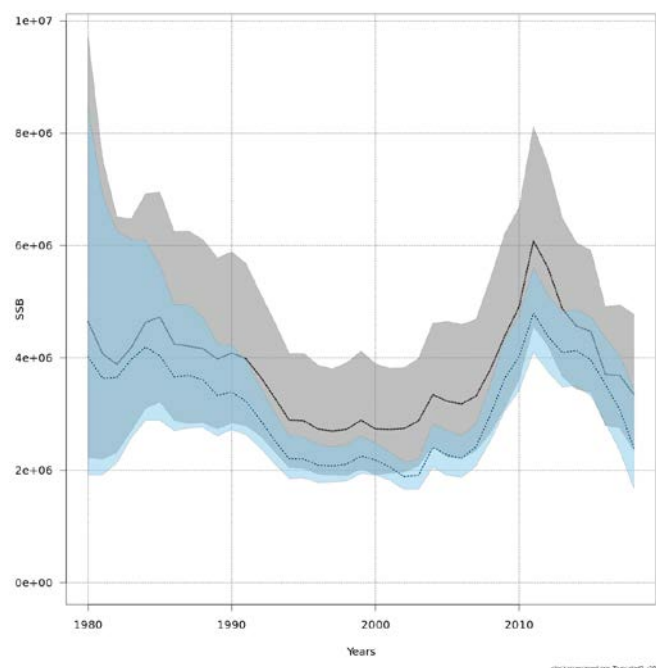


Figure 1. Output comparing the baseline run (WGWISE2018.V2) in blue and a version where catch data prior to 2017 are down weighted.

Conclusion

To what extend the changing trajectories of SSB, F and R reflects the exclusion of specific years (the later period) from the catch time series or if it reflects the time series gets shorter and therefore the catch data itself are being given less weight by the model is unclear, but given the gradual scaling of eg. SSB when expanding the truncation from 2010 to 2017, the latter might be the most likely explanation.

In 2014 truncating the timeseries from 2000 was decided although, truncating from 2005 would be more aligned with the four “misreporting time periods” identified during WGWISE 2013 (ICES 2013c). Here the “golden age” without unallocated catches are set to be the period after 2005. The choice does matter as it can be seen by comparing the base run with the “DW>2005” run. It might not have a large effect on the SSB in 2018 but the stock recruitment pairs are most likely changed and hence the Fmsy value etc. could also change. It might be worth to consider truncating the timeseries further, such that catch data prior to 2005 is excluded. This would be more in accordance ICES 2013c.

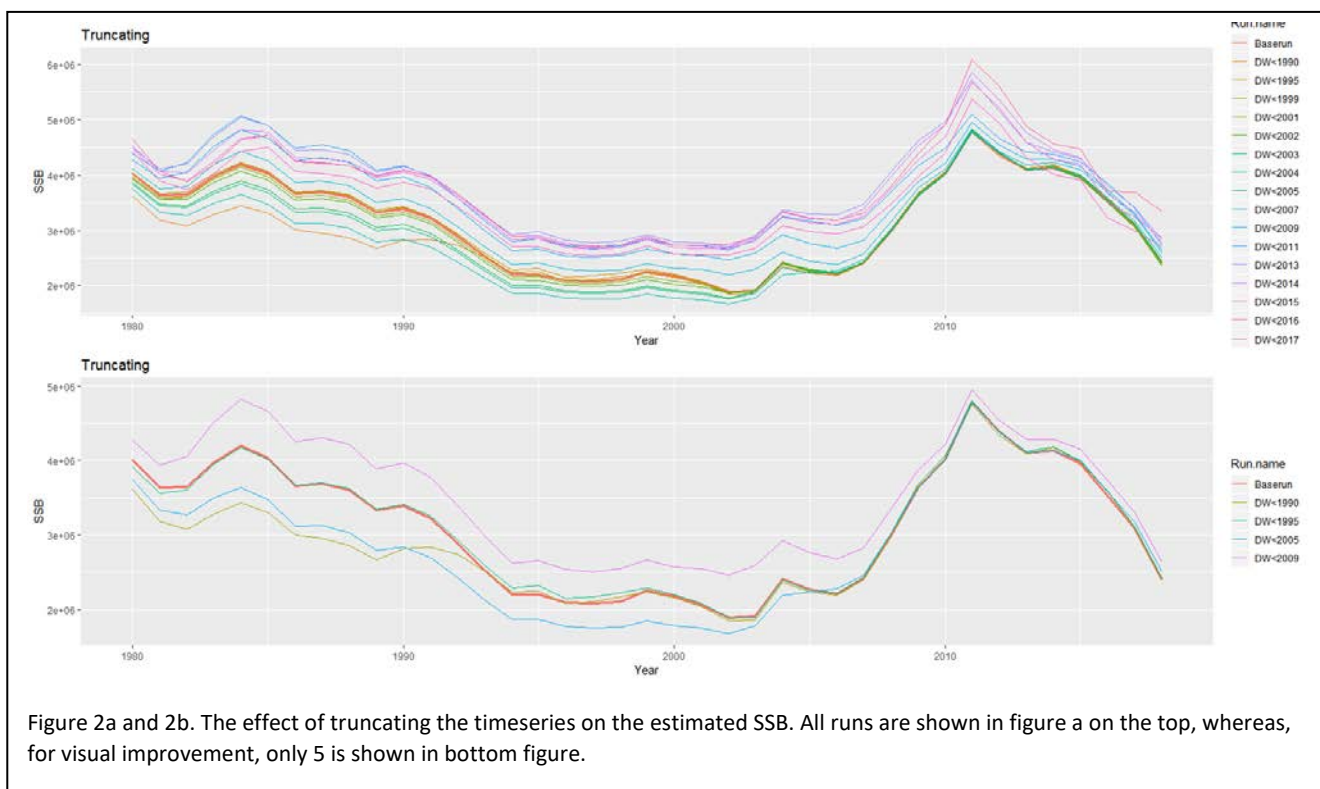


Figure 2a and 2b. The effect of truncating the timeseries on the estimated SSB. All runs are shown in figure a on the top, whereas, for visual improvement, only 5 is shown in bottom figure.

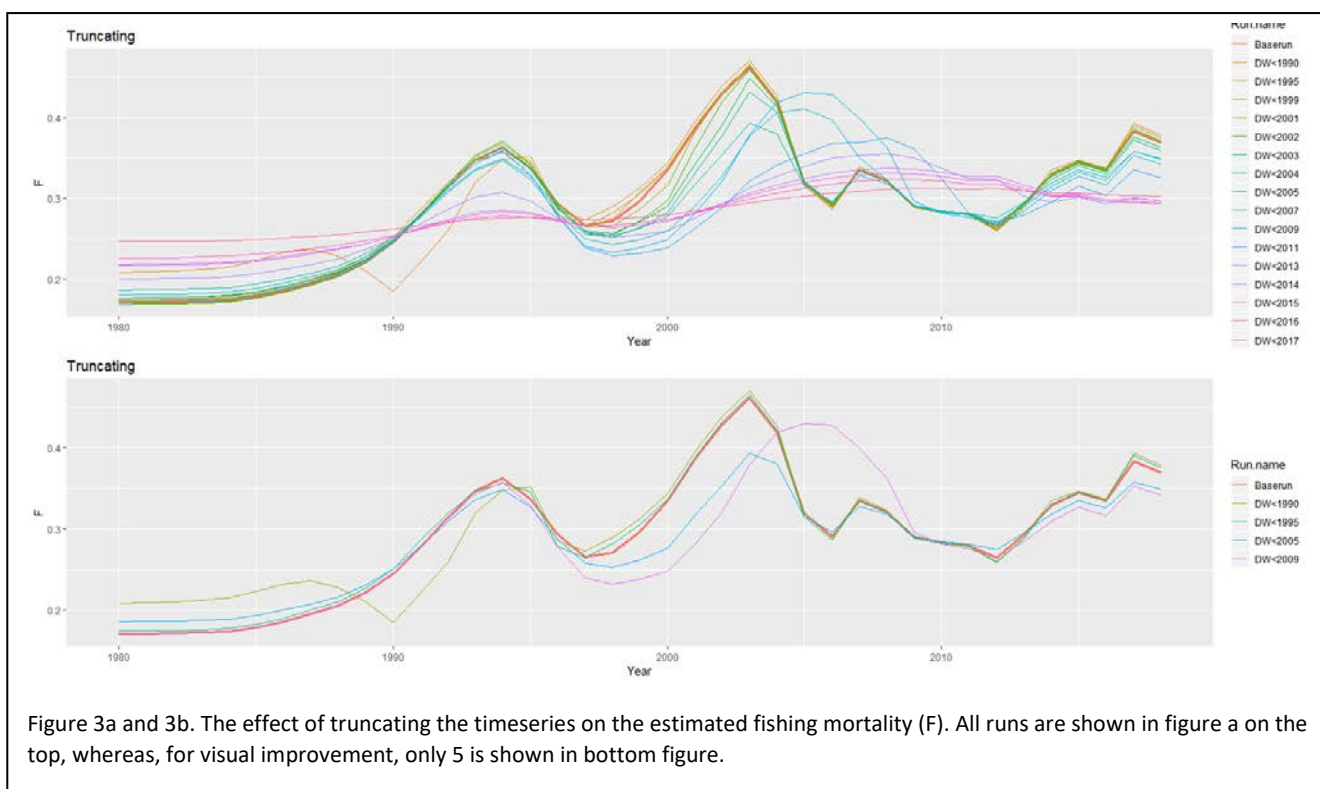
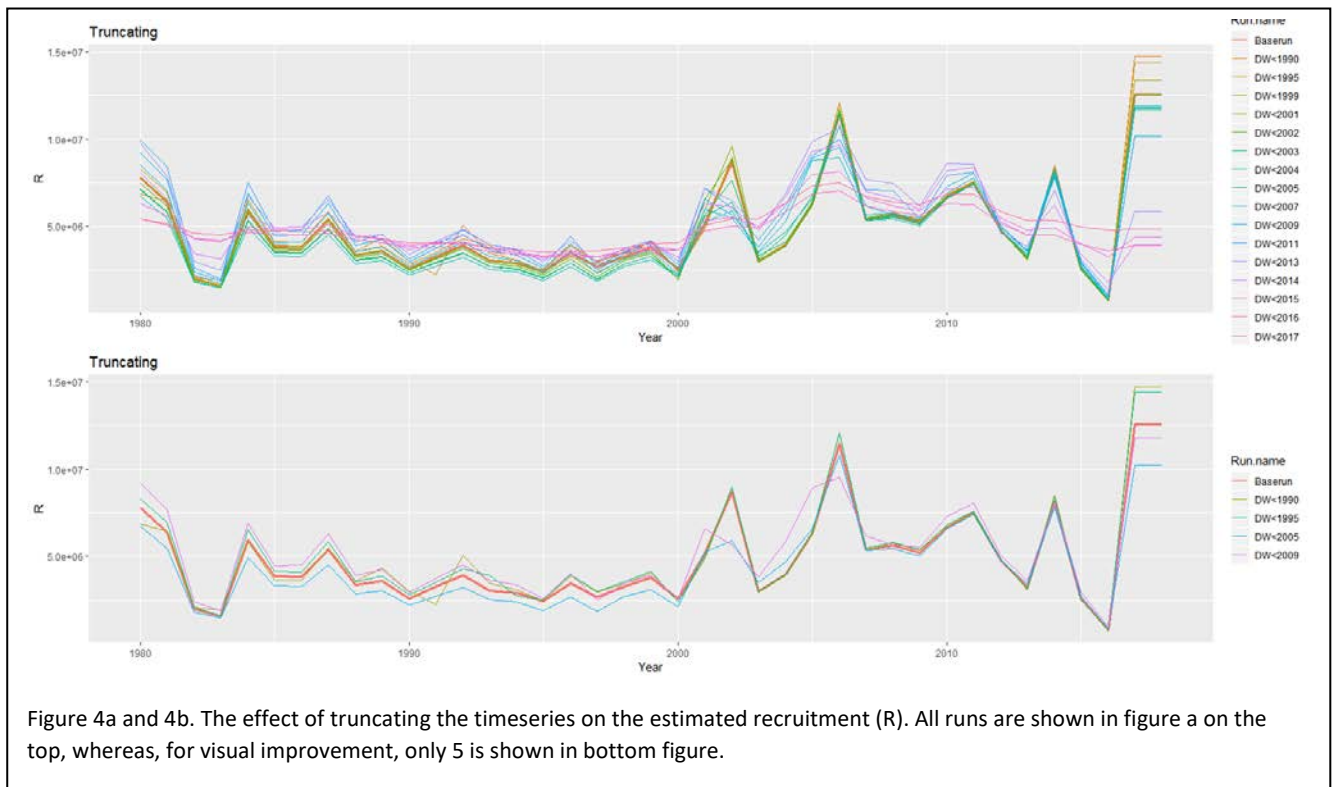


Figure 3a and 3b. The effect of truncating the timeseries on the estimated fishing mortality (F). All runs are shown in figure a on the top, whereas, for visual improvement, only 5 is shown in bottom figure.



References.

ICES. 2013a. Report of the Workshop on Northeast Atlantic Mackerel monitoring and methodologies including science and industry involvement (WKNAMMM). ICES CM 2013/SSGESST:18. 37pp.

ICES. 2013b. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 27 August–02 September 2013, ICES Headquarters, Copenhagen. ICES CM 2013/ACOM:15.

ICES 2013c:

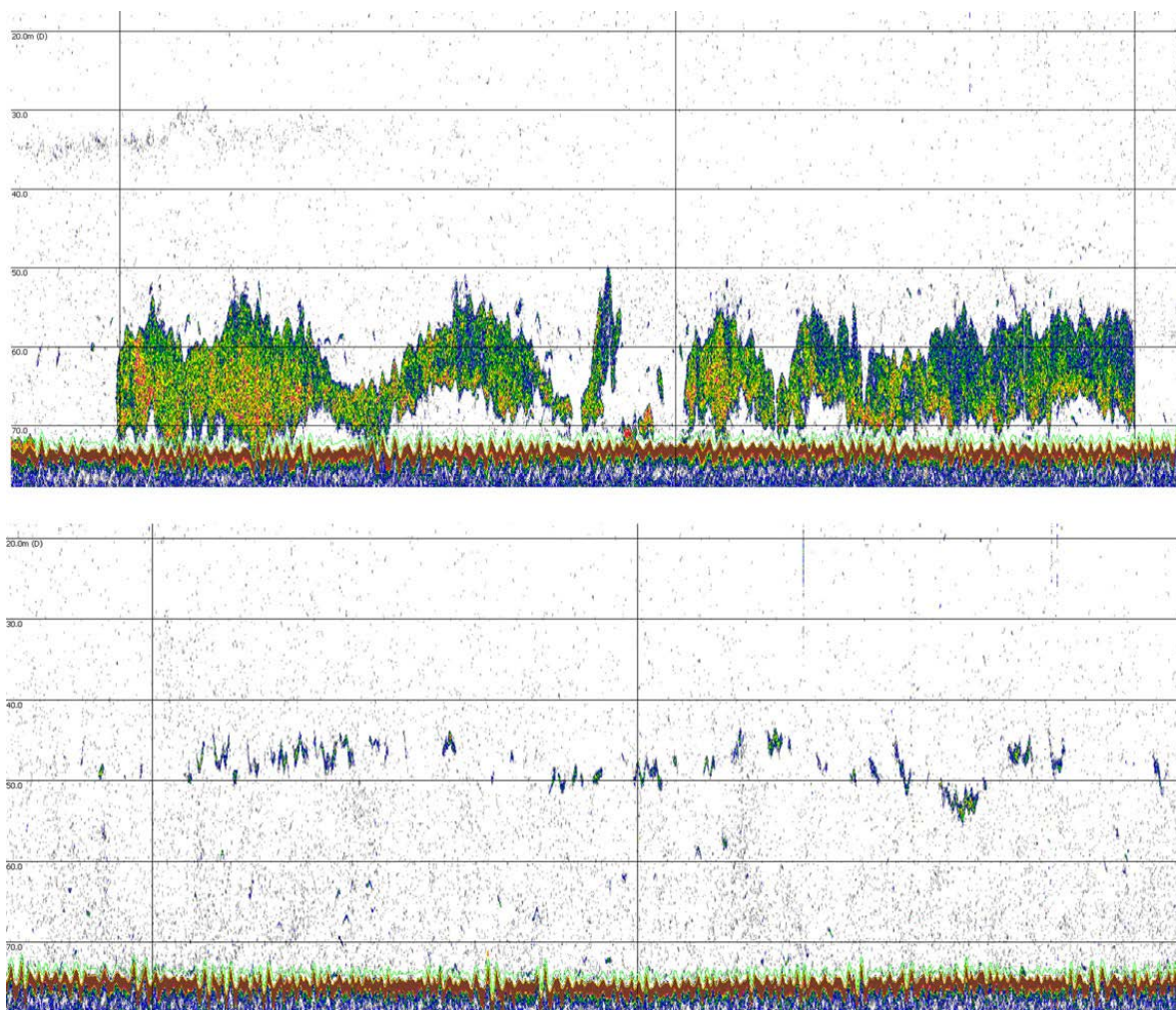
<http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2013/WGWIDE/19%20WGWIDE%20Report%20-%20Annex%20003%20Special%20Requests.pdf>

Update of the abundance index of age 0 NEA mackerel based on demersal trawl surveys in October – March

ICES IBPNewMac 2019 Working document

Teunis Jansen^{1,2}

- 1) GINR – Greenland Institute of Natural Resources, PO Box 570, 3900 Nuuk, Greenland
- 2) DTU AQUA – National Institute of Aquatic Resources, Kemitortvet, 2800 Lyngby, Denmark



Abstract

The index of survivors in the first autumn-winter (recruitment index) has not been updated since WGWIDE 2016 because of input data quality issues. This was solved, and the last data quality assurance features requested in ICES DATRAS system are expected to be online in due time for input data preparations for WGWIDE 2019. Meanwhile, was the model software outdated. A new implementation was therefore developed, tested and used to derive the index updated to include year classes from 1998 to 1997. The revision of 1998 to 2015 was minor. Latest year classes are estimated to be the most abundant in the time series.

Introduction

At WGWIDE 2018, the index of survivors in the first autumn-winter (recruitment index) could not be updated due to input data quality issues in the ICES DATRAS system that had not been updated as recommended by WKWIDE 2017 and WGWIDE 2017. The outdated time series from WGWIDE 2016 was therefore used in the assessment (ICES, 2018, 2017a, 2017b, 2016).

The present document describes the progress with the data quality and the updated model output.

Materials and methods

An index of survivors in the first autumn-winter (recruitment index) was derived from a geostatistical model fitted to catch data from bottom trawl surveys conducted during autumn and winter. A complete description of the data and model can be found in Jansen *et al.* (2015) and the NEA mackerel stock annex.

Data

The data were compiled from several bottom trawl surveys conducted between October and March from 1998–2018 by research institutes in (Denmark, England, France, Germany, Ireland, Netherlands, Norway, Scotland and Sweden). Surveys conducted on the European shelf in the first and fourth quarters are collectively known as the International Bottom Trawl Survey (IBTS), although several of the surveys use different names. All surveys sample the fish community on the continental shelf and upper shelf slope. IBTS Q4 covers the shelf from the Bay of Biscay to North of Scotland, excluding the North Sea, while IBTS Q1 covers the shelf waters from north of Ireland, around Scotland, the North Sea, Skagerrak and Kattegat.

Trawl operations during the IBTS have largely been standardized through the relevant ICES working group (ICES, 2013b). Furthermore, the effects of variation in wing-spread and trawl speed were included in the model (Jansen *et al.*, 2015). Trawling speed was generally 3.5–4.0 knots, and trawl gear is also standardized and collectively known as the Grande Ouverture Verticale (GOV) trawl. Some countries use modified trawl gear to suit the particular conditions in the respective survey areas, although this was not expected to change catchability significantly. However, in other cases, the trawl design deviated more significantly from the standard GOV type, namely the Spanish BAKA trawl, the French GOV trawl, and the Irish mini-GOV trawl. The BAKA trawl had a vertical opening of only 2.1–2.2 m and was towed at only 3 knots. This was considered substantially less suitable for catching juvenile mackerel and, therefore, was excluded from the

analysis. The French GOV trawl was rigged without a kite and typically had a reduced vertical opening, which may have reduced the catchability of pelagic species like mackerel. Catchability was assumed to equal the catchability of the standard GOV trawl because testing has shown that the recruitment index was not very sensitive to this assumption (Jansen *et al.*, 2015). Finally, the Irish mini-GOV trawl, used during 1998–2002, was a GOV trawl in reduced dimensions which was accounted for by inclusion of the wing-spread parameter in the model.

Statistical modelling

A geostatistical log-Gaussian Cox process model (LGC) with spatiotemporal correlations was used to estimate the catch rates of mackerel recruits through space and time. The modelled average recruitment index (squared CPUE) surface was mapped in Figure 8.6.2.1. The time-series of spatially integrated recruitment index values was used in the assessment as a relative abundance index of mackerel at age 0 (recruits) – see Figure 8.6.2.2.

Results

Input data updates

Following data was corrected:

- ICES DATRAS database system
 - o Up to 63 % of the NS IBTS stations were lost in some years. They were lost internally in the data processing that take place between the uploaded exchange data and the downloadable raised data products. The bugs that caused this have been fixed.
 - o Conversion of all lengths to “pinched tail length” using conversion tables from (Hansen et al., 2018).
 - o Revision of the length to age conversion algorithms (inter- and extrapolation rules).
 - o New and updated mackerel data products *CPUE by Age-Haul* developed for EVHOE, SWC_IBTS, SCOWCGFS and IE_IGFS, but not all are online yet.
 - o Data quality assurance reports developed, but not yet online.
- Scottish data revision of 14 hauls from 2005 and 2006
- Irish data have been included in DATRAS from 2003 onwards. This included a revision of some of the historical data.
- The Irish survey “ISCGS” used a downscaled trawl from 1998 to 2002. This trawl was fished slightly slower (~3.5 kn vs. ~4.0 in IBTS) due low vessel power, and the wing spread (and consequently the height of the opening) was substantially less than the GOV trawl (8.5 m vs. 19-20 m). This was accounted for in the data by “rasing” with i) 4.0/3.5 and, ii) the relation between the areas of semicircles with base line equal to the wing spreads.

An overview of the IBTS database is given in Figure 1.

Modelling

Model fitting with the original lgc-package was not possible. The developer (Kasper Kristensen) suggested that it could be due to either a change in hardware or software. Instead of spending time on (possibly) finding a workable setup using old versions of r and r-packages, he developed an updated software implementation based on TMB and uploaded the code to GitHub. The new software fitted the model approximately 10 times faster. The new software implementation was tested by loading the original data from (Jansen et al., 2015). The result was nearly exactly the same (details not appended here, but can be required from the author). During this work, it was discovered that the software defaulted missing values to 0 instead of omitting those data points as is normally done in r. Also, that there were records missing effort information. However, the effect on the index of this issue was minor. The results below describe the model fitted to the updated data from for year classes from 1998 to 2017.

Haul duration was removed from the model in WGWIDE 2016 because it was not significant ($p > 0.05$). Furthermore, the variation in haul duration is not random. Some countries stop towing when there are indications of a large catch. The relation between hauls duration and catch is therefore affected by two opposing processes which cannot readily be separated. It was therefore decided to keep haul duration exclude, and account for it in the data.

Haul speed was surprisingly estimated to be negative. The haul speed varied between vessels and areas. The estimated negative effect could therefore be an artefact, and was therefore removed from the model. Speed was then accounted for in the data.

Consequently, the catch expressed as catch by swept nautical mile, and the model only used wingspread as effort.

The wing spread was estimated to be negative. This was assumed to reflect an underlying negative relation between wing spread and vertical opening, combined with a positive relation between vertical opening and catch rate. Wing spread was therefore kept in the model, but has close to no effect on the model fit and resulting index.

Parameter estimates and standard errors in the final model are listed in Table 1.

Large scale spatial correlation (H) was substantially higher than the fit with the old data, and these spatial patterns were also more consistent through time (T). Furthermore, the variance of the nugget effect was higher which led the model to give more trust to large catches.

Distribution maps (observed and modelled) are given as an average over all year classes in Figure 2, for the most recent year class (2017) in Figure 3, and for each year in the time series as supplementary material 1.

Updated index for assessment

The revision of the recruitment time series i.e. the annual mean of square rooted expected catch was minor. The pattern in the index from 1998 to 2015 was similar to the index provided to WGWIDE 2016 ($r = 0.93$) and the index in (Jansen et al., 2015) (1998 – 2012, $r = 0.95$) (Figure 4). The estimates of added year

class 2016 and 2017 are high indicating good recruitment in recent years. Uncertainty of the index values were not provided before the meeting.

Correlation between WGWIDE assessment estimates of age 0 and the index remained moderate ($r=0.504$). Interestingly, the correlation was higher with a simple mean of square rooted observed catch rates ($r=0.66$), suggesting that the type of model and the model structure should be evaluated at the next benchmark.

Discussion

Survey Coverage

The combined demersal surveys have insufficient spatial coverage in some areas that can be important for the estimation of age-0 mackerel abundance, namely: (i) Since 2011, the English survey (covering the Irish sea and the central-eastern part of the Celtic sea including the area around Cornwall) has been discontinued, (ii) the Scottish survey has not consistently covered the area around Donegal Bay, (iii) the IBTS has observed high catch rates in some years at the north-eastern edge of the survey area (towards the Norwegian trench) in winter. It is therefore possible that some recruits are also overwintering on the other side of the trench along the south western shelf edge of Norway. Consequently, the NS-IBTS in Q1 should be extended to include the south-western Norwegian shelf and shelf edge in proximity to the Norwegian trench.

Data Quality

WGWIDE 2018 encouraged studies of vertical distribution and catchability of age-0 mackerel in the Q4 and Q1 surveys. This recommendation is echoed here.

Age readings are missing from Ireland in some years. I recommend this is done, with priority to the most abundant year classes and for year classes (± 1) where the present index and the WGWIDE assessment estimate of age 0 abundance deviates substantially.

The ICES DATRAS data products “CPUE per age and haul”, “CPUE per length and haul” and the data quality reports (comparisons of summary statistics for data uploaded to DATRAS and the two aforementioned data products) should be online and tested for all relevant surveys before 1st of June 2019 to ensure a timely and quality assured update of the index for WGWIDE 2019.

Acknowledgements

I wish to thank Kasper Kristensen (DTU AQUA) for developing a new software solution for this work.

Figures

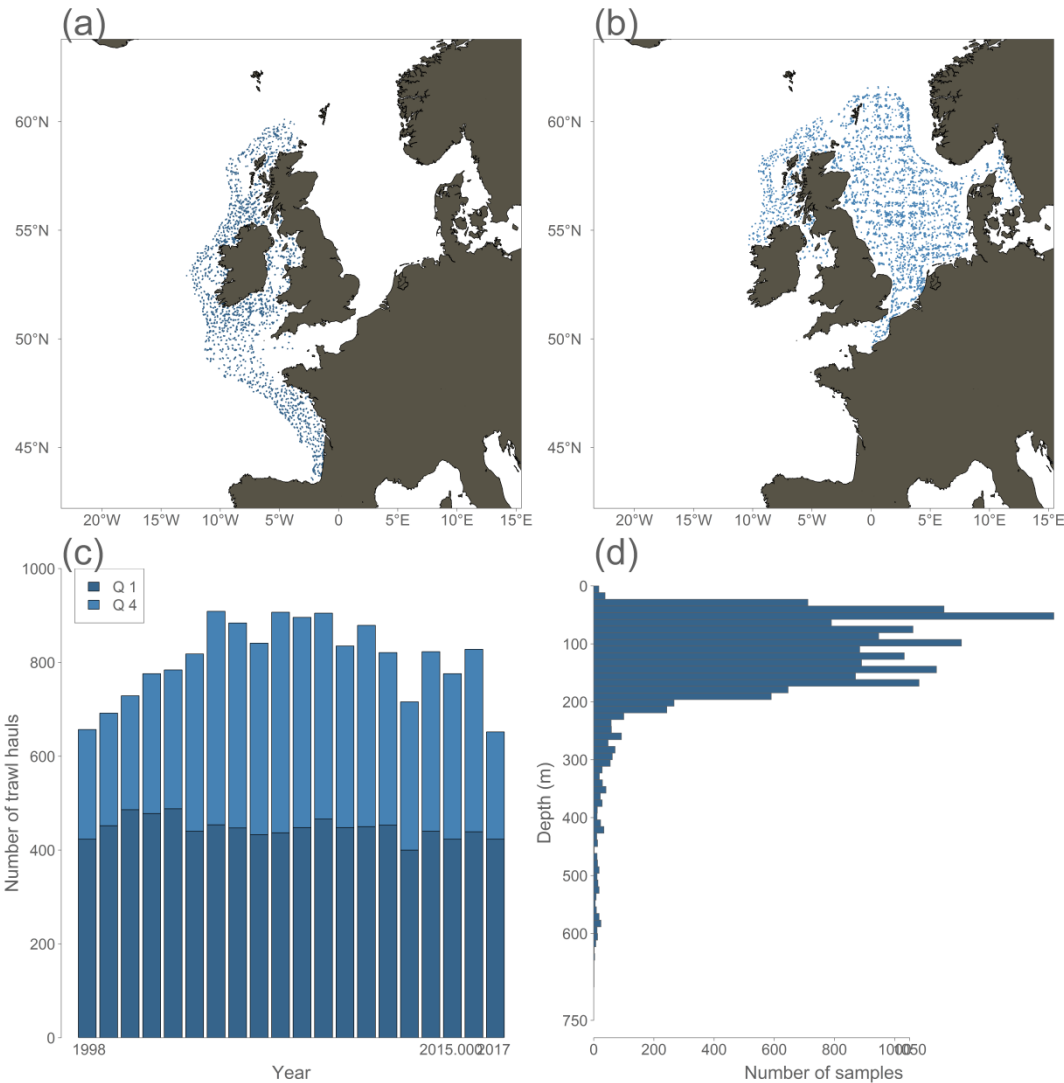


Figure 1. Demersal trawl survey data used to derive the abundance index of age-0 mackerel. (a) Trawl sample locations in the fourth quarter (Q4, October - November, blue dots); (b) trawl sample locations in the first quarter (Q1, January - March, light blue dots); (c) number of samples by year and quarter; and (d) depth.

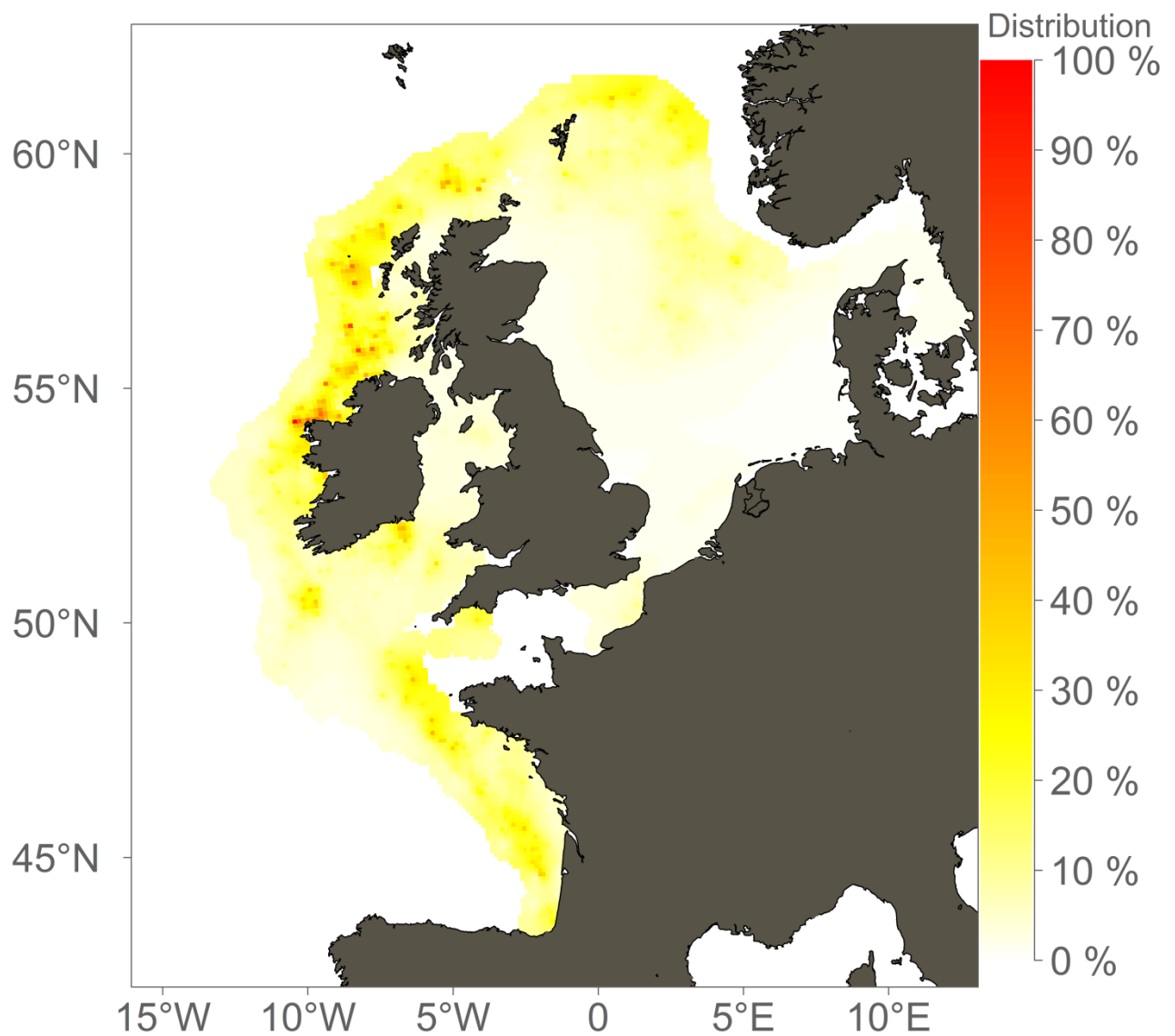


Figure 2. Spatial distribution of mackerel juveniles at age 0 in October to March. Average of year classes 1998 - 2017. Modelled squared catch rates per 10 x 10 km rectangle. Each rectangle is coloured according to the expected squared catch rate in percent of the highest value.

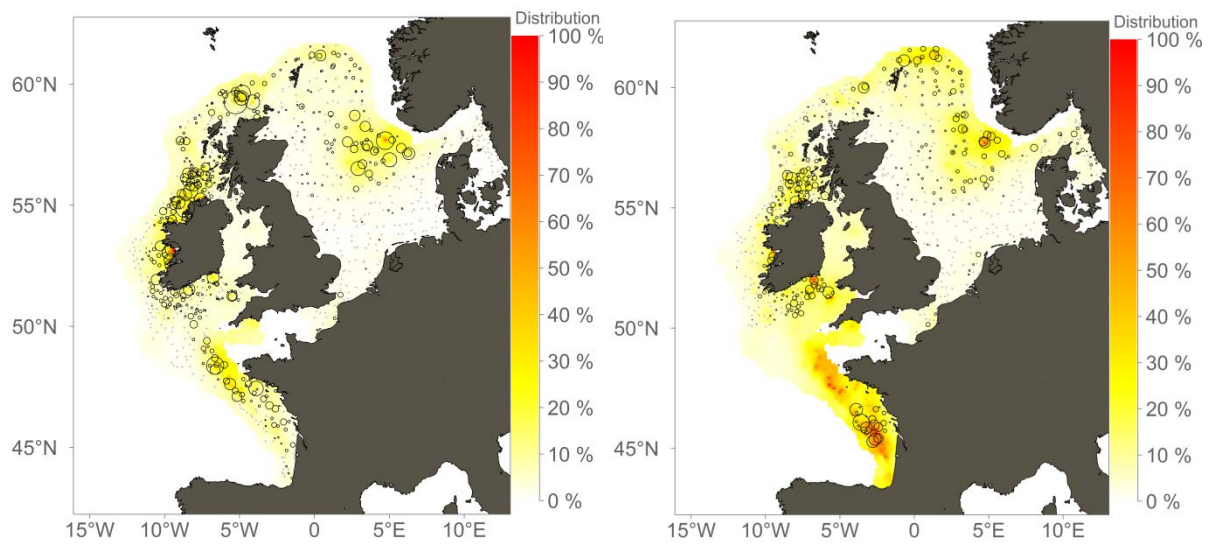


Figure 3. Spatial distribution of mackerel juveniles at age 0 in October to March. Left: The 2016 year class, right: the 2017 year class. Mackerel squared catch rates by trawl haul (circle areas represent catch rates in kg/km²) overlaid on modelled squared catch rates per 10 x 10 km rectangle. Each rectangle is coloured according to the expected squared catch rate in percent of the highest value for that year.

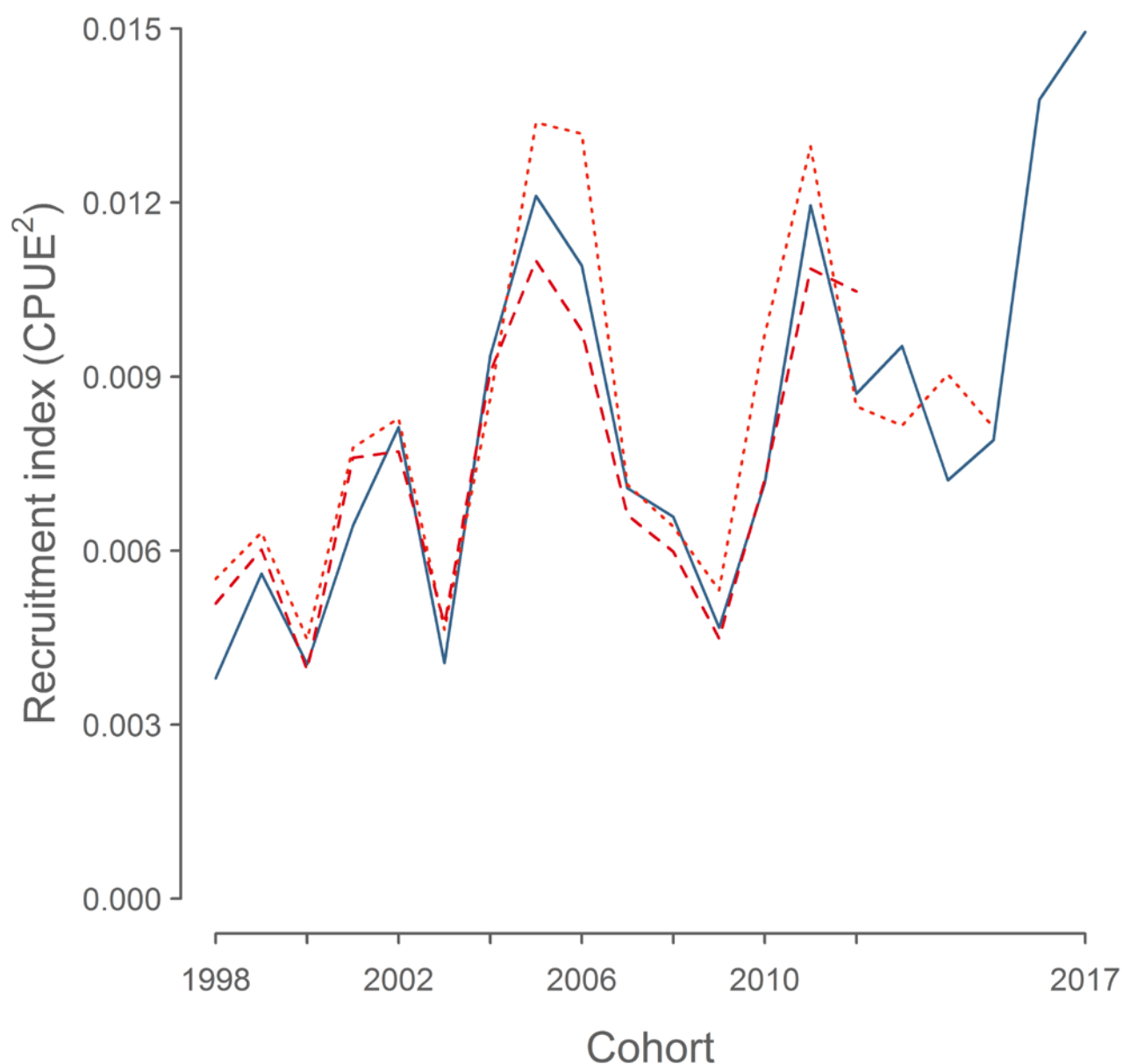


Figure 4. Index of mackerel juveniles at age 0 in October to March proxied by annual integration of square root of expected catch in demersal trawl surveys (Blue line). Old versions of the index rescaled for reference: as red stippled line, Jansen et al. (2015) and WGWISE 2016 as red dotted line.

Tables

Symbol	Description	Unit	Estimate	St.err.	CV (%)
T	Decorrelation time	year	2.2	0.46	21
H	Spatial decorrelation distance	km	511.2	110.12	22

σ_{xy}^2	Spatial variance parameter (year specific surfaces)	1	ToDo		
$\sigma_x^{2^2}$	Spatial variance parameter (intercept surface)	1	ToDo		
σ_N^2	Variance of the nugget effect	1	3.9		
WS	Log Wing spread	Nmi	-1.5	0.69	46

Table 1. Model parameter estimates and standard errors

Supplementary information

Supplementary information 1. Spatial distribution of mackerel juveniles at age 0 in October to March. One map per year class for the entire time series. Mackerel squared catch rates by trawl haul (circle areas represent catch rates in kg/km²) overlaid on modelled squared catch rates per 10 x 10 km rectangle. Each rectangle is coloured according to the expected squared catch rate in percent of the highest value for that year.

References

- Hansen, F.T., Burns, F., Post, S., Thygesen, U.H., Jansen, T., 2018. Length measurement methods of Atlantic mackerel (*Scomber scombrus*) and Atlantic horse mackerel (*Trachurus trachurus*) – current practice, conversion keys and recommendations. *Fish. Res.* 205, 57–64.
- ICES, 2018. Report of the Working Group on Widely Distributed Stocks (WGWIDE). ICES C. 2018/ACOM23.
- ICES, 2017a. Report of the Working Group on Widely Distributed Stocks (WGWIDE). ICES C. 2017/ACOM23.
- ICES, 2017b. Report of the Benchmark Workshop on Widely Distributed Stocks (WKWIDE). ICES C. 2017/ACOM36.
- ICES, 2016. Report of the Working Group on Widely Distributed Stocks (WGWIDE). ICES C. 2015/ACOM16.
- Jansen, T., Kristensen, K., van der Kooij, J., Post, S., Campbell, A., Utne, K.R., Carrera, P., Jacobsen, J.A., Gudmundsdottir, A., Roel, B.A., Hatfield, E.M.C., 2015. Nursery areas and recruitment variation of North East Atlantic mackerel (*Scomber scombrus*). *ICES J.Mar.Sci.* 72(6), 1779–1789. <https://doi.org/10.1093/icesjms/fsu186>

Fishery/Survey Timing
WKIBPNEAMac
Wageningen Marine Research
4-7 March, 2019

Andrew Campbell
Fisheries and Ecosystems Advisory Services,
Marine Institute,
Galway, Ireland

Introduction

The proportion of the total annual fishing mortality occurring prior to the recording of any fishery independent index data is an important consideration for the stock assessment.

The mackerel fishery has changed substantially in recent years with an increase in exploitation during the summer months potentially changing the temporal development of the fishery throughout the year when compared to years prior to the geographic expansion of the stock distribution.

Available Data

In addition to providing catch and sample data by quarter and ICES division for input to the stock assessment, the distribution of catches by ICES statistical rectangle is also provided. The greater spatial resolution of this data allows the WG to monitor and investigate the distribution of the fishery from year to year. Of relevance to this exercise is the frequency of this data which was increased from quarterly to monthly in 2010. Available data for years prior to 2010 was also collated. For years prior to 2010 where only quarterly data is available, the quarterly catch has been assumed to be distributed evenly throughout the quarter for this analysis.

The catch by statistical rectangle is, in some cases, derived from a source other than official logbook declarations. This can lead to minor discrepancies between the total ICES catch and the total catch by statistical rectangle. In addition, some minor fleets (*e.g.* countries with minor catches, by-catch and discard fleets) do not supply catch by statistical rectangle.

Table 1: WG, Total Catch Reported by Statistical Rectangle.[illegible]

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
DE	M	M	M	M	M	M	M	M	M	M
DK	M	M	M	M	M	M	M	M	M	M
ES	M	M	M	M	M	M	M	M	M	M
FO	Q	M	M	M	M	M	M	M	M	M
FR	M	M	M	M	M	M	M		M	M
GL	-	-	-	M	M	M	M	M	M	M
IC	M	M	M	M	M	M	M	M	M	M
IE	M	M	M	M	M	M	M	M	M	M
NL	M	M	M	M	M	M	M	M	M	M
NO	M	M	M	M	M	M	M	M	M	M
PT	M	M	M	M	M	M	M	M	M	M
RU	Q	M	M	M	M	M	M	M	M	M
SE	M	M	M	M	M	M	M	M	M	M
UKE	Q	Q	M	M	M	M	M	M	M	M
UKN	M	M	M	M	M	M	M	M	M	M
UKS	M	M	M	M	M	M	M	M	M	M

Table 2: Reporting of Catch by Statistical Rectangle (- no catch).

An important fishery independent index for the NEA Mackerel assessment is available from the IESSNS summer survey. Currently, data from surveys in 2010 and 2012-2018 is used in the assessment. The start and end dates for these surveys are shown in table 3. In general, the survey runs for the full month of July such that mid-July can be considered as proxy for the survey time. The proportion of the total catch taken by mid-July (*i.e.* the proportion of the total represented by Jan-Jul) is also given.

It can be seen that this proportion has reduced from approximately 50% during 2010-2013 to 35%.

Year	Start	End	FProp July	Year	Start	End	FProp July
1998	NA	NA	56%	2009	NA	NA	62%
1999	NA	NA	52%	2010	08/07/10	21/08/10	49%
2000	NA	NA	50%	2011	NA	NA	48%
2001	NA	NA	52%	2012	02/07/12	10/08/12	46%
2002	NA	NA	49%	2013	02/07/13	09/08/13	46%
2003	NA	NA	44%	2014	02/07/14	12/08/14	35%
2004	NA	NA	51%	2015	01/07/15	10/08/15	34%
2005	NA	NA	57%	2016	01/07/16	31/07/16	36%
2006	NA	NA	51%	2017	03/07/17	04/08/17	35%
2007	NA	NA	48%	2018	30/06/18	06/08/18	NA
2008	NA	NA	53%				

Table 3: IESSNS dates and proportion total catch by mid-July.

Figures 1 - 5 indicate the development of the fishery throughout the year for 1998-2017. The solid black line represents the total fishery with the proportion of total national catch also shown. Where appropriate, the timing of the IESSNS is represented by the shaded rectangle. Relative shares have remained largely stable in recent years, despite the lack of agreement by the fishing parties on a

TAC. The reduced proportion is as a result of a later start for the northern fisheries (Iceland, Russia and particularly Norway which accounts for the largest national catch) and a reduction in the proportion of national quota taken in Q1 by Scotland and Ireland.

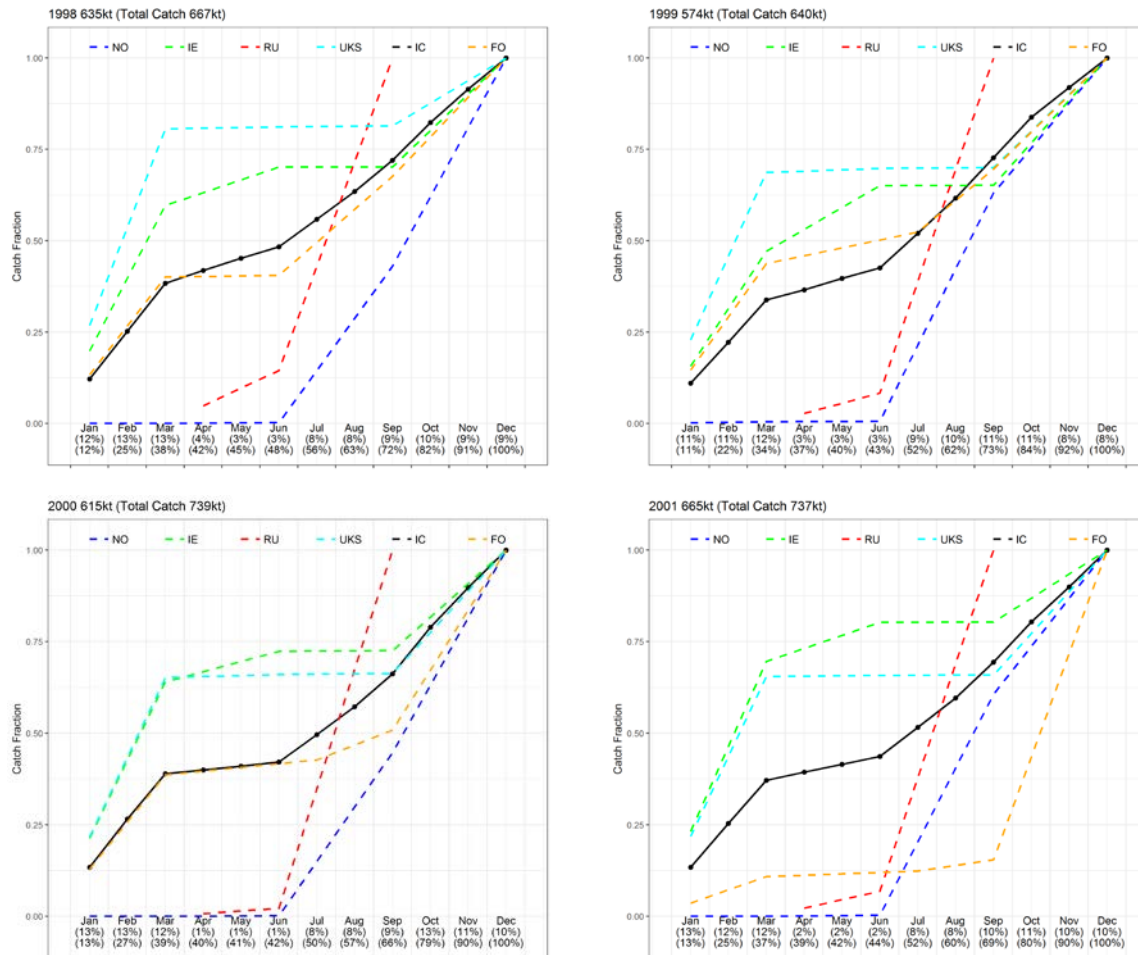


Figure 1: 1998-2001.

Working Document to WKIBPNEAMac 2019 – Fishery/Survey Timing

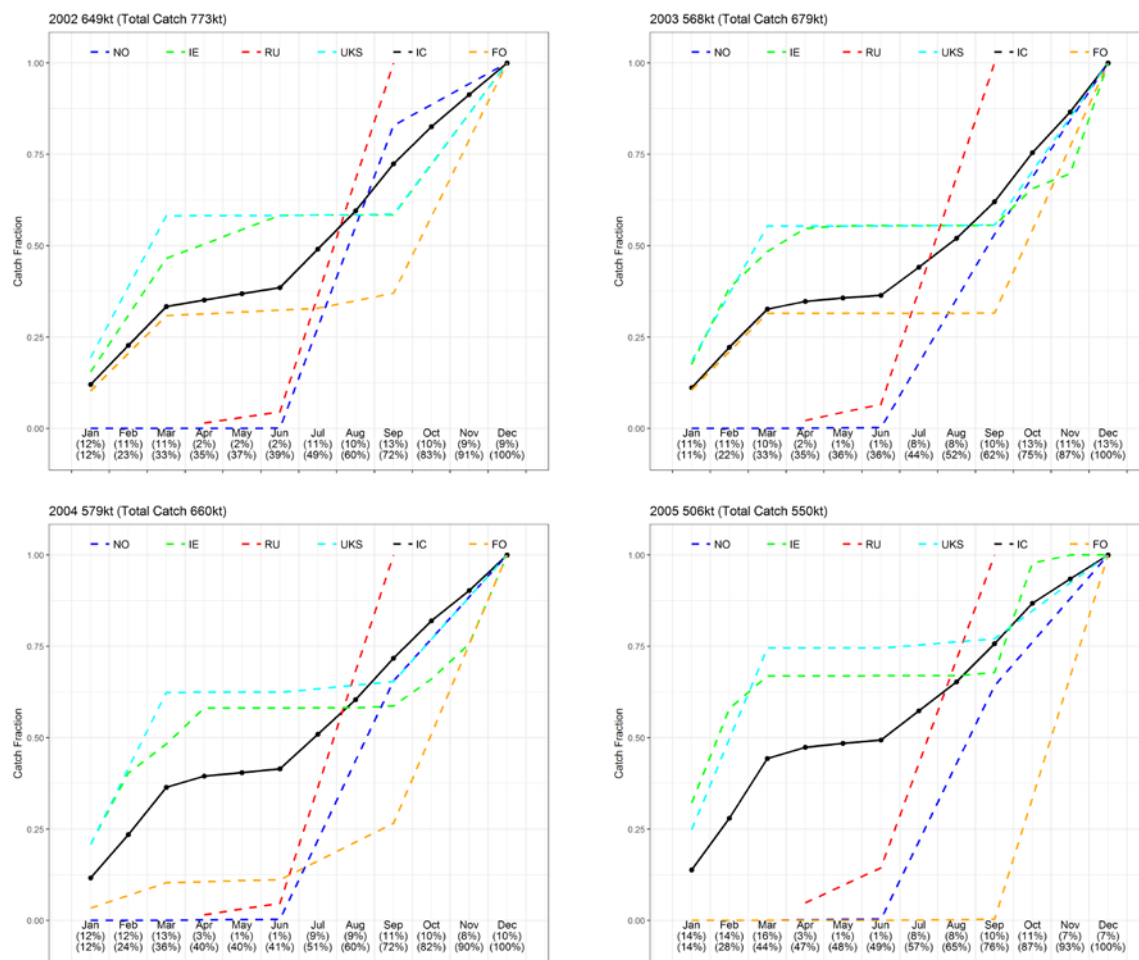


Figure 2: 2002-2005.

Working Document to WKIBPNEAMac 2019 – Fishery/Survey Timing

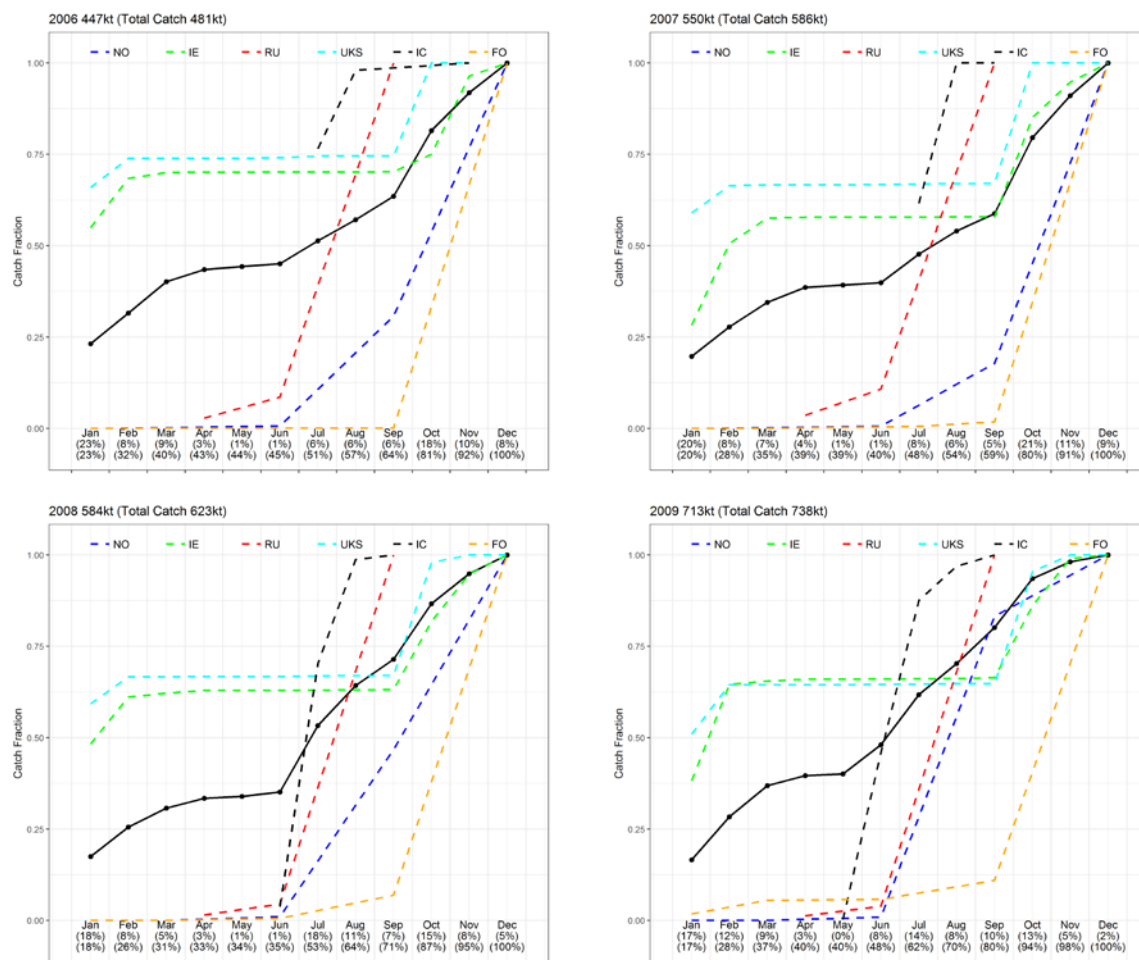


Figure 3: 2006-2009.

Working Document to WKIBPNEAMac 2019 – Fishery/Survey Timing

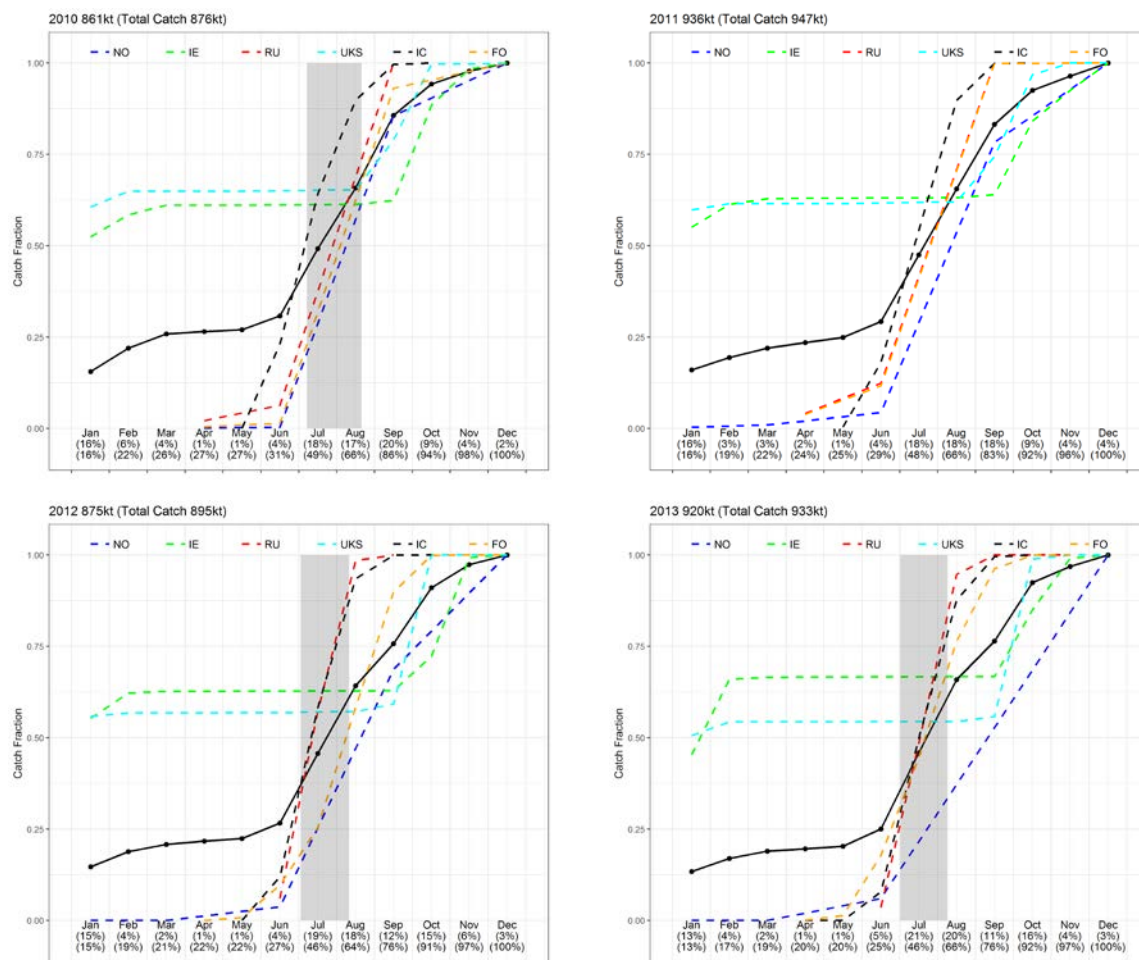


Figure 4: 2010-2013.

Working Document to WKIBPNEAMac 2019 – Fishery/Survey Timing

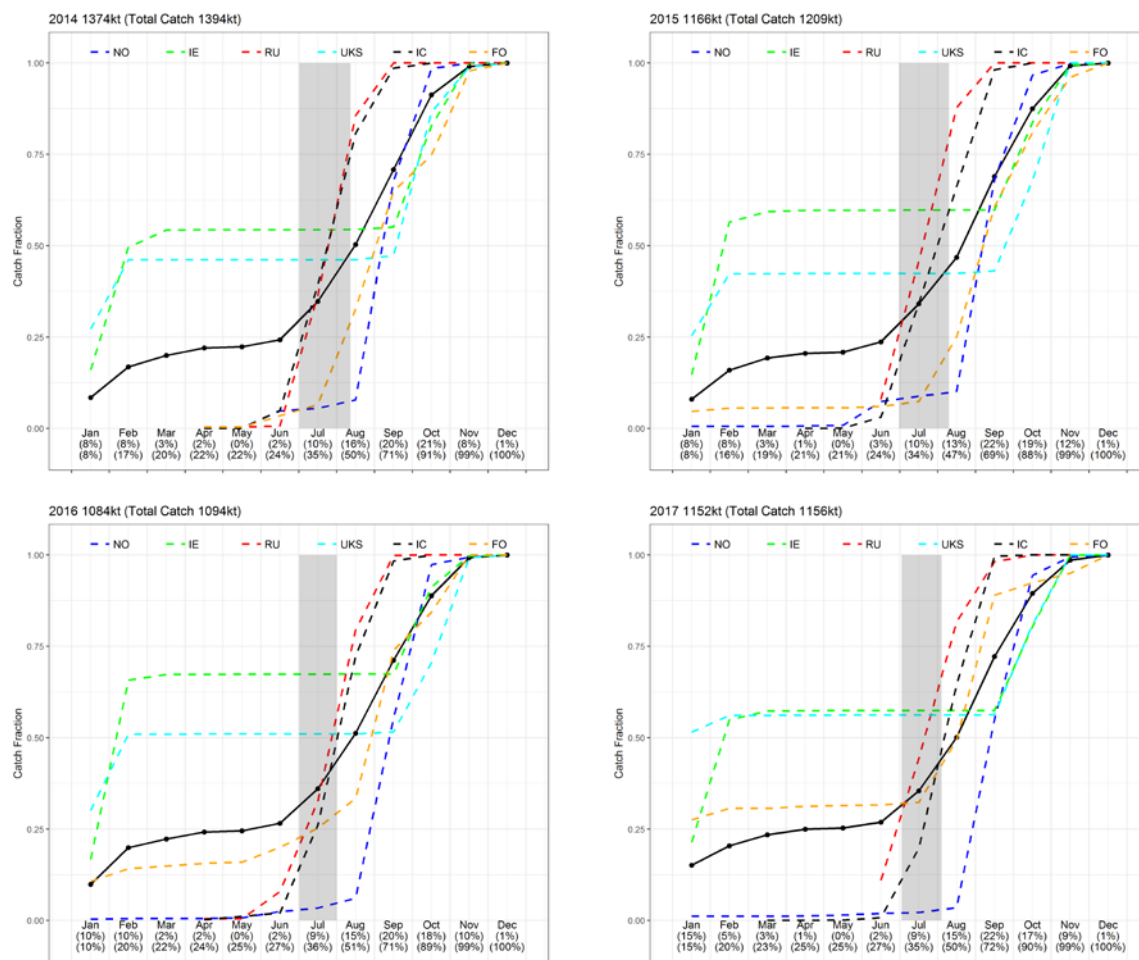


Figure 5: 2014-2017.

Issues regarding updated version of RFID-tag data 2019

Aril Slotte

Contact: Email: Aril@hi.no, Institute of Marine Research, Bergen, Norway

Background and objectives

The RFID tagging started in 2011, the first factories with RFID antennas started working in 2012. Since then there has been a development in both the distribution of factories with RFID antenna systems scanning mackerel landings, but also a development in software solutions to monitor the factories. This WD is describing data excluded from the estimation of the tag table going into the assessment, and the reasoning behind it. It is also describing the differences in recapture rates between factories over the time series.

Exclusions of experiments and factories having issues with efficiencies in estimation of tag file

Experiments excluded from the assessment are 1 experiment off the Norwegian coast in 2011, only targeting young mackerel, mostly the 2010 year class. Also excluded are all experiments in Iceland 2015-2017.

Factories excluded from the time series due to very low efficiency in WGWIDE 2018 were:

- Pelagia Austevoll 2012-2017
- Pelagia Egersund 2014-2017
- Lunar Freezing Fraserburgh 2014-2017

These three factories were included again in 2018 scannings due to high efficiency in the large scale testing (Table 1), and normal recapture rates (Figure 8)

In autumn 2018 a large scale test program were initiated, where each factory got a test material of 100 tags, 10 tests of 10 tags and a tagging gun. They got instructions to do 10 different tests where they tagged 10 fish and released them into the RSV tanks of the vessel landing the catch. In table 1 these tests are summarized. Based on these tests it was decided to exclude 4 factories from 2018:

- Brødrene Sperre
- Vikomar (new factory)
- Grøntved Pelagic (new factory)
- Lofoten Viking (new factory)

In addition, Nortbay factory was excluded as catch data were not delivered in time for the update estimation.

After excluding these factories, the mean efficiency was estimated to 93%, which is an acceptable efficiency. However, one cannot be sure this efficiency has been the same backwards in time for all factories, there has been some services etc. There are 3 factories that are not touched, fixed or serviced after initiation, and that still had high efficiencies in the test, these are Vopnafjord at Iceland, Pelagia Shetland in Scotland, Pelagia Liavaagen in Norway. This is something one may have in mind when comparing recapture rates between factories back in time.

Differences in recapture rates between factories

In Figure 1 is given an overview of the recaptures rate (in terms of tons scanned per recaptured fish) development at the different factories over the time series, here including recaptures from all experiments. The figure suggests a potential problem in 2012, with very variable recapture rates compared with the other years; a solution may be to exclude this year from the data. The more detailed figures per year 2012-2018 (Figures 2-8) indicate that there still is variability also in years 2013-2018, but at a level perhaps more acceptable. The variability may be due to differences in efficiencies like shown in Table 1, but also differences in year classes scanned, this is not addressed when looking more roughly at recapture rates.

Corrections in numbers scanned in the updated RFID data for interbenchmark

Over the time series there has been development in the software used to monitor all the factories. In autumn 2013 one decided to store conveyor belt tag data for those factories having tags incorporated into the conveyor belt. The idea was that one would follow the production making sure the system was operating during a landing by looking at how many times the tag in the conveyor belt was passing the antenna. It turned out that the software solution, a web solution to monitor the factories could cope properly with all the data produced, also there were other issues with slow processes between data going in and out of database and all analyses and estimation need for producing the tag table going into the assessment. Therefore a process with development of new quicker web solutions was started, and the new solutions was ready in 2018. This solution handles all conveyor belt statistics very well, and therefore for the interbenchmark process these data were looked at with the purpose of finding mis-match, or discrepancies in the data. Factories that seemed to have comparatively lower recapture rates than others were looked at and landing data corrected. For all figures 1-8, data shown is based on an update data set where some discrepancies have been fixed.

What has been done is as follows. In factories with conveyor belt tags, if there were periods where the conveyor tag was not detected, unstable, and where no recaptures were coming in, but landing data had been reported, then landing data were removed from the database. This counts for the following years and factories and will reduce the numbers scanned in tag table:

- 2013, Skude Factory, landings removed from 8-28.1, in total 4410 t
- 2014, Brødrende Sperre, landings removed from 23.2 and 30.9-22.10, in total 11495 t
- 2014, Pelagia Måløy, landings removed from 3-4.2 and 27.8-24.9, in total 9793 t
- 2015, Pelagia Florø, landings removed from 25.10-18.11, in total 6318 t
- 2015, Skude Factory, landings removed from 30.10-20.11, in total 2204 t
- 2016, Pelagia Selje, landings removed from 10-16.1, in total 2827 t
- 2016, Vardin, landings removed from 16.9-6.10, in total 2200 t
- 2017, Brødrene Sperre, 10-25.1, and 12.9-16.10, in total 21433 t

Use of age samples to estimate numbers scanned by year class in landings

Up to now Icelandic biological samples with age data has been used to allocate to the landings in the estimation of numbers scanned per year class in their landings. Same for Faroes, the nation's own biological data. However, Scottish biological data have not been available in right format yet, and Norwegian biological data from same area and period has been allocated to both landings at Scottish factories and Norwegian. Hopefully at the time of benchmark 4-7.March 2019 we may be able to change allocation and look at sensitivity to using Scottish biological data seemingly with younger fish.

Table 1. Overview of tests of efficiency of RFID antenna systems. Red, not included due to issues with antenna systems. Green not included, mainly herring landings.

Factory	N-tests	Efficiency	Potential problem
DK01 Sæby	0		Not online -not included
FO01 Vardin Pelagic	0		Burned down
GB01 Denholm Coldstore	10	84,0	Some issues with noise -unstability - still included
GB01 Denholm Factory	9	95,6	
GB02 Lunar Freezing Peterhead	10	99,0	
GB03 Lunar Freezing Fraserburgh	8	91,3	Not included up to 2017 - included from 2018
GB04 Pelagia Shetland	10	99,0	
GB05 Northbay Pelagic	10	88,0	Some issues with noise -unstability
IC01 Vopnafjord	10	98,0	
IC02 Neskaupstad	5	74,0	Increased production speed with 36% in 2018, some detection problems, still included
IC03 Höfn	0		Antenna problems -not included
NO01 Pelagia Egersund Seafood	1	91,0	
NO02 Skude Fryseri	1	10,0	New engine prior to Q3-4 create noise problem, include only Q1
NO03 Pelagia Austevoll	10	96,0	
NO04 Pelagia Florø	0		Closed down
NO05 Pelagia Måløy	10	93,0	
NO06 Pelagia Selje	10	97,0	
NO07 Pelagia Liavågen	8	98,8	
NO08 Brødrene Sperre	2	50,0	Unstable - New noise problems autumn 2018, not included
NO09 Lofoten Viking	5	22,0	Unstable - New factory - noise problems autumn 2018, herring focus, not included
NO10 Pelagia Træna	10	96,0	New factory 2018 - high effectivity - focus on herring - not included
NO11 Nergård Sild Senjahopen	0		
NO12 Pelagia Lødingen	2	30,0	New factory - noise problems detected
NO13 Pelagia Tromsø	0		
NO14 Nils Sperre	10	97,0	New factory 2018 - high effectivity - included
NO15 Grøntvedt Pelagic	2	15,0	Noise problems detected, not included
NO16 Vikomar	6	68,3	One of two ring antennas has stopped working 2018, not included
Mean efficiency		93,0	

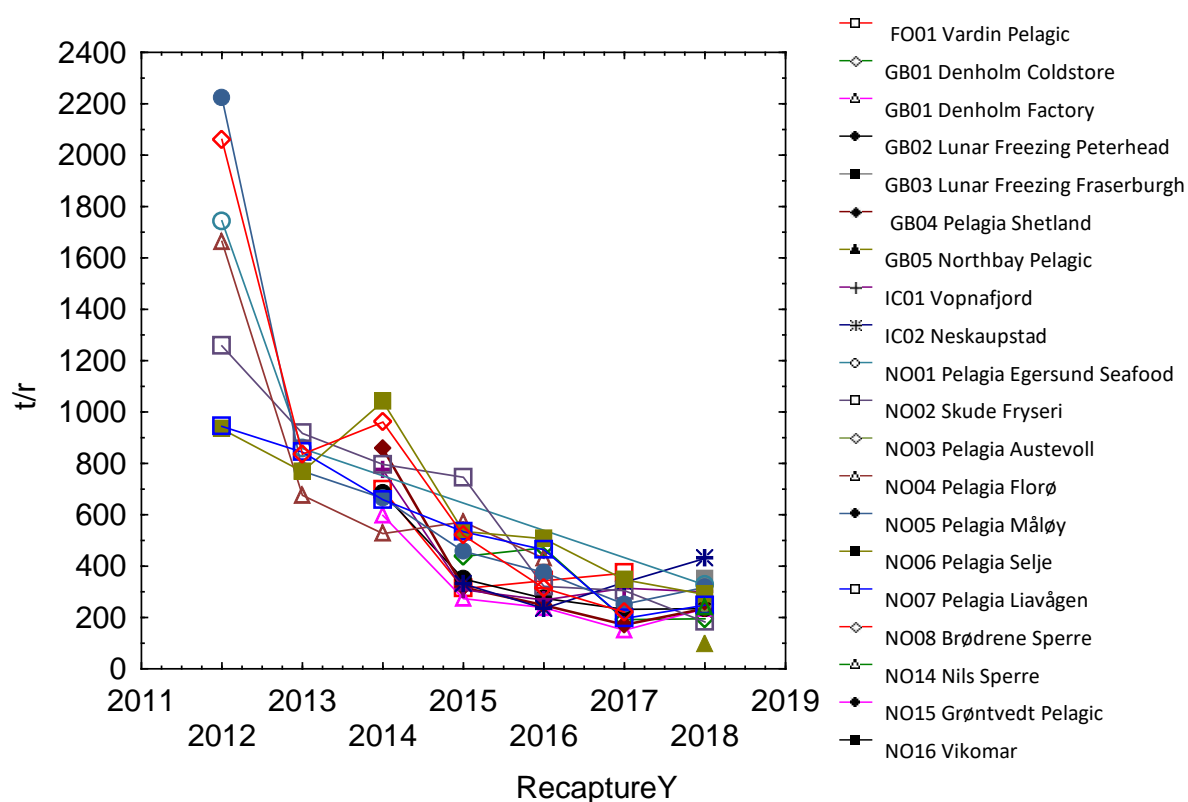


Figure 1. Number of tons scanned per recaptured mackerel (regardless of release year and areas), per recapture year and factory, see Figures 2-8 for details.

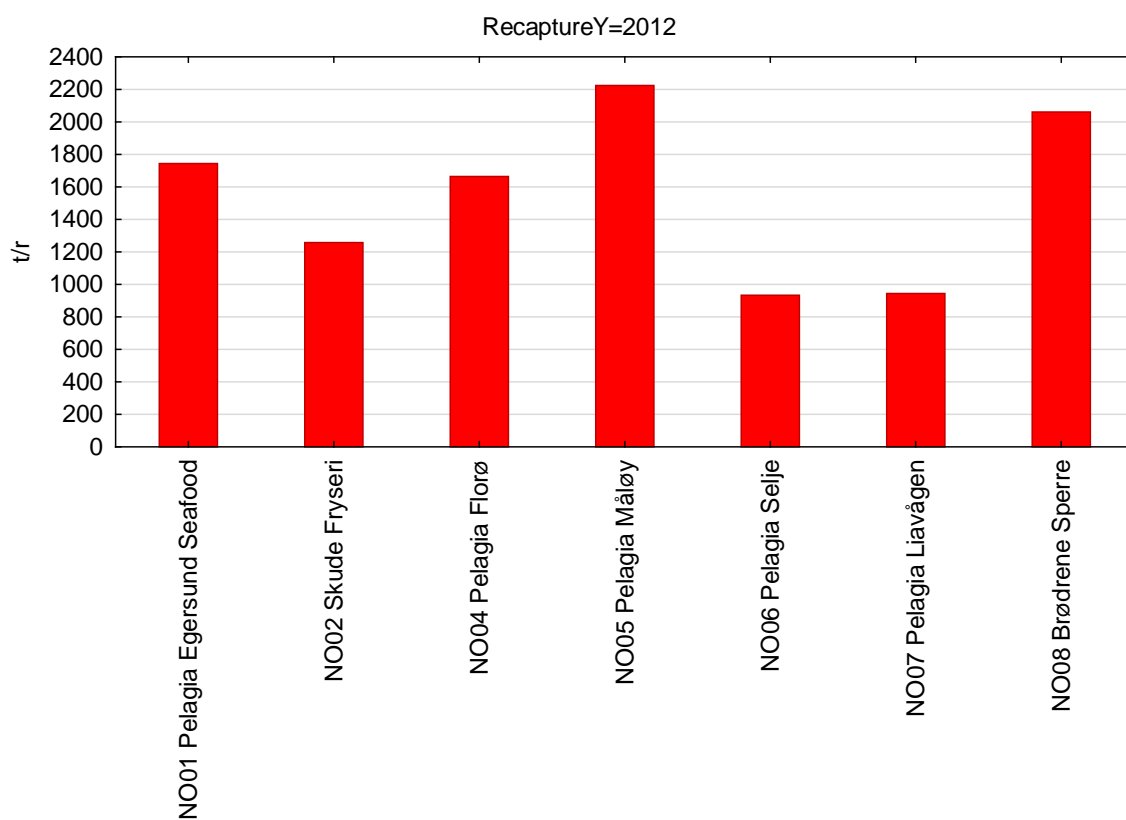


Figure 2. Number of tons scanned per recaptured mackerel (regardless of release year and areas) by factory in 2012.

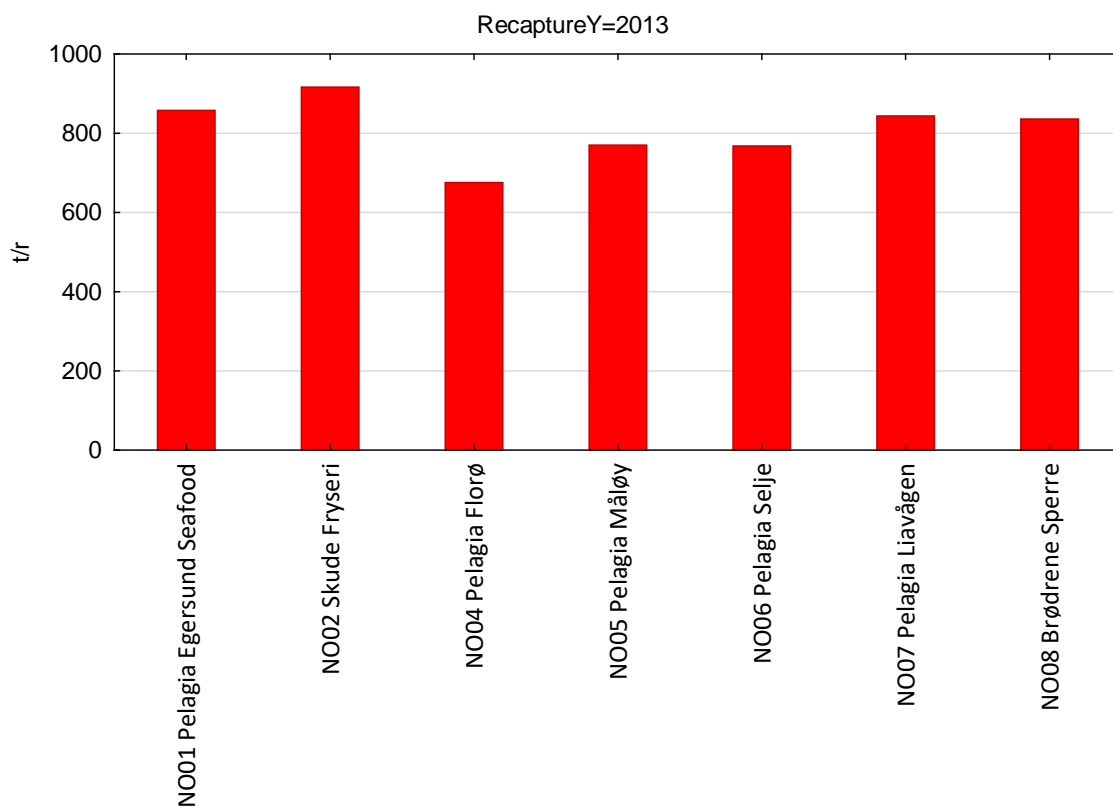


Figure 3. Number of tons scanned per recaptured mackerel (regardless of release year and areas) by factory in 2013.

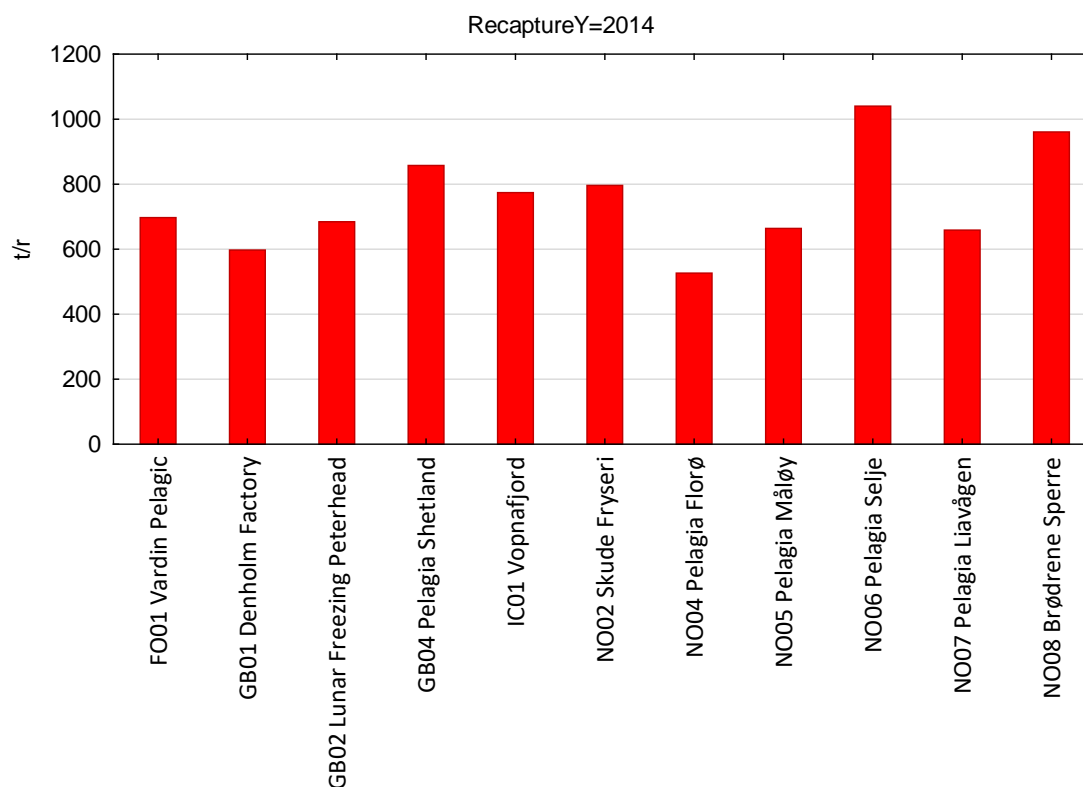


Figure 4. Number of tons scanned per recaptured mackerel (regardless of release year and areas) by factory in 2014.

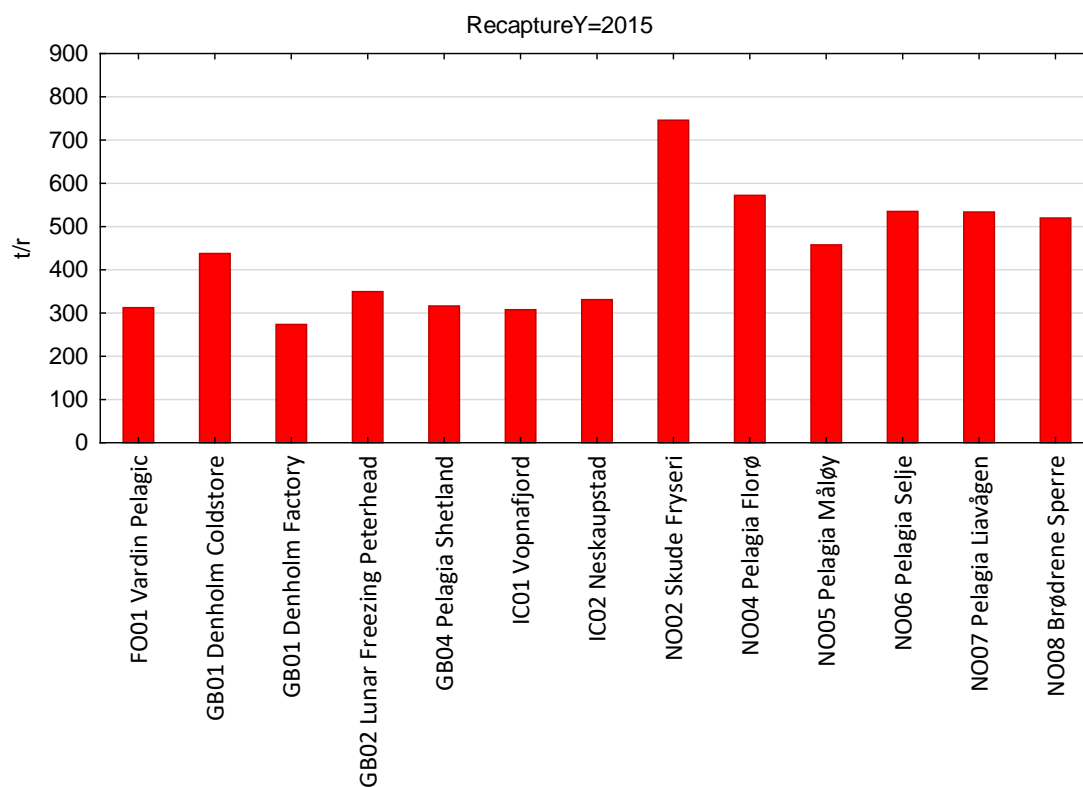


Figure 5. Number of tons scanned per recaptured mackerel (regardless of release year and areas) by factory in 2015.

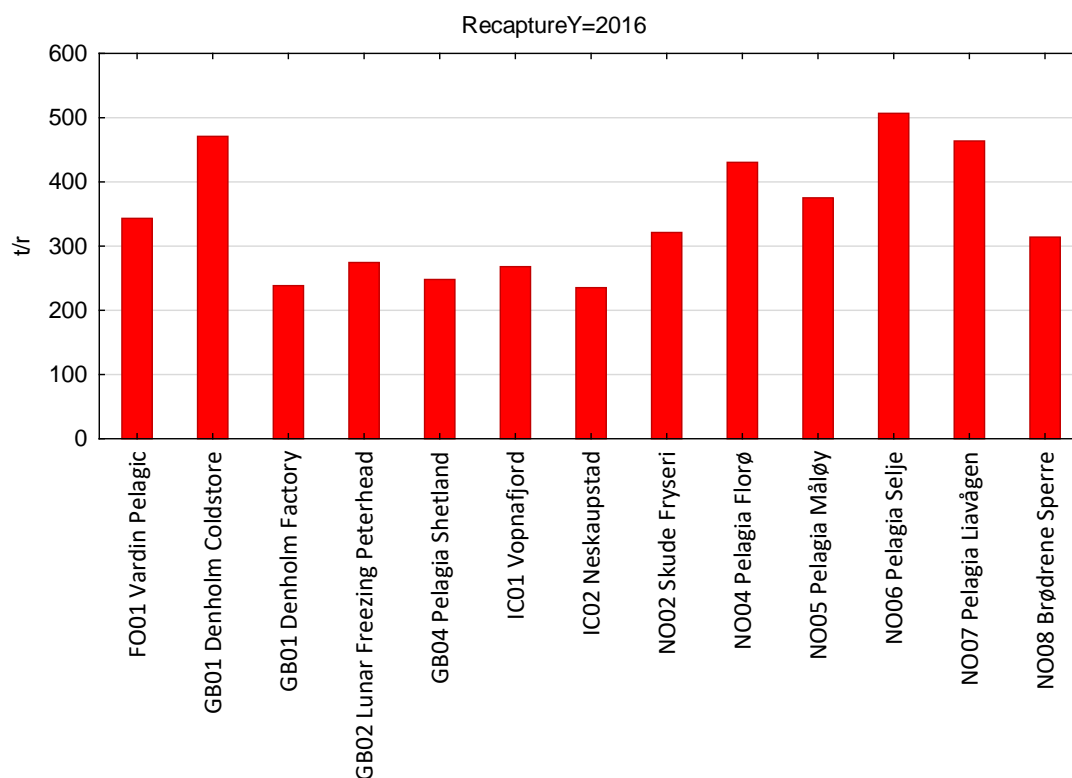


Figure 6. Number of tons scanned per recaptured mackerel (regardless of release year and areas) by factory in 2016.

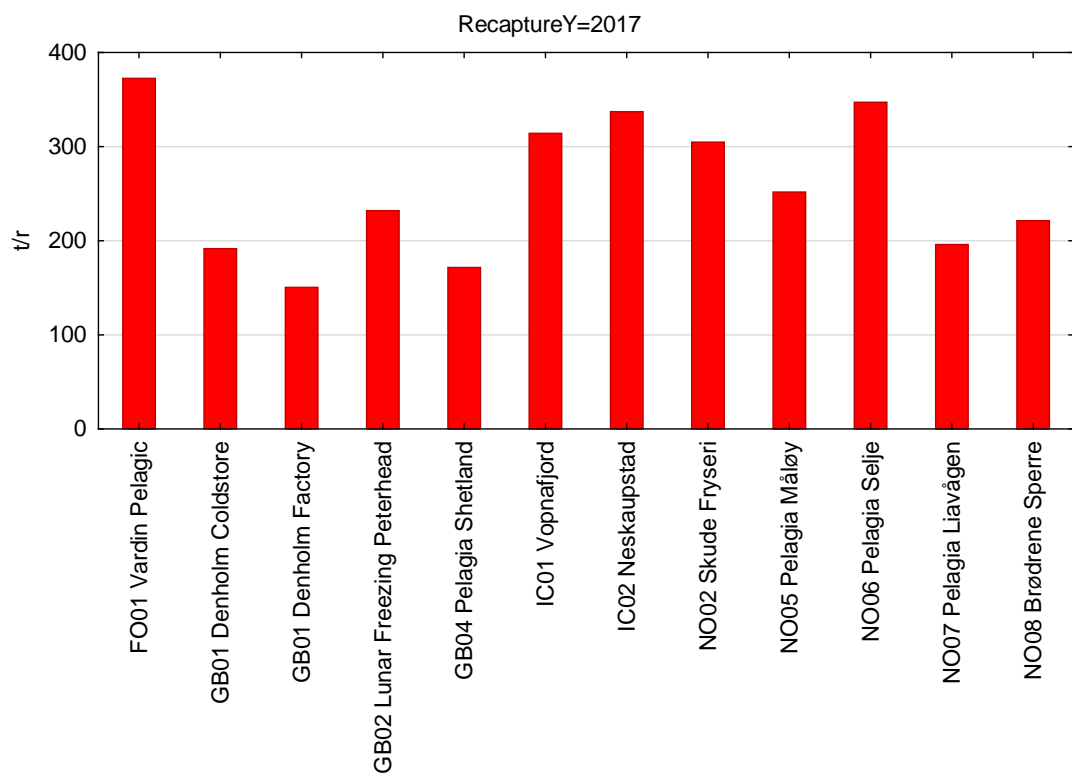


Figure 7. Number of tons scanned per recaptured mackerel (regardless of release year and areas) by factory in 2017.

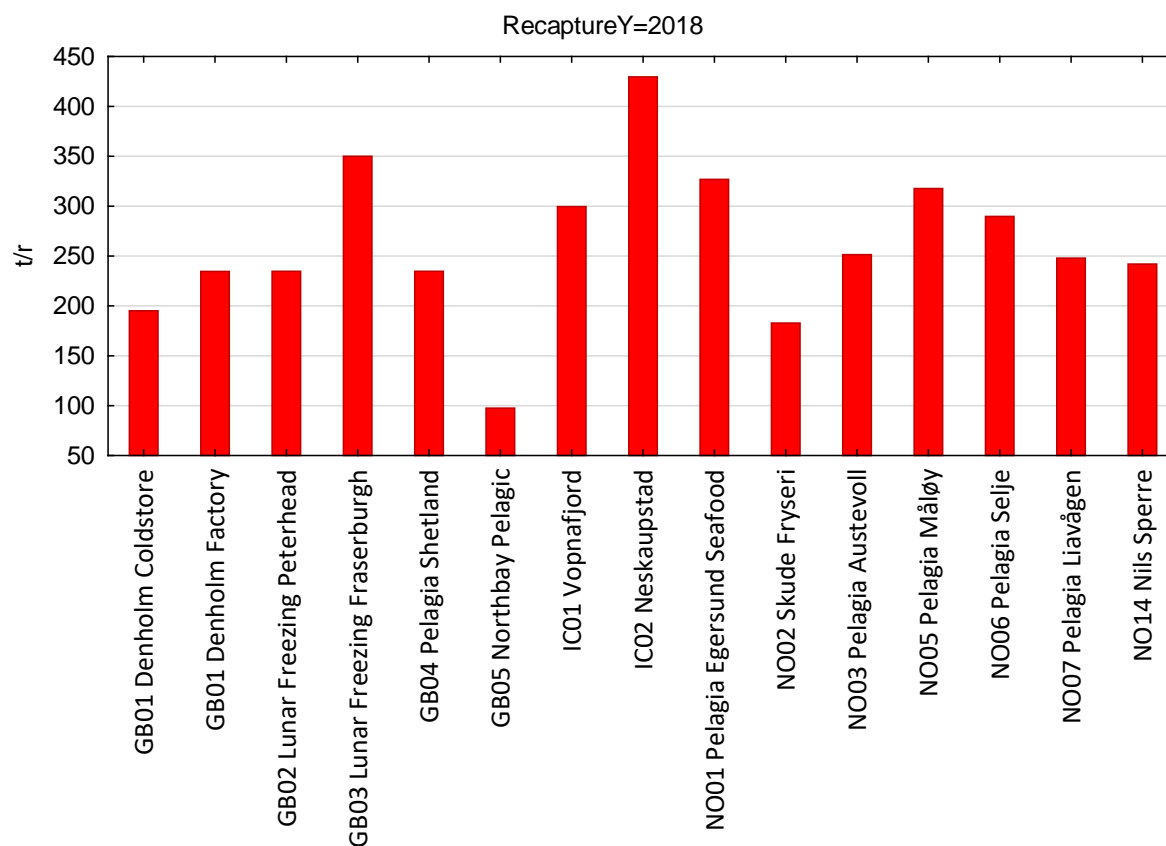


Figure 8. Number of tons scanned per recaptured mackerel (regardless of release year and areas) by factory in 2018.

NEA Mackerel inter-benchmark 2018/19 Working Document

Investigation of changes in apparent ‘survival’ in the mackerel tagging time series data.

Steven Mackinson, Aril Slotte, Andy Campbell, Anna Olafsdottir, Niels Hintzen

1. Rationale

“Leaving out the RFID tagging data from the assessment, or making other assumptions about how these data are used in the assessment, changes the assessment estimates of stock size considerably. A closer investigation of how the model ascribes weights to each data source is required” (ICES advice mackerel 28/9/18).

2. Purpose

- To identify possible causes for the changes in apparent survival of the steel and RFID tagging series.
- To apply the understanding to interpretation of the tagging data and how it is used in the mackerel assessment model.

3. Research questions

- Does the recapture rate of tags meet what is expected based on the concentration of tags released and assumed equal mixing?
- Where it does not, what might be the reasons for this?

4. Approach and methods

This analysis looks at changes in concentration of tags released in to the population and subsequently recaptured in the scanned fishery catches to identify changes in the ‘apparent’ survival behaviour of the steel and RFID tagging data. It then considers possible explanations for this using evidence from correlations to corroborate or refute these.

The principles of mark-recapture experiments assume that the tagged fish mix equally in the population and that after any mortality caused by the tagging procedure, the survival rate of tagged fish is equal to untagged fish. This implies that the concentration of tags released into the population is directly proportional to the concentration of tags recaptured from the fishery catches. And further, the concentration of tags recaptured for each age group should be equal to the concentration of those released – provided the selectivity of the fishery (i.e. the proportion of each age group caught) is the same in the fishery as it is in the tagging survey. i.e.

$$\frac{\text{Number of mackerel tagged survey (tags Released)}}{\text{Total number of mackerel (Population, } N)} = \frac{\text{Number of recaptured tagged mackerel (tags recaptured)}}{\text{Total mackerel catches screened for tags (catch screened, } S)}$$

$$\frac{\text{Released age group } i, \text{ year } j}{N \text{ age group } i, \text{ year } j} = \frac{\text{recaptured age group } i, \text{ release year } j}{\text{Screened age group } i, \text{ release year } j}$$

And, if we assume that the tagging procedure has a constant mortality an index of the expected rate of tag recaptures is

$$r = S * \frac{R}{N}$$

Based on these principles and assumptions, we would expect to see that the ratio of the concentrations $R/N : r/S$ should be about constant over time. Where it departs from this, it implies that any of the following might be the reason:

1. The tags do not mix equally in the population. One possible reason for this could be related to changes in patterns of movements of the stock (or specific components or ages).
2. That some tags might not be available to be caught by the fishery (e.g. outside range). Scanned fish could be different from the from those being caught in the fishery (see WD Sindre Vatnehol 'Trends in age distribution in tag')
3. The selectivity of the fishery means that while all tags might be available to be caught, they are not. (i.e. selectivity of the fishery targeting or overlap with population)
4. Changes in the survival of the tagged fish over time (either short term after the tagging procedure, or longer term) compared to non-tagged fish mean that the concentration of tags available for recapture is no longer the same as that of those released. For example, older fish might shed tags.
5. The numbers of fish scanned could be wrong (see WD Aril Slotte, in progress)
6. Actual survival could have changed (Discussed meeting 1). Previous survival experiments point to expectations for higher survival than estimated in the assessment model, at least in the short-term, and more experiments are planned to try and understand better the survival related to the practices of steel vs RFID tagging.

Table 1 lists the issues investigated in this working document and identifies the data sources used.

Table 2 list the recaptures rates for the steel and RFID tagging data used in the stock assessment.

Table 1. Issues and data sources. **NOTE:** Tagging data are from tag data input files from the WGWIDE 2018 assessment.

Issue	Factors	Information resources
1. Mixing and changes in the population movements	a) Stock size/ rate of population change b) Changes in growth c) Stock distribution – extent and direction.	a) Abundance age 2+ from 2018, 2017 and 2012 stock assessments b) Growth rate (WGWIDE stock weigh at age, and Olafsdottir et al. 2016) c) Centre of gravity and area of stock (Olafsdottir et al. 2018, Lidar data from Utne et al. 2016)
2. Availability of tags to fishery	d) Fishery spatial extent e) Timing of fisheries f) Principal fishing locations	d) Area of fishing (based on area of ICES rectangles with reported catch) (source Andy Campbell, stock coordinator) e) Catches by quarter (source Andy Campbell, stock coordinator) f) Catches by ICES area (source Andy Campbell, stock coordinator)
3. Tag and recapture system	g) Number of factories scanning for tags	g) WGWIDE 2018 and Tenningen, M. Masters thesis

Table 2. Recapture numbers and proportion in each recapture year (data from WGWIDE 2018)

	RFID		Steel	
Recapture yr	Recaptures	Propn recap	Recaptures	Propn recap
Ry+1	1315	56%	713	35%
Ry+2	628	82%	513	60%
Ry+3	289	95%	327	76%
Ry+4	86	98%	210	86%
Ry+5	36	100%	134	92%
Ry+6	5	100%	76	96%
Ry+7	2359		42	98%
Ry+8			20	99%
Ry+9			13	100%
Ry+10			4	100%
Ry+11			3	100%

5. Results

5.1 Comparison of concentration of tags released to those recaptured the following year (i.e '1 year out') and those recaptured after all those fish from a single release year are captured (i.e. 'all years out')

Figure 1 shows that the concentration of released to recaptured steel tags remained fairly constant until around 2007, after which relatively fewer tags were recaptured than released. From 2001, the RFID tagging data is quite consistent with the steel tag data. These patterns were also shown in the working document by T. Brunel (ref: 'Investigating different model configurations to estimate post release survival in the mackerel assessment.docx')

The mackerel assessment uses only steel tag data to 2006 and then RFID data from 2012, thus emphasising the differences in average release : recapture concentration between the two series, which is characterised in the assessment model as tag survival. It is these differences in tag survival, and the influence that the tagging data has on the assessment models predictions that have put the spotlight on the tagging data.

Because tagging survival cannot be measure directly, here we use the difference in the number of releases to number of recaptures divided by the number of releases to represent the changes 'apparent' survival, which is then used to correlate with external variable. Figure 2 shows a decline in apparent survival rate was visible in the steel tagging series from around 2007, which has remained low in the RFID tagging data.

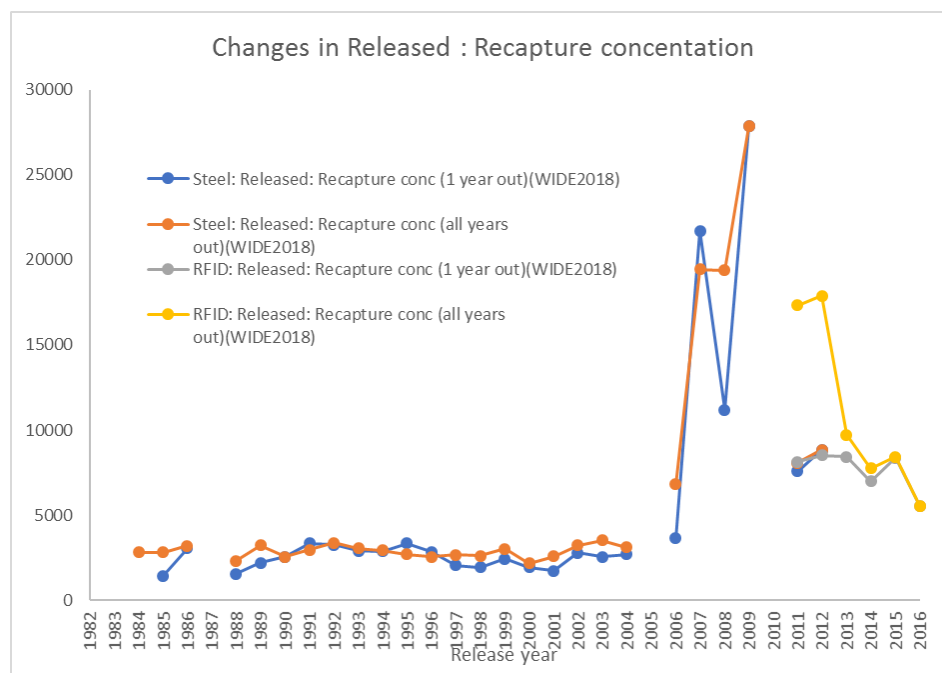


Figure 1. Concentration of tags released compared to the concentration of those released in the following year (1 year out) and after all the tags of that release year are recovered (all years out).

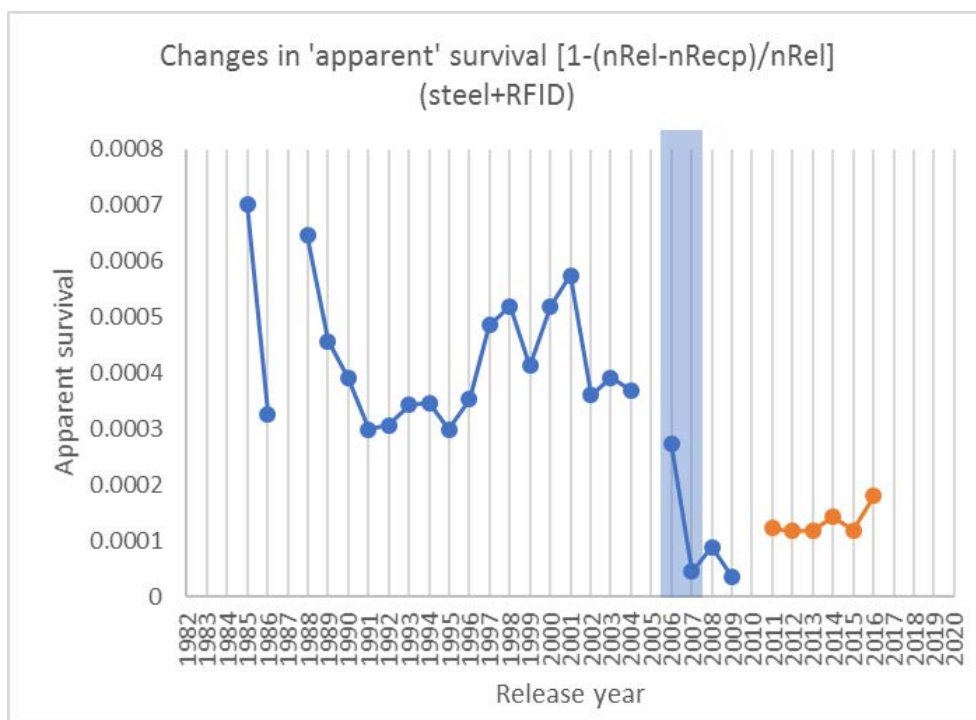


Figure 2. Changes in the apparent survival of tagged fish

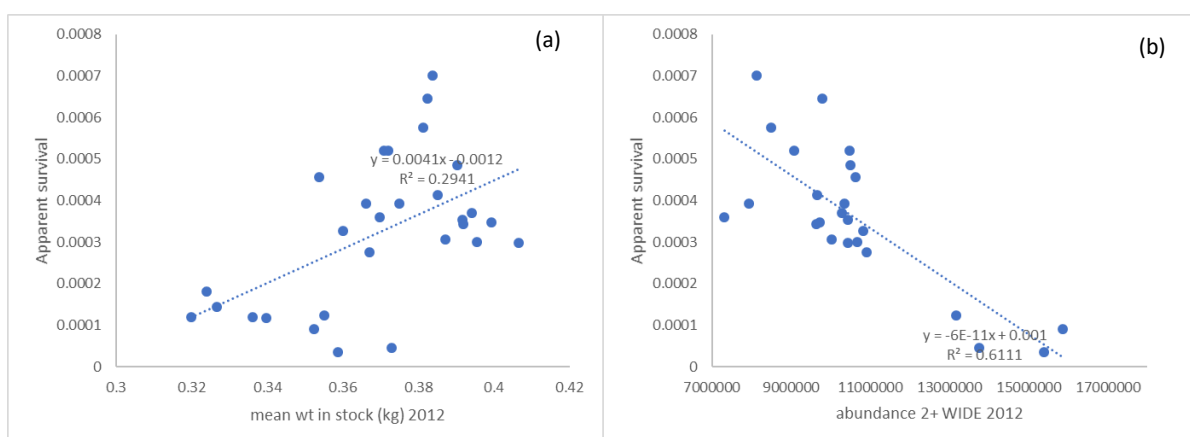
5.2 Correlations between changes in apparent survival and various factors

Factors relating to mixing : stock size, distribution, growth

Timing of the decline of apparent survival are consistent with the increase in stock size and decline in mean weight (Figure 3 & 4), but the most prominent relationship is with the change in the Northern and the Western extent of the stock distribution (Figure 5 and 6a) and the total area of stock distribution (Figure 6d). Note the relationship with the westward centre of gravity of the stock (Figure 6b) is driven by the 2007 data point. Prior to 2007, limited mackerel was detected westward of longitude -10W (Olafsdottir 2018), hence we can assume that prior to 2007 the COG was located further eastward, and that the 2007 data point indicates a point when geographical range expansion began. For this reason, we include it in the correlation analysis¹.



Figure 3. Changes in abundance (2012, 2017, 2018 WGWIDE assessments) and means weight in the stock (from WGWIDE 2018)



¹ The first scientific survey measuring mackerel westward of longitude 10W was in 2007. Even though there is no scientific survey data before 2007, we can with high confidence say that westward COG was to the east of longitude -10W. If years prior to 2007 were included COG would be similar to 2007. (Olafsdottir pers.com.)

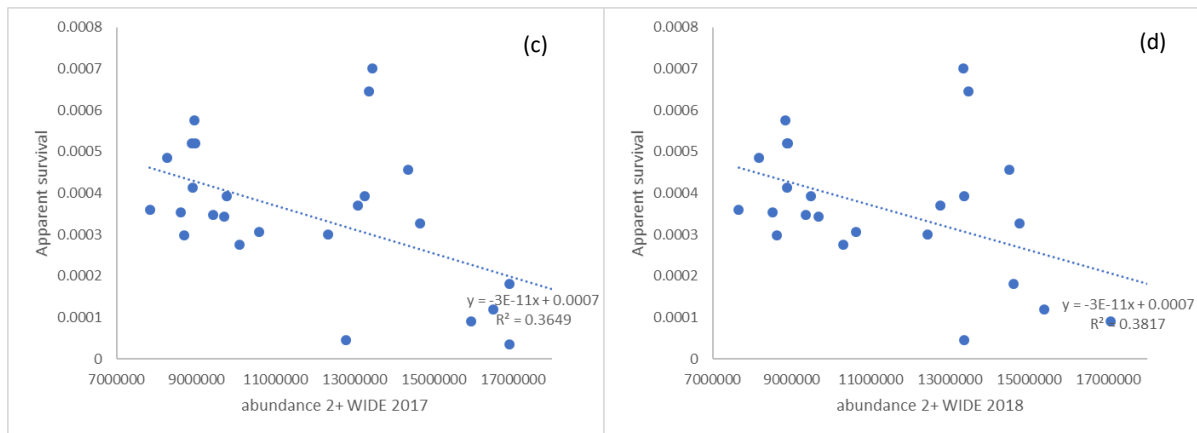


Figure 4. Relationship between apparent survival, mean weight in the stock (from WGWIDE 2008 assessment) (a) and stocks abundance (b-d).

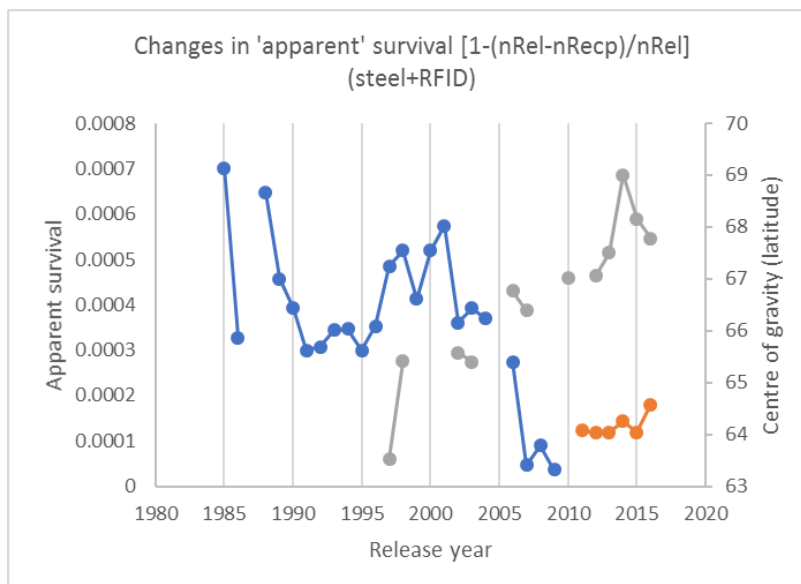


Figure 5. Overlay of apparent survival and change in latitudinal centre of gravity of the stock in the Norwegian Sea (calculated from IESSNS abundance data for the area northward of latitude 62N and eastward of longitude -10W), (years for where information is available).

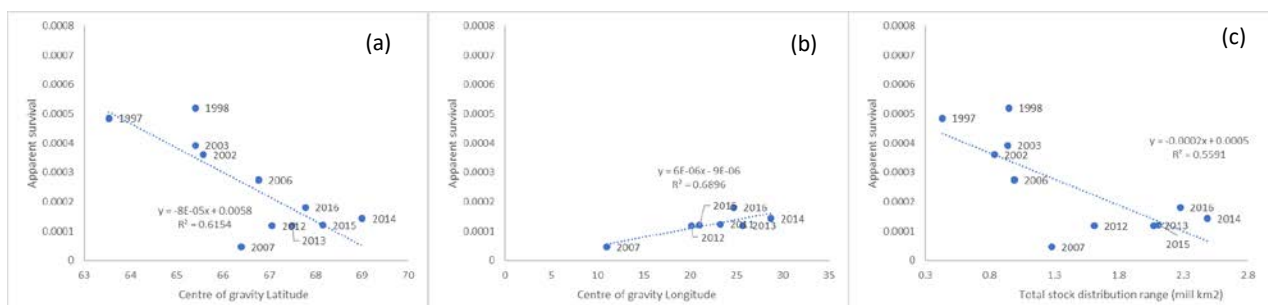


Figure 6. Relationship of apparent survival and latitude of COG in the Norwegian Sea (a; calculated from IESSNS abundance data for the area northward of latitude 62N and eastward of longitude -10W), longitude of COG for the westward expansion area (b; calculated from IESSNS

abundance data for the area westward of longitude -10W), and estimated stock size (c) from 1997 to 2016. For details of COG calculations see Olafsdottir et al (2018).

Factors relating to fisheries: timing, location and extent of fishing

Around 2006/07 the proportion of catches from Q3 increases sharply, mirroring the increasing catches from Area 2a at this time (Figure 7). Increases in Q3, area 2 around this time are balanced by reductions in catches in 4a and Q4. However, neither of these show a strong or convincing relationship with changes in apparent survival of tags (Figure 8, 9).

Because Norway was the only country screening for tags when steel tags were in use, we looked at changes in the extent and location of Norwegian fishing and its relation to changes in apparent survival. While survival does decrease with increase in fishing area, there is no noticeable change coincident with the timing of change in tagging mortality (Figure 10a). However, visually, the western extension of the fishing area is seen to increase after 2006.

The extent of the major mackerel fisheries from all countries (classified as rectangles with catches >1000t) was also considered (Figure 10b), but showed little relation.

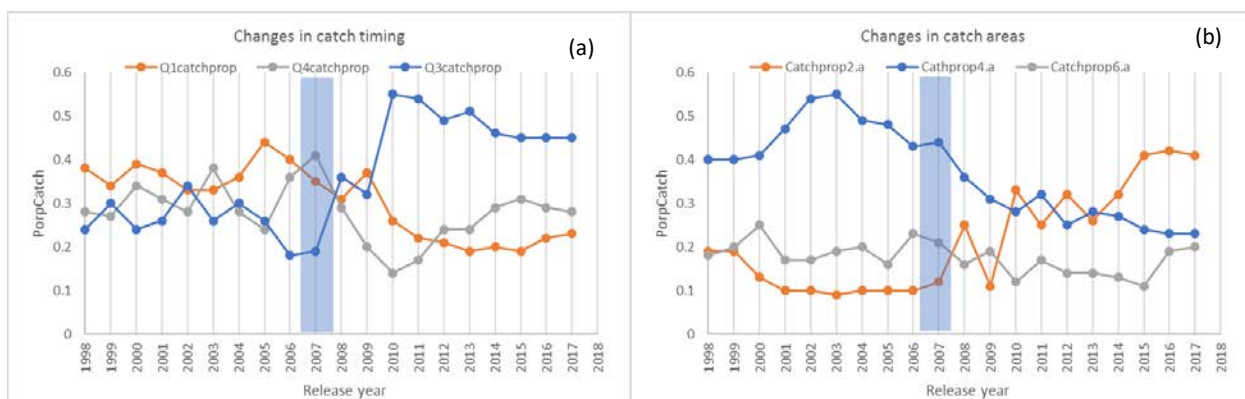


Figure 7. Changes in proportion of catches by quarter (a) and from ICES area (b)

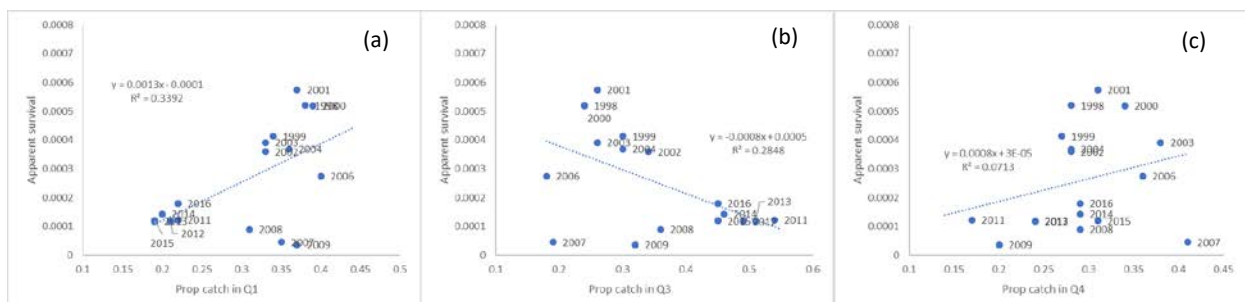


Figure 8. Relationship of apparent survival and proportion of catch by quarter

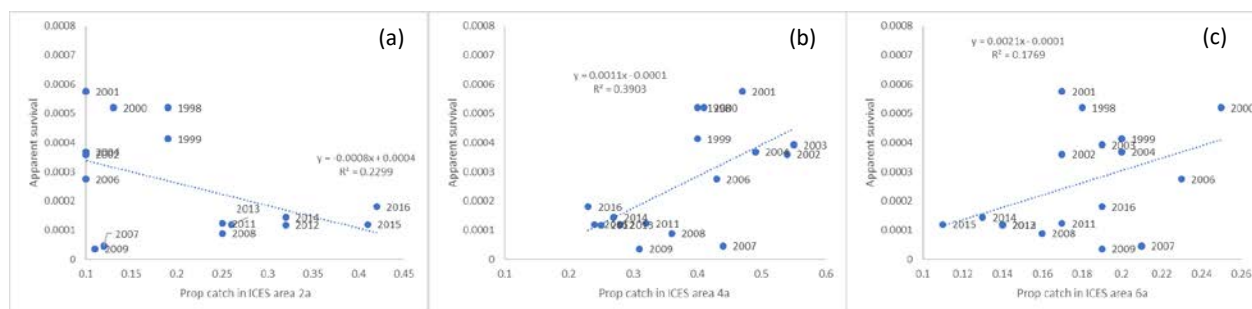


Figure 9. Relationship of apparent survival and proportion of catch by ICES area

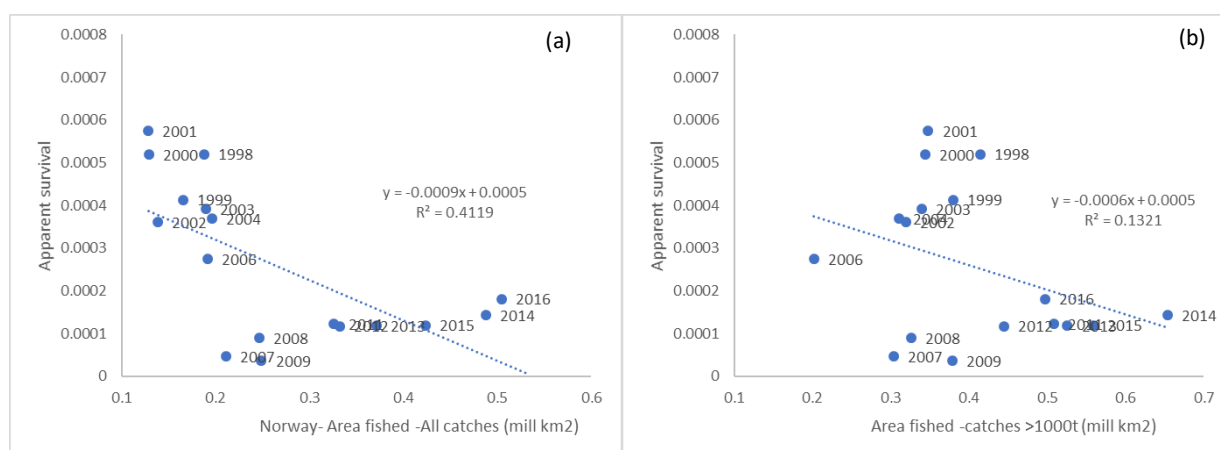
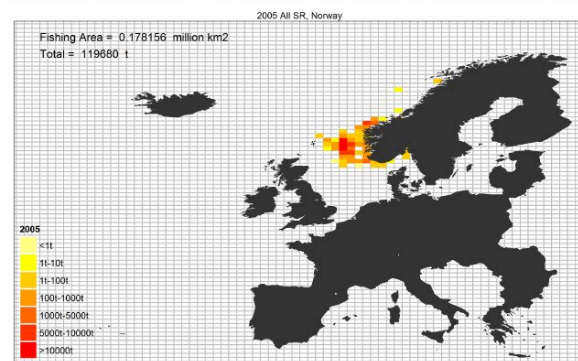
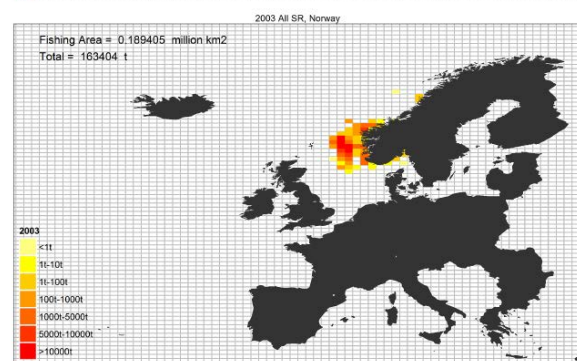
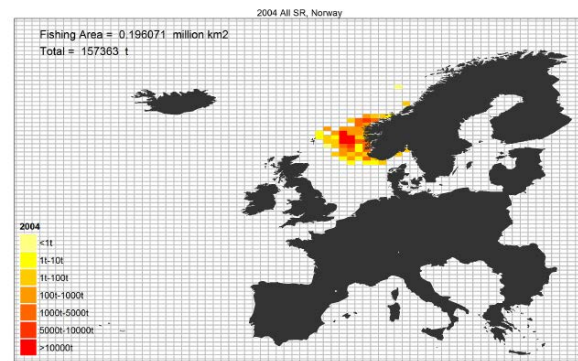
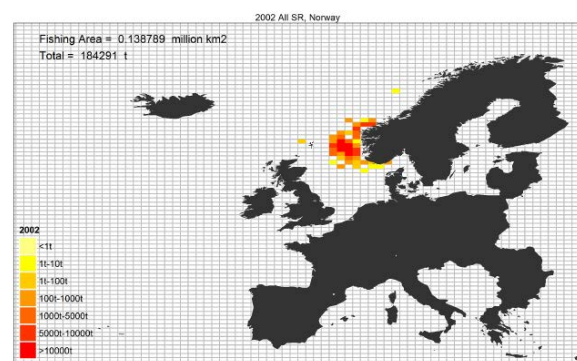
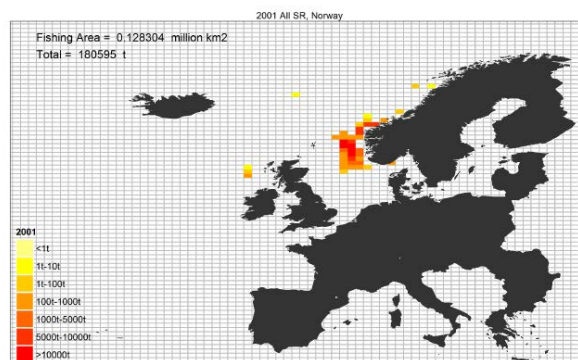
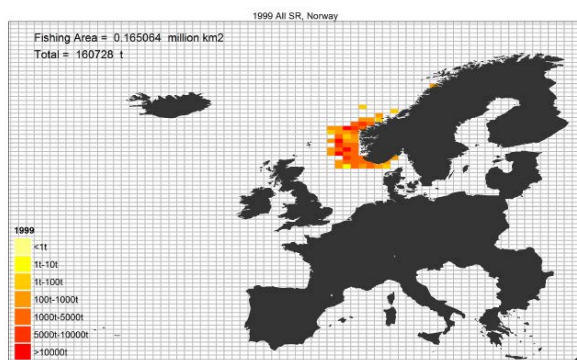
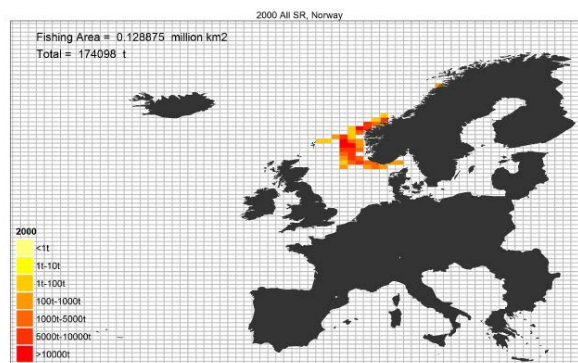
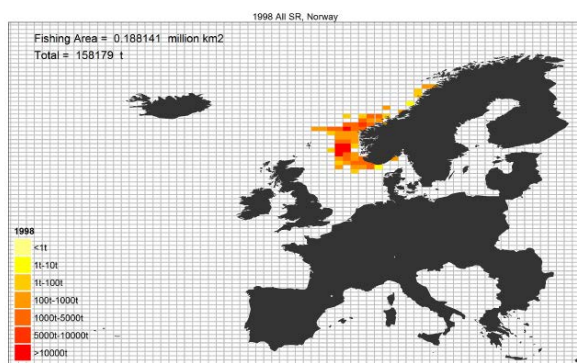
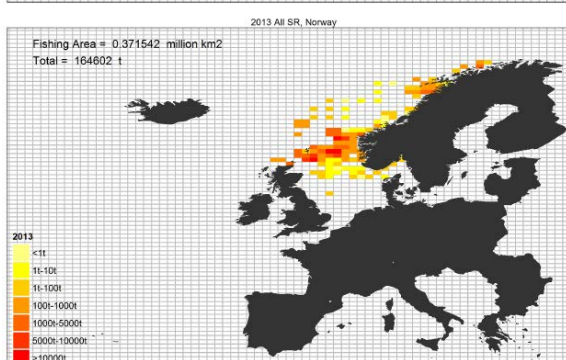
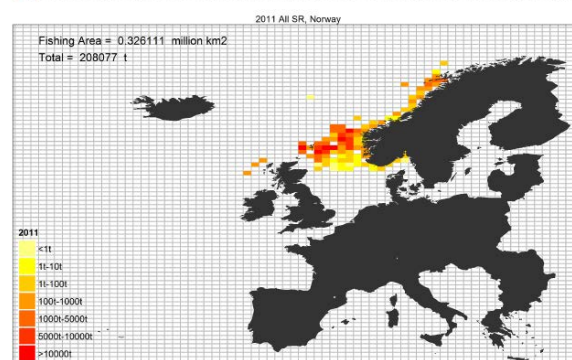
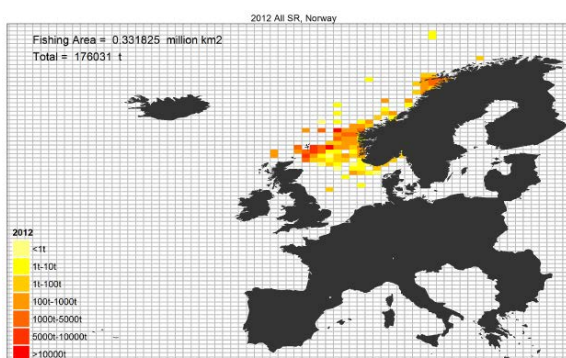
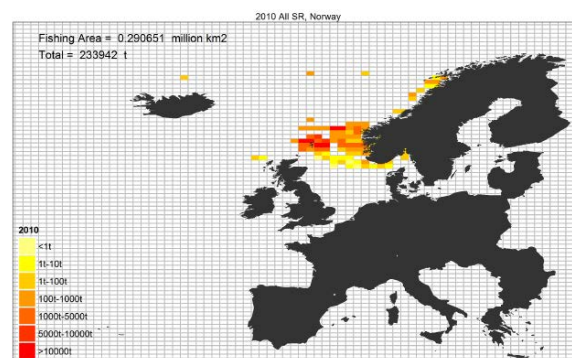
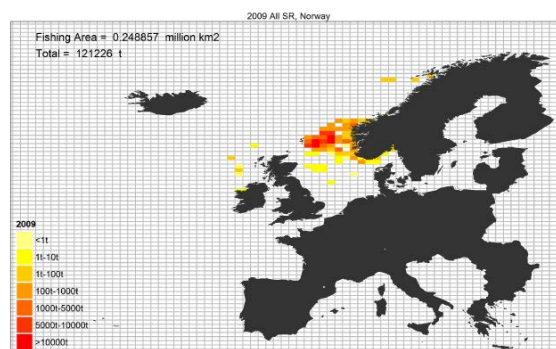
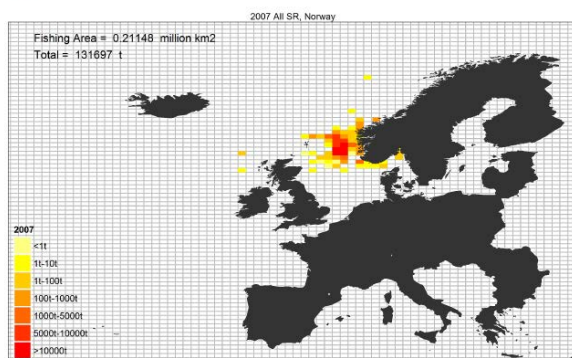
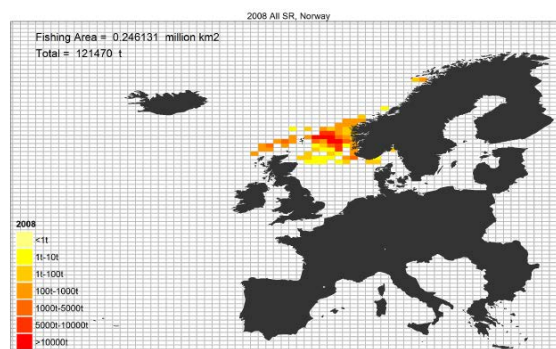
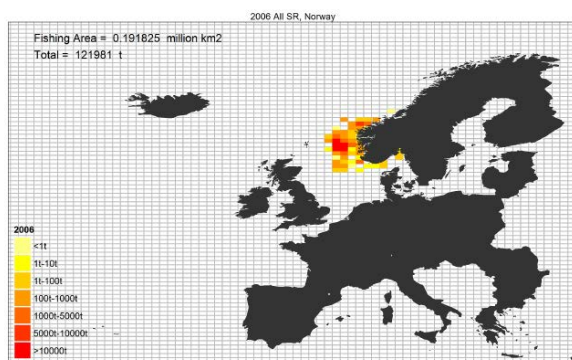


Figure 10. Relationship of apparent survival and area fish by Norway (a) and all countries where catches > 1000t (b)





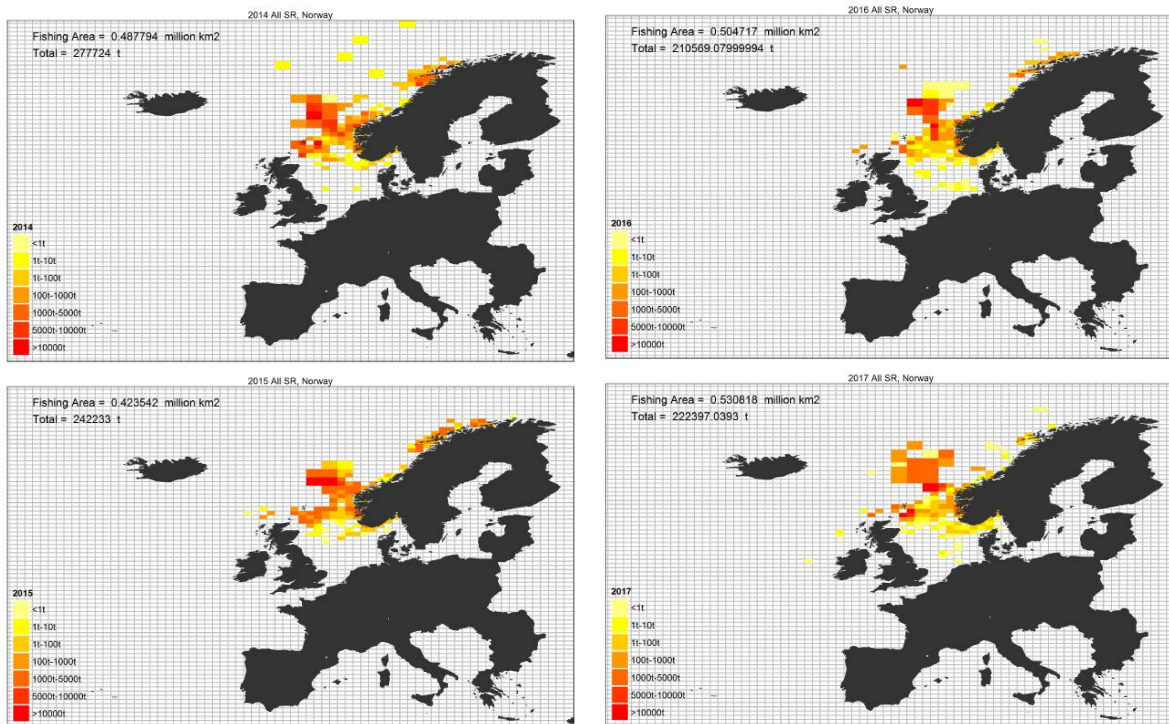


Figure 11. Distribution of Norwegian mackerel fishing

Factors relating to tag recapture systems: number of factories scanning for tags

During the steel tagging time series, 1-4 Norwegian factories were involved in tag recovery (Tenningen, Masters thesis). Since the RFID tagging has been in place the number of factories (and countries) involved has increased rapidly leading to a large change in the distribution and abundance of catches scanned for tags (see WGwide 2018, Figure 8.6.4.4). From 2014 onward, this amounted to a tripling of the catch scanned. Information on the number of factories scanning is absent from 2005-2012, so no meaningful comparison can be made with the data on changes on apparent survival (Figure 12). WGwide, p290 notes that differences in mixing could cause changes in the recapture rates in different areas which may affect the way the tagging data is used in the model. An investigation into RFID tag survival rates in different areas that undertaken during WGwide 2018 showed small difference between areas, but it could not account for the difference between steel tag and RFID survival rate estimates by the SAM model (see presentation from the IBP 'Tag data – potential issues for benchmark 2019', Aril Slotte).

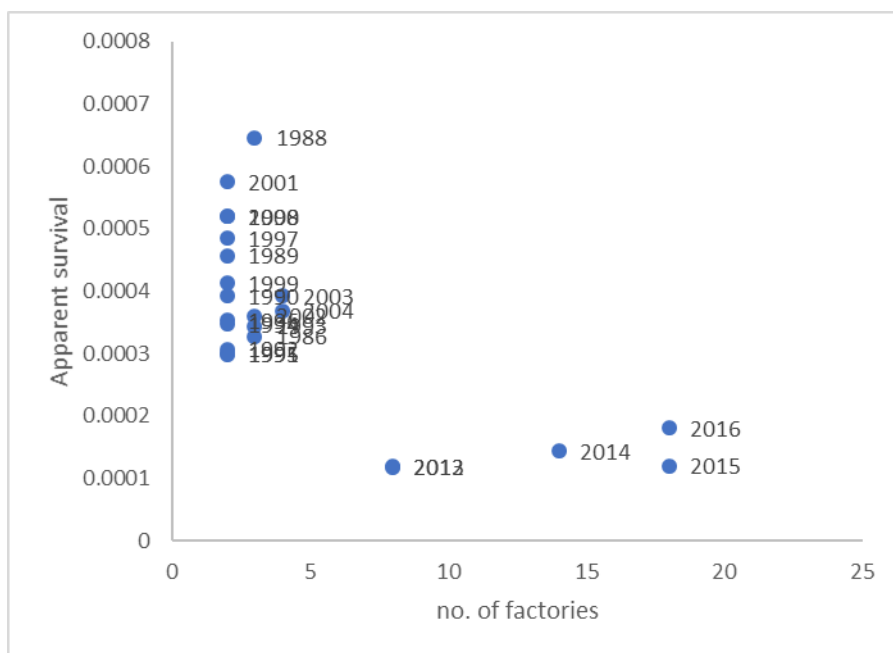


Figure 12. Apparent survival plotted against number of factories scanning for tagged fish.

Table 3 summarises the correlations of a external factors with 'apparent' tag survival

Table 3. Correlations with apparent tag survival

	<i>Steel+RFID apparent survival (1 year out)</i>
Steel+RFID apparent survival (1 year out)	1
No. factories tag scanning	-0.70
Abund (2+) WIDE2012	-0.78
Abund (2+) WIDE2017	-0.60
Abund (2+) WIDE2018	-0.62
mean stock weight (2-13, WGWIDE 2018)	0.54
Stock area (Lidar)	-0.17
Core area (Lidar)	-0.24
cog_latitude_24	-0.78
longitude_12_cog West	0.83
total stock distribution range (million km^2)	-0.75
Q1catchprop	0.58
Q3catchprop	-0.53
Q4catchprop	0.27
Catchprop2.a	-0.48
Cathprop4.a	0.62
Catchprop6.a	0.42
FishAreaOver1t(Jul-Aug)	-0.73
FishAreaOver10t(Jul-Aug)	-0.65
FishAreaOver100t(Jul-Aug)	-0.64
FishAreaAllRects (Jul-Aug)	-0.78
Norway_FishAreaAllRects	-0.64
FishArea_All (>1000t catches)	-0.36

6. Conclusion

Data on the concentration of release:recaptures of steel tags indicate that apparent survival of tags was decreasing before the arrival of the RFID tagging data and thus the issue of interest is not simply why the RFID tags have lower survival than steel tags. Rather, it is about how and why apparent survival of tags has changes over time.

Most notably, changes in the apparent survival of tagged fish occurring between 2006 and 2007 are correlated to, and coincident with changes in the stock distribution, with survival declining as the stock expanded in area and was distributed further to the North and expanding westward into areas previously not occupied. Changes are also coincident with changes in the fishery distribution and timing of fisheries, but these signals are less clear because of the diversity of fishing activities. The combination of these factors may have influence both the mixing rates and the recapture rates as the overlap of the fishery with the core area of the stock is changes.

A closer look at Norwegian catches during the steel tagging period shows changes in westward extent of fishing taking place when apparent tag survival declined. A more detailed investigation of the how changes apparent survival relate to the changes in fishing distribution of those nations involved in scanning for tagging mackerel is needed but at this time, the RFID time series is insufficient to do this justice.

Changes in the stock and fishery complicate application of the 'equal mixing assumption' necessary to estimate abundance from tagging data. At the same time it reveals important information about the dynamics of the stock and its representation through the tag 'survival' parameter used in the SAM stock assessment model. How to address this without the need to have to estimate an annual survival parameter and thus risk over-parameterising the assessment model is a point for discussion now and in the future when more is learned from the RFID tagging data.

References

- Olafsdottir, A. H., Slotte, A., Jacobsen, J. A., Oskarsson, G. J., Utne, K. R., and Nøttestad, L. 2016. Changes in weight-at-length and size at-age of mature Northeast Atlantic mackerel (*Scomber scombrus*) from 1984 to 2013: effects of mackerel stock size and herring (*Clupea harengus*) stock size. – *ICES Journal of Marine Science*, 73: 1255–1265.
- Olafsdottir, A. H., Utne, K. R., Jacobsen, J. A., Jansen, T., Óskarsson, G. J., Nøttestad, L., ... Slotte, A. (2018). Geographical expansion of Northeast Atlantic mackerel (*Scomber scombrus*) in the Nordic Seas from 2007 – 2016 was primarily driven by stock size and constrained by low temperatures. *Deep-Sea Research. Part 2: Topical Studies in Oceanography*. DOI: 10.1016/j.dsr2.2018.05.023
- Utne, K.J., Geir Huse , Geir Ottersen , Jens Christian Holst , Vladimir Zabavnikov , Jan Arge Jacobsen , Guðmundur J. Óskarsson & Leif Nøttestad (2012) Horizontal distribution and overlap of planktivorous fish stocks in the Norwegian Sea during summers 1995–2006, *Marine Biology Research*, 8:5-6, 420-441, DOI: 10.1080/17451000.2011.640937.
- Tenningen, M. (previously Antsalo, M). 2006. Abundance estimation of the Northeast Atlantic mackerel with use of Norwegian tag data. Master Thesis, Bergen University.

Note SAM configuration for Mackerel

The basic state-space assessment model (SAM) is described in Nielsen & Berg (2014). The model has been continuously developed and adapted for different stocks. The current implementation (<https://github.com/fishfollower/SAM>) is an R-package based on Template Model Builder (TMB) (Kristensen et al. 2016).

The data set used to assess Mackerel include catches-at-age $(C_{a,y})_{a=0...12+, y=1980...2017}$, and egg-based SSB index (every third year from 1992), a recruitment index (from 1998-2015), an age-specific swept-area index (of ages 3-11 in years 2010, 2012-2018), and finally tagging data is used (more details later). In addition to these data a set of biological parameters are available, these are: Mean weight in stock, mean weight in catch, proportion mature, and an estimate of natural mortality.

Model as configured by last benchmark

The model for Mackerel is a state-space model. The states α are the log-transformed stock sizes $\log N_0, \dots, \log N_{12+}$ and fishing mortalities $\log F_0, \dots, \log F_{7-12+}$ corresponding to total age specific catches. Notice that it is assumed that $F_7 = \dots = F_{12+}$. In any given year y the state is the combined vector $\alpha_y = (\log N_0, \dots, \log N_{12+}, \log F_0, \dots, \log F_{7-12+})'$. The transition equation describes the distribution of the next years state from a given state in the current year. The following is assumed:

$$\alpha_y = T(\alpha_{y-1}) + \eta_y$$

The transition function T is where the stock equation and assumptions about stock-recruitment enters the model. The equations are:

$$\begin{aligned} \log N_{1,y} &= \log(N_{1,y-1}) \\ \log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} , & 1 \leq a < A \\ \log N_{A,y} &= \log(N_{A-1,y-1} \exp^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} \exp^{-F_{A,y-1} - M_{A,y-1}}) & A = 12 \\ \log F_{a,y} &= \log F_{a,y-1} , & 0 \leq a \leq 7 \end{aligned}$$

Here $M_{a,y}$ is the age and year specific natural mortality parameter, which is assumed known from outside sources. $F_{a,y}$ is the total fishing mortality.

The prediction noise η is assumed to be Gaussian with zero mean, and three separate variance parameters. One for recruitment ($\sigma_{N_{a=1}}^2$), one for survival ($\sigma_{N_{a>1}}^2$), one for fishing mortality at age (σ_F^2). The elements of η are assumed uncorrelated.

The observation part of the state-space model describes the distribution of the observations for a given state α_y . Here the vector of all observations from

a given year y is denoted x_y . The elements of x_y are age-specific log-catches $\log C_{a,y}$, age-specific log-indices from scientific surveys $\log I_{a,y}^{(s)}$, biomass survey $\log B_y^{(s)}$, and the recaptures from the tagging experiments $r_y^{(j)}$. The expectation in the combined observation equation is:

$$Ex_y = O(\alpha_y)$$

The observation function O consists of the catch equations for total catches and scientific surveys. An expanded view of the observation equation becomes:

$$\begin{aligned}\log C_{a,y} &= \log \left(\frac{F_{a,y}}{Z_{a,y}} (1 - e^{-Z_{a,y}}) N_{a,y} \right) \\ \log I_{a,y}^{(s)} &= \log \left(Q_a^{(s)} e^{-Z_{a,y} \frac{D^{(s)}}{365}} N_{a,y} \right) \\ \log B_y^{(s)} &= \log \left(Q^{(s)} SSB_y \right) \\ r_{ay}^{(j)} &= p_{\text{surv}}(\text{tagtype}^{(j)}) n_{\text{scan},a,y} \frac{R^{(j)}}{N^{(j)}}\end{aligned}$$

Here Z is the total mortality rate $Z_{a,y} = M_{a,y} + F_{a,y}$, $D^{(s)}$ is the number of days into the year where the survey s is conducted, $Q_a^{(s)}$ are model parameters describing catchability coefficients. For the one age-specific survey it is assumed that the catchability is the same for the two oldest ages. $B_y^{(s)}$ is the index for the biomass survey. There are two tagging data series here referred to as tag-types. The number of tag-recoveries $r_{ay}^{(j)}$ from age a in year y from the j 'th release event is predicted by the number scanned $n_{\text{scan},a,y}$ times the fraction of released $R^{(j)}$ at the j 'th release event into the cohort $N^{(j)}$ times the probability of surviving the tagging event $p_{\text{surv}}(\text{tagtype}^{(j)})$. This probability is estimated separately for each of the two tag-types.

The observations from the catch-at-age, egg-based-ssb index, and the recruitment index are all assumed independent and normal at the logarithmic scale. The logarithm of the swept-area age-specific index is assumed normal with an AR(1) correlation structure across ages. The tag returns are assumed to follow a negative binomial distribution with an estimated over-dispersion parameter for each of the two tag-types.

0.1 Model with correlated tags

The recaptures $r_{ay}^{(j)}$ are all assumed independent and to follow a negative binomial distribution. To study the effect of this assumption the model was extended to allow tag returns to be correlated. This was done by introducing

a random effect ε in prediction of the recaptures. The recapture prediction is changed to:

$$r_{ay}^{(j)} = p_{\text{surv}}(\text{tagtype}^{(j)})n_{\text{scan},a,y} \frac{R^{(j)}}{N^{(j)}} e^{\varepsilon(G_{a,y}^{(j)})} \quad , \quad \text{where} \quad \varepsilon(G_{a,y}^{(j)}) \sim N(0, \sigma_G^2)$$

here G is a grouping factor, which can be used to define the groups within which the recaptures are allowed to be correlated. Groups are defined based on shared release event, shared recapture event, or shared release and recapture event.

Model	log(L)	#par	AIC
Spaly	-2785.8	23	5617.6
Spaly common over-dispersion	-2785.9	22	5615.8
Correlated tag-release	-2770.8	23	5587.7
Correlated tag-recapture	-2766.7	23	5579.4
Correlated release×recapture	-2766.9	23	5579.8

Table 1: Summary of model runs with different tag-correlation configurations.

When running the models with correlated tags it was no longer possible to estimate a separate over-dispersion for each tag-type, so the spaly model (from last benchmark) was first compared to a model with common over-dispersion for both tag-types. This was an acceptable model reduction (Tab. 1). From the three different correlated tag runs grouping according to shared recapture gave the optimal likelihood, but all three tag-groupings gave a significant improvement over the uncorrelated tag model (Tab. 1).

Model with release time set to spawning time

The spaly model (from last benchmark) used the beginning of the year estimated numbers in the cohort, as the denominator in the equation to predict the recaptures. It was suggested that since the releases are done at spawning time, then these estimated numbers should be adjusted to be numbers at spawning time. The adjustment to the prediction of releases are:

$$r_{ay}^{(j)} = p_{\text{surv}}(\text{tagtype}^{(j)})n_{\text{scan},a,y} \frac{R^{(j)}}{N^{(j)} e^{-F\text{pf} - M\text{pm}}}$$

Here F is natural mortality, M is natural mortality, pf proportion F applied before spawning time, and pm is proportion of M applied before spawning. All of these are corresponding to age and year of the relevant cohort.

The model with spawning time adjusted did not provide an improved fit to the observations (Tab. 2).

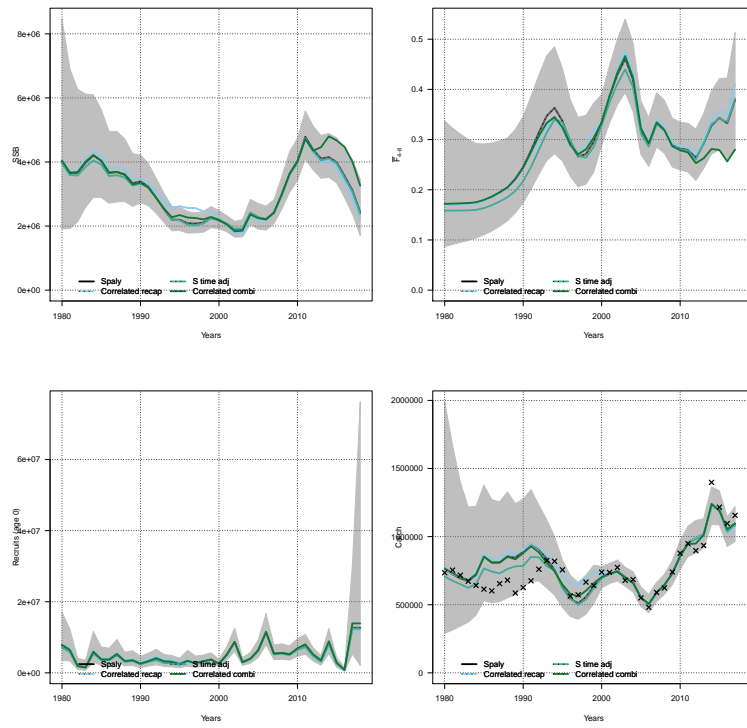


Figure 1: Spawning stock biomass, average fishing mortality, recruitment, and estimated catch for four different model options compared.

Model	log(L)	#par	AIC
Spaly common over-dispersion	-2785.9	22	5615.8
Spawning time adjusted N	-2790.2	22	5624.4

Table 2: Summary of model run with spawning time adjusted N .

Spawning time adjusting tag-releases did not change the main results of the model (Fig. ??) and the likelihood deteriorated (Tab. 2). Allowing correlation to the tag releases did improved the likelihood, and it appeared from the that the release event was the important event to allow correlation within. The model results are however similar between the spaly run and the run with correlated within release event (Fig. ??). The model with correlation within all combinations of release and recapture events did not improve the likelihood compared to the model with correlations within release, but the results were substantially different.

1 Rounding

The later tagging series are not giving tagging returns as integers. To explore the maximal potential effect of this rounding 0.5 is added to all new tag returns, which will insure that they are rounded upwards.

References:

- Berg, C.W., A. Nielsen 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES Journal of Marine Science: Journal du Conseil 73 (7), 1788-1797
- Berg C. W., A. Nielsen, K. Kristensen 2014. Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. Fisheries Research , 151: 91-99.
- Kristensen, K, A. Nielsen, C.W. Berg, H.J. Skaug, B. Bell. 2016. TMB: Automatic differentiation and laplace approximation. Journal of Statistical Software 70 (5), 1-21
- Nielsen, A. and C.W. Berg 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research 158, 96-101
- Thygesen, U.H., C.M. Albertsen, C.W. Berg, K. Kristensen, and A. Nielsen 2017. Validation of state space models fitted as mixed effects models. (Subm. EES).

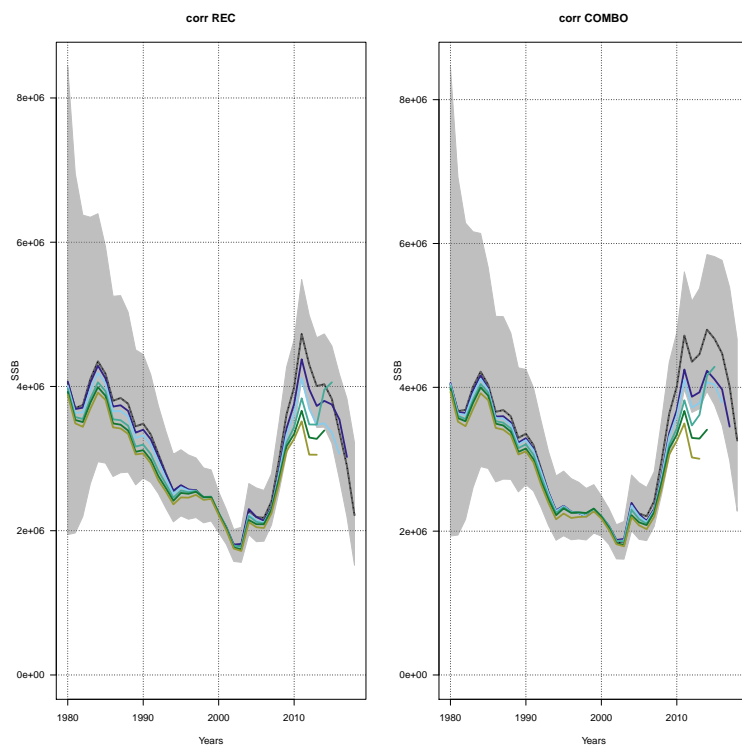


Figure 2: Spawning stock retrospective for the correlated release and the correlated combo run.

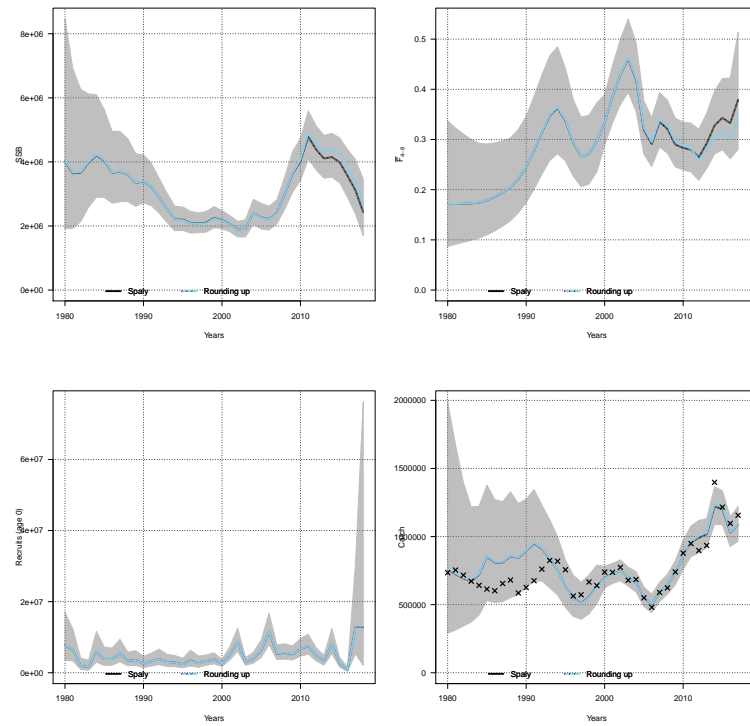


Figure 3: Spawning stock biomass, average fishing mortality, recruitment, and estimated catch for standard rounding and always upwards rounding compared.

NEA Mackerel Catch Sampling

Wageningen Marine Research

4-7 March, 2019

Andrew Campbell

Fisheries and Ecosystems Advisory Services,

Marine Institute,

Galway, Ireland

Introduction

The current assessment method for NEA mackerel is age-based. In support of this, ICES issues a data call detailing the data reporting requirements. Using information gathered during catch monitoring and national sampling programs, each nation submits data to ICES, ideally in the requested format. For NEA Mackerel this includes quarterly estimates of total catch by ICES division. For sampled catches, estimates of catch abundance at age along with mean weight and length at age are also provided along with an indication of the level of sampling undertaken. Sampling levels for mackerel are generally high with typically over 85% of the total catch sampled annually.

The design and operation of sampling schemes and the resulting calculation of catch numbers at age is a national competence, to which the WG currently has little visibility. Upon submission of catch and sampling information, it is the responsibility of the stock coordinator to compile a single annual estimate of catch at age from the stock for input to the stock assessment. This exercise will involve the allocation of samples to unsampled (generally relatively minor) catches. Aggregated length frequency profiles are also provided and are used by the stock coordinator as an aid to allocation. LF data is usually provided on a quarterly basis, occasionally disaggregated by ICES division but often combined.

During preparation for the current inter-benchmark exercise, it was noticed that data on proportions at age appeared to be particularly inconsistent for certain fleet and area combinations. The Scottish, Irish and Norwegian fleets operating in ICES division 4a in quarter 4 target similar aggregations and have historically reported similar catch at age profiles with high levels of catch sampling. However, in 2017 this does not appear to be the case.

This WD describes a short exercise to explore these data and the possible implications of potential error on the stock assessment.

WG Data Submissions

The data submissions of Ireland, Scotland and Norway for division 4a, Q4 in 2017 are shown below

Age	Norway			Scotland			Ireland		
	Num	Wgt	Len	Num	Wgt	Len	Num	Wgt	Len
	(000s)	(kg)	(cm)	(000s)	(kg)	(cm)	(000s)	(kg)	(cm)
0	0	NA	NA	0	NA	NA	0	NA	NA
1	3710	0.227	29.7	3212	0.185	28.5	1618	0.177	28.3
2	260	0.271	31.5	15313	0.291	32.6	1270	0.274	32.2
3	18510	0.308	32.8	31303	0.324	33.7	16699	0.310	33.4
4	5050	0.339	33.8	34942	0.370	35.0	7883	0.368	35.2
5	8410	0.368	34.7	32991	0.376	35.2	14264	0.386	35.7
6	17770	0.394	35.5	26926	0.404	36.0	19463	0.400	36.1
7	18070	0.417	36.2	26383	0.427	36.6	12429	0.432	36.9
8	13480	0.439	36.8	14677	0.472	37.8	9140	0.470	37.9
9	11010	0.460	37.3	7686	0.489	38.2	4459	0.484	38.2
10	9370	0.480	37.8	5815	0.483	37.9	3034	0.498	38.5
11	4040	0.499	38.3	2390	0.505	38.7	879	0.516	39.0
12	2280	0.516	38.7	1301	0.554	39.8	182	0.533	39.3
13	530	0.534	39.1	409	0.605	40.8	38	0.634	41.5
14	120	0.550	39.5	0	NA	NA	104	0.494	38.5
15+	140	0.616	41.0	0	NA	NA	0	NA	NA
Catch	45kt			78kt			36kt		
Samples	12			21			17		
Measured	358			2299			2954		
Aged	358			591			776		

Table 1: WGWIDE data submission, ICES division 4a, Q4, 2017.

Catches from all three fleets in 2017 were substantial and well sampled.

The distribution of catch in each month of the quarter is shown in figure 1.

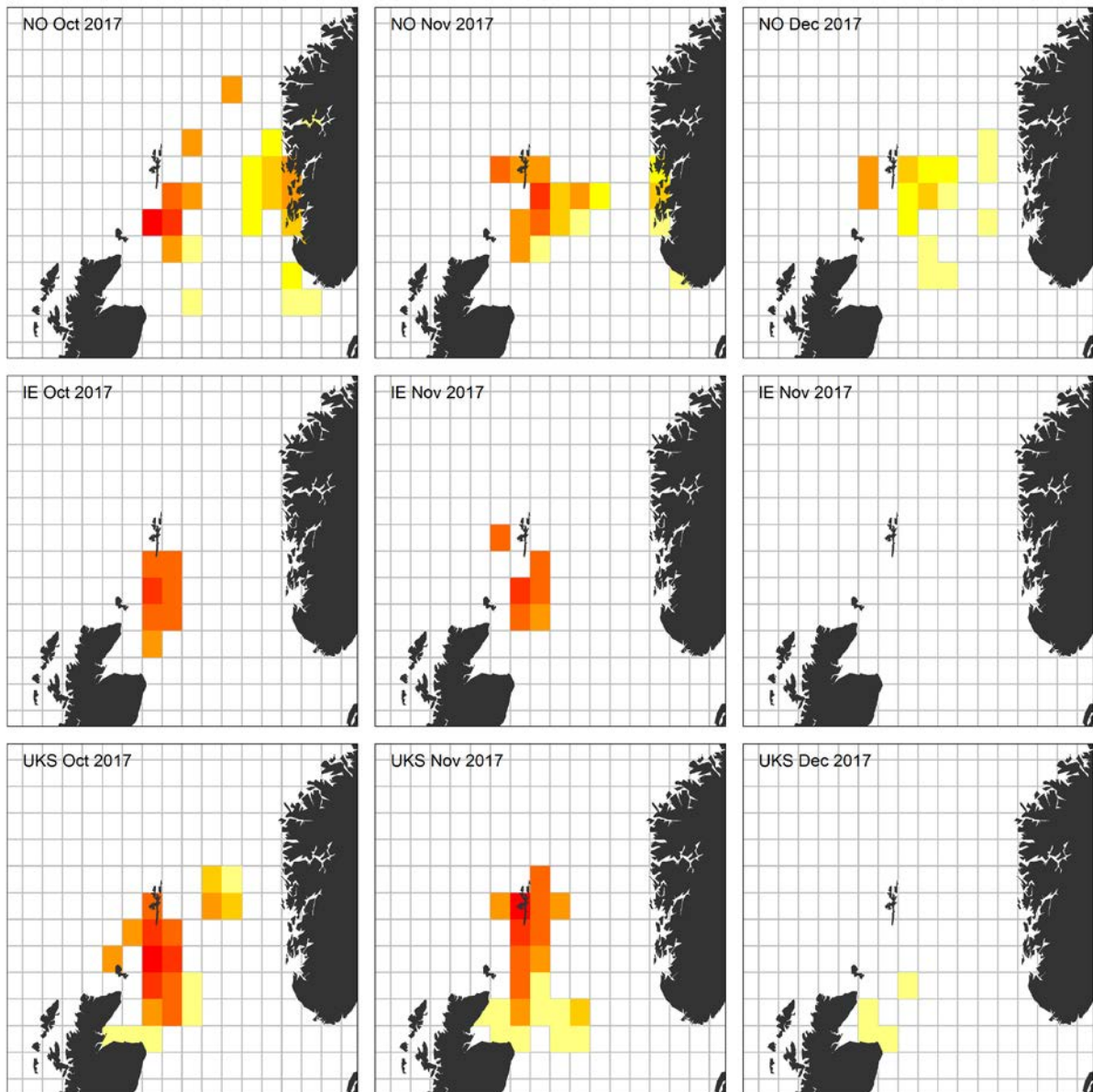


Figure 1: Ireland, Scotland and Norway catch by SR by Month, 4a, Q4, 2017.

For all fleets the bulk of the catch is taken in October and November, south of Shetland. The Norwegian fleet has additional catch from their coastal waters although these are minor compared to those catches further west. The Scottish fleet has access inside the 12-mile limit with significant catches taken close to Shetland. The Norwegian fleet operates primarily purse seiners whereas large modern RSW pelagic trawl vessels comprise the Scottish and Irish fleets. Although the contribution to the overall catch from this component of the fishery has reduced in recent years with the northern expansion of the distribution and fishery, the winter fishery in 4a (which continues into 6a in the first quarter for Irish and Scottish vessels) is a significant component of the total catch.

The (normalised) annual length frequency data submitted by Scotland and Ireland for 2012-2017 is shown in figure 2.

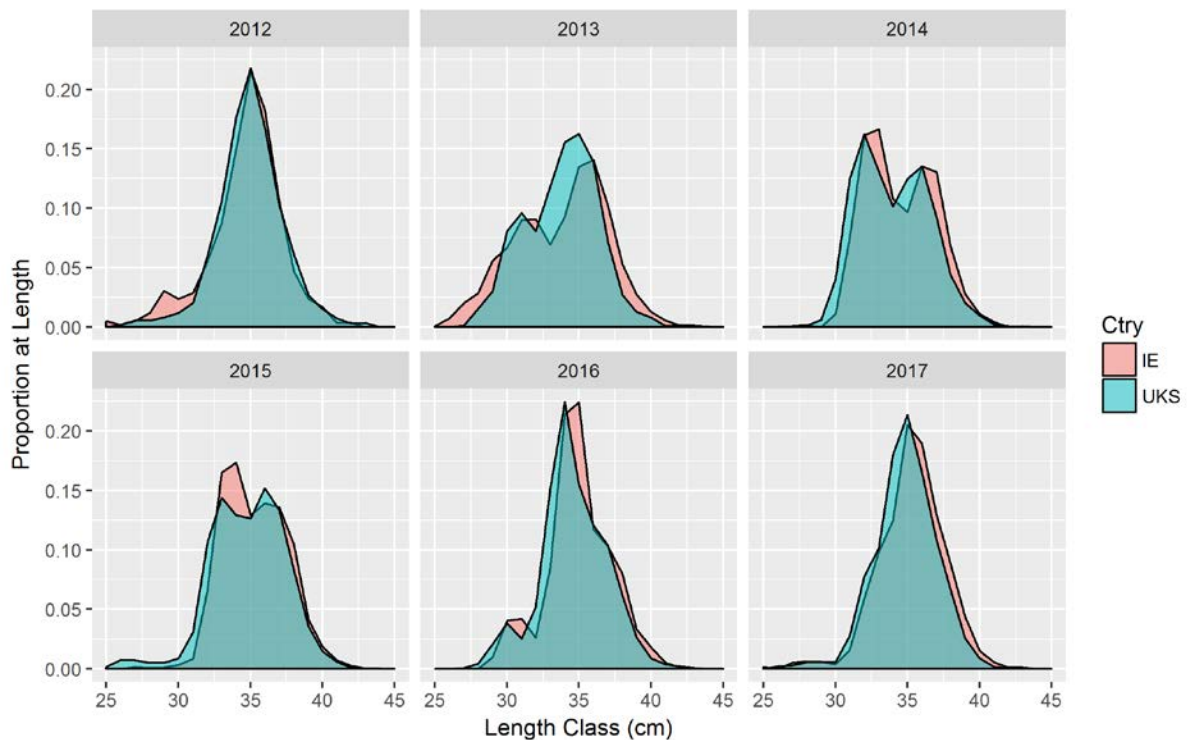


Figure 2: Ireland, Scotland LF, 4a Q4, WG Submissions 2012-2017.

Overall, there is close agreement between the two fleets as might be expected from two fleets that operate with similar practices. Bimodal profiles are recorded for both in each of 2013-2016 and in general peaks are coincident or within 1cm. A peak at 28cm (1-year old) seen in the Irish catch in 2012 is not seen in the Scottish data. Of concern is an apparent systematic tendency for Scottish catch to be slightly smaller. This is seen in each year and is unlikely to be due to random variation suggesting a possible difference in sampling and/or recording protocol.

Accompanying the sampling of catch for length is an ageing protocol which is carried out in order to estimate the catch numbers at age. A number of fish are selected from each length class, the otolith is extracted and read by an expert age reader in order to determine the age of the individual fish. The length frequency is then passed through the resulting age-length key to generate the numbers at age. Profiles of catch at age corresponding to the LFs in figure 2 are shown in figure 3.

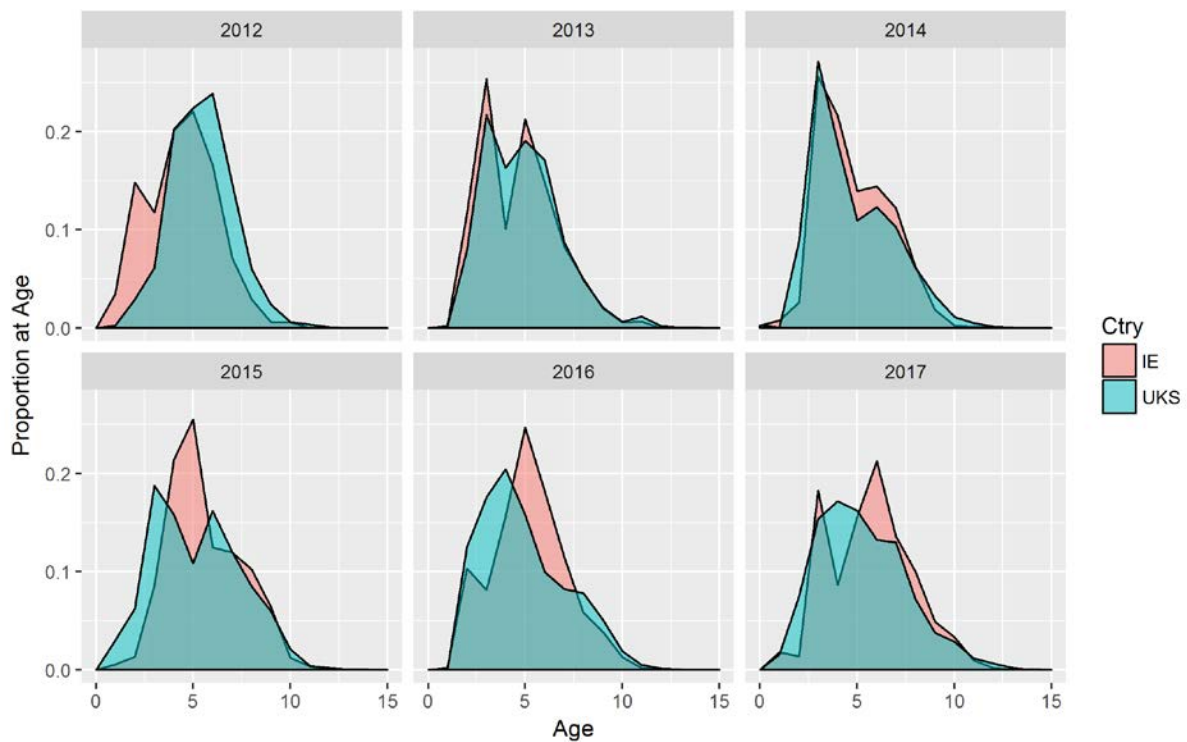


Figure 3: Ireland and Scotland Proportion at Age, 4a Q4, WG Submissions 2012-2017.

The similarities noted in the LF data are translated into broadly similar profiles of catch at age in 2012-2014. The peak at 28cm in the Irish catch translates into higher proportions of juvenile fish than seen in the Scottish catch. There are however, significant differences in proportions at age in each of the last 3 years. In particular Irish (and also Norwegian) data suggests a lack of age 4 fish in the catch that is not seen in the Scottish data for which this is the modal age. The same year class (2013) is also seen as potentially weak in the Irish and Norwegian data from 2016 but not in the Scottish sampling. Further comparisons including the Norwegian sampling are shown in figure 4. The differences noted above between the Irish and Scottish age profiles are evident when Scottish data is compared to the Norwegian submissions.

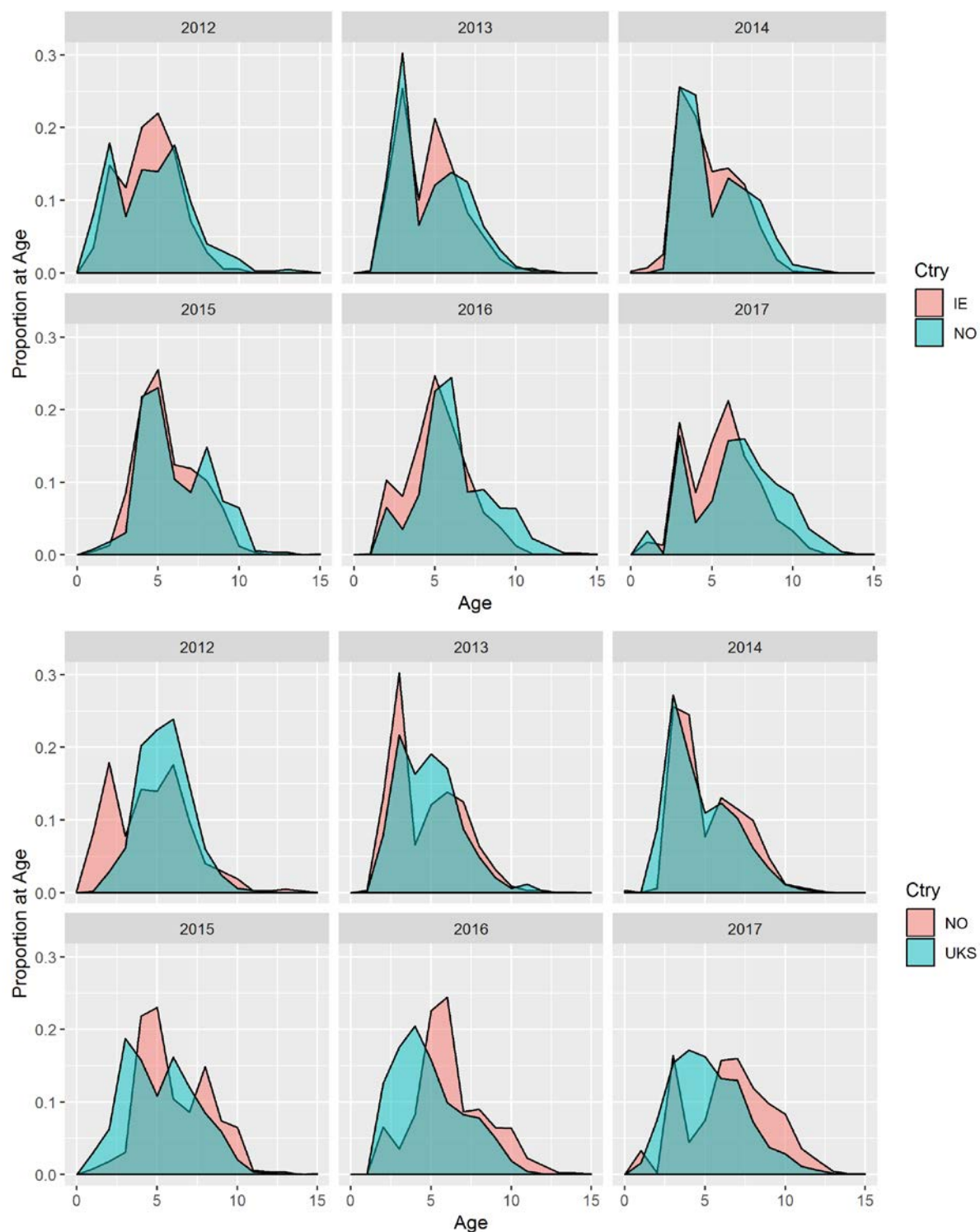


Figure 4: Norway vs Ireland (upper) and Norway vs Scotland proportion at age (lower), 4a, Q4, WG Submissions 2012-2017.

Average length and weight at age is also provided annually to the WG. A comparison of Irish and Scottish data is shown in figure 5. Aside from some differences for very young or old fish, data for the ages most numerous ages in the total catch (2-8) are very similar.

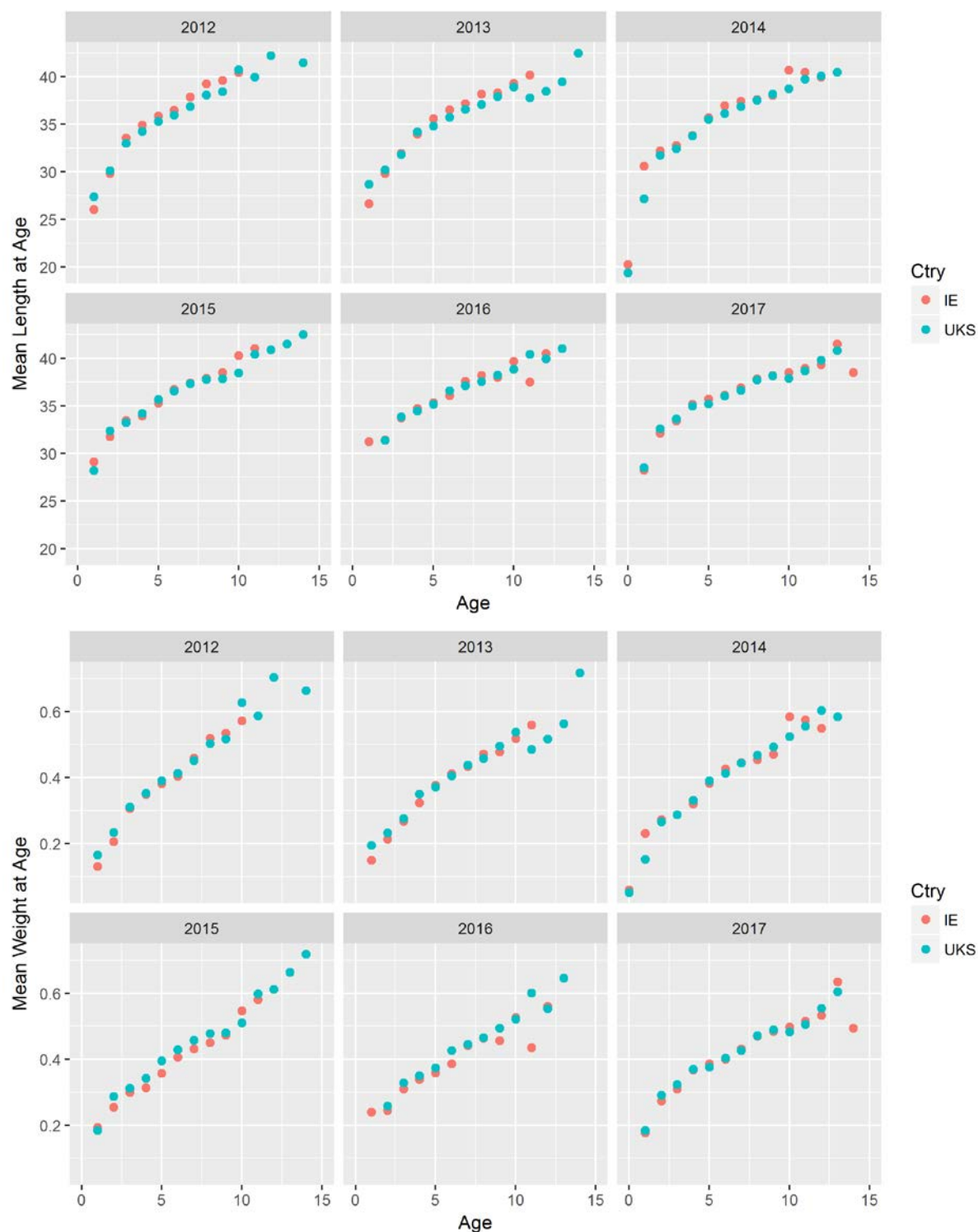


Figure 5: Average length (upper) and weight (lower) at age, 4a, Q4, WG Submissions 2012-2017.

Ageing of Mackerel

The ageing of mackerel involves visual inspection of the otolith and, for best results is best carried out by an experienced reader. Since individual nations use their own reader(s), the experience of a particular individual may introduce inconsistencies when national results are compared, as here. In order to minimise potential bias, regular workshops are held where readers compare results and

discuss methods. The most recent workshop was held in 2018. This workshop was attended by 11 expert readers and 7 trainee readers and was preceded by an otolith exchange which indicated that levels of agreement between readers was lower than for the previous exercise in 2014. During the workshop itself, all readers present analysed a set of 28 otoliths of known age.

Taking all otoliths into account the results of the workshop show a relatively poor low agreement for all readers (experts and trainees) with only slight improvement when only expert reader results are compared. However, by splitting the otolith readings into fish younger and older than 6 years, it can be clearly seen that the agreement drops dramatically from 79.1% (experts) and 67.5% (trainees) for the younger fish to 40.7% and 15.9% for fish older than 6.

Disaggregated Data

In order to further investigate the possible reasons for the observed differences in catch numbers at age, it is necessary to explore disaggregated data *i.e.* individual length-frequency and age samples. Unfortunately, due to a lack of a current data sharing arrangement, it was not possible to assemble data for each of Scotland, Ireland and Norway in time for the IBP. Irish and Norwegian biological samples from 2015-2017 for area 4a are compared in figure 6.

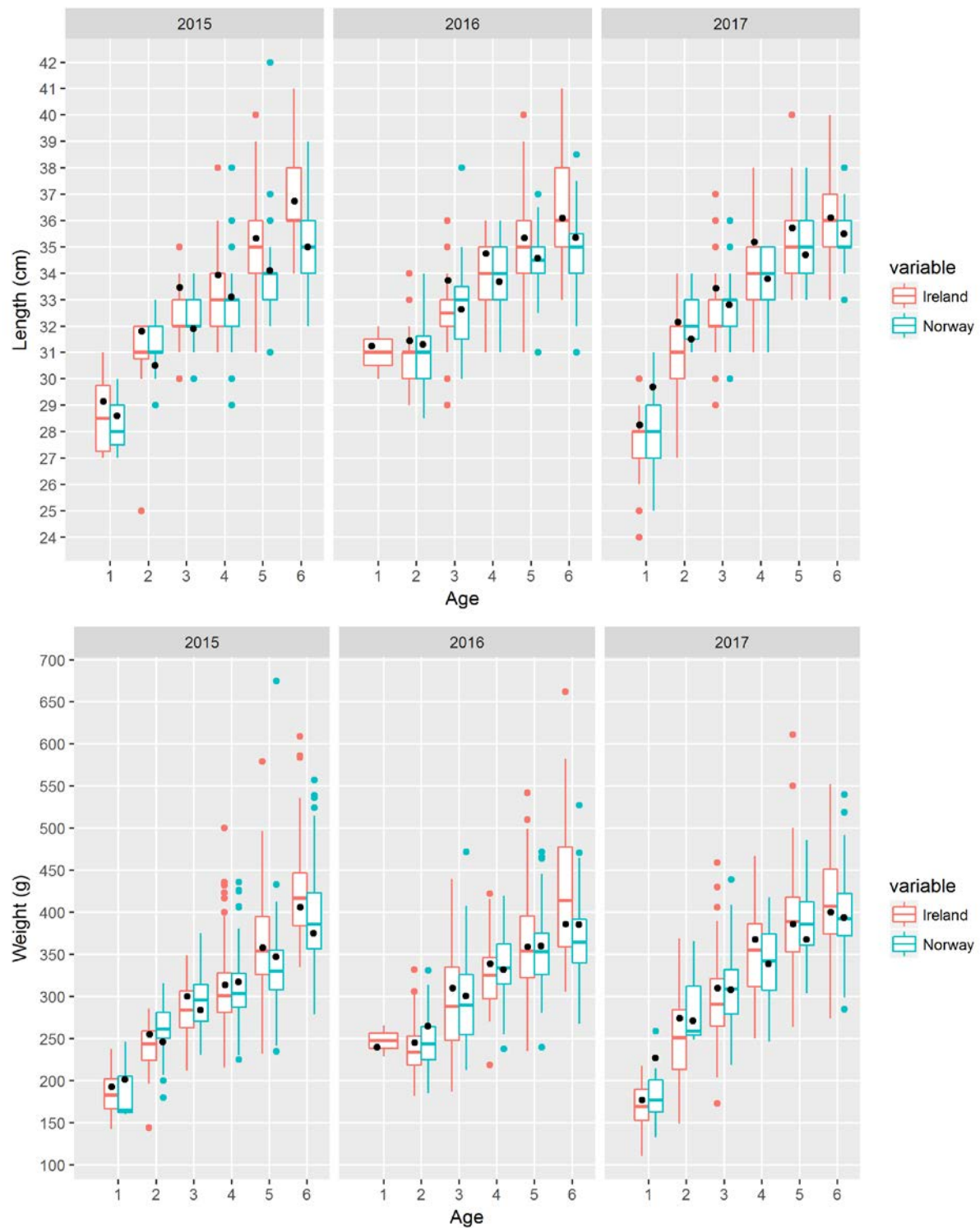


Figure 6: Length (upper) and Weight (lower) at age, 4a, Q4. Black dots represent the mean value submitted to the WG.

A comparison of the proportions at age for each length class from the Irish, Norwegian and Scottish national sampling programs in 2017 (4a, Q4) is shown in figure 7. Results are consistent for the youngest fish. However, for the most important length classes (*i.e.* those of which the majority of the catch is comprised, 31-36cm) the Scottish data differs in that a wider range of ages is associated with each length class with a maximum proportion for any one age of 0.5. In contrast, approximately 75%

of fish of length 31-33cm are aged 3 in Irish and Norwegian data. For larger (older) fish Norwegian samples tend to be older and Scotland younger. Agreement on fish older than 6 is known to drop significantly

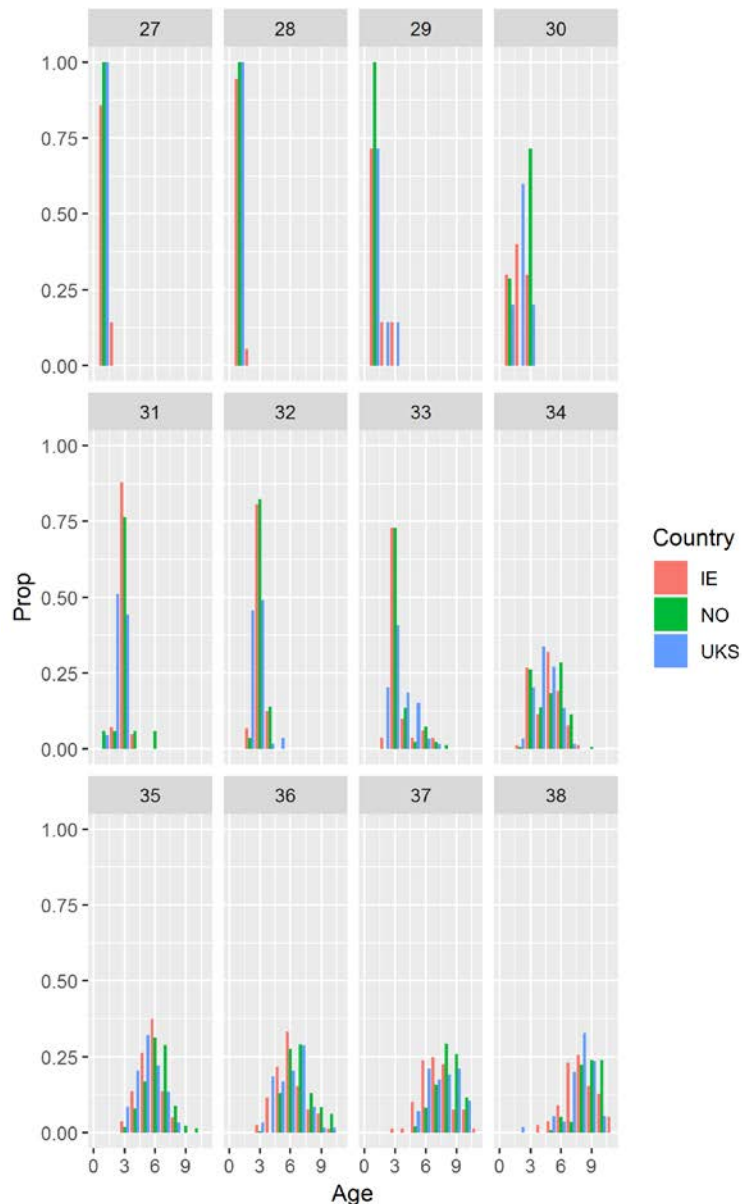


Figure 7: Proportion at Age by Length Class.

Assessment Sensitivity

As an exploration into the sensitivity of the assessment a number of alternative SAM runs were made under alternative sampling scenarios. The final WGWIDE 2018 assessment was considered as a baseline. Total catch at age estimates were then reworked under 3 alternative sampling schemes, progressively replacing estimates of catch at age from the Scottish fleet calculated using Scottish sampling data back to 2015 with alternative estimates calculated using Irish sampling information. This was done for all areas and quarters within each year and results from the following 3 assessments were compared with the baseline

- Alt1 - Using Irish samples to raise Scottish catch in 2017
- Alt2 - Using Irish samples to raise Scottish catch in 2016-2017
- Alt3 - Using Irish samples to raise Scottish catch in 2015-2017

Where Scottish catch samples were previously used to raise unsampled catch in combination with another sample, the Scottish sample was removed. Where the Scottish sample was the sole sample assigned then it was replaced with the Irish equivalent. In general, only minor catches require allocations to be assigned (approximately 15% of the total catch). The overall contribution of Scottish and Irish catch to the total catch in each of these years is given in the table below.

Year	Total WG Catch	Scotland	Ireland
2015	1,208,990	202,296 (17%)	88,744 (7%)
2016	1,094,066	190,818 (17%)	76,526 (7%)
2017	1,155,944	182,096 (16%)	84,915 (7%)

Table 2: WG Catch

The updated total numbers at age (compared to the baseline) are given in table 3 and figure 8. The general effect of this is to reduce the numbers of younger fish (2-4) and increase older (5-8) ages as Scottish sampling indicates an overall younger catch than Irish (figs 2 and 4).

Given the relative importance of Scottish catch in the total catch, reductions of the order of up to 20% are seen.

	2015					
	0	1	2	3	4	5
Baseline	6,732	104,019	124,411	248,852	579,835	646,894
Alt	6,732	95,220	108,720	196,250	581,244	707,050
Diff	0%	-8%	-13%	-21%	0%	9%
	2016					
	0	1	2	3	4	5
Baseline	716	45,199	203,753	257,293	424,843	589,549
Alt	716	45,750	196,874	231,670	389,723	616,889
Diff	0%	1%	-3%	-10%	-8%	5%
	2017					
	0	1	2	3	4	5
Baseline	28,306	43,458	87,739	458,301	351,779	396,862
Alt	28,382	43,250	71,847	452,036	308,396	378,460
Diff	0%	0%	-18%	-1%	-12%	-5%

	2015						
	6	7	8	9	10	11	12+
Baseline	450,344	415,107	355,997	205,691	107,685	26,939	22,700
Alt	430,778	428,645	363,702	213,404	101,680	26,215	21,588
Diff	-4%	3%	2%	4%	-6%	-3%	-5%
	2016						
	6	7	8	9	10	11	12+
Baseline	532,890	340,155	269,962	170,373	94,778	33,896	24,420
Alt	590,248	347,672	269,925	163,271	93,695	30,005	23,208
Diff	11%	2%	0%	-4%	-1%	-11%	-5%
	2017						
	6	7	8	9	10	11	12+
Baseline	503,601	431,014	261,959	188,950	138,143	59,211	51,090
Alt	513,685	448,340	269,133	195,990	156,390	67,571	56,874
Diff	2%	4%	3%	4%	13%	14%	11%

Table 3: Catch at Age adjustments for sensitivity runs

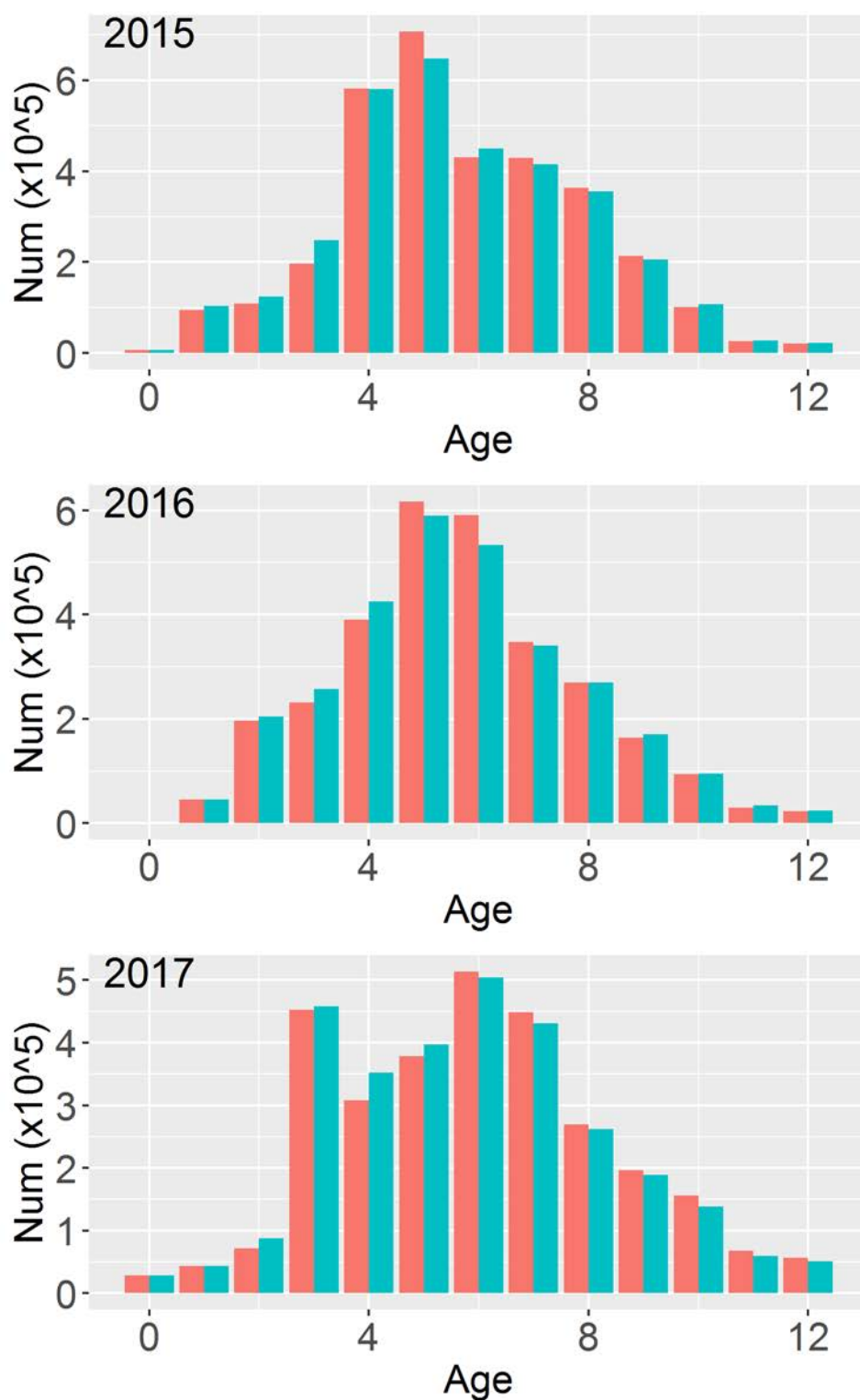


Figure 8: Catch numbers by age class, WGWIDE2018 and re-estimated (excluding Scottish sampling in red).

Figures 9 and 10 compare the outputs and parameter estimates from the 4 assessments. The residuals are in the appendix. There is little evidence of sensitivity of this particular assessment to the sampling information associated with Scottish catch in 2015-2017.

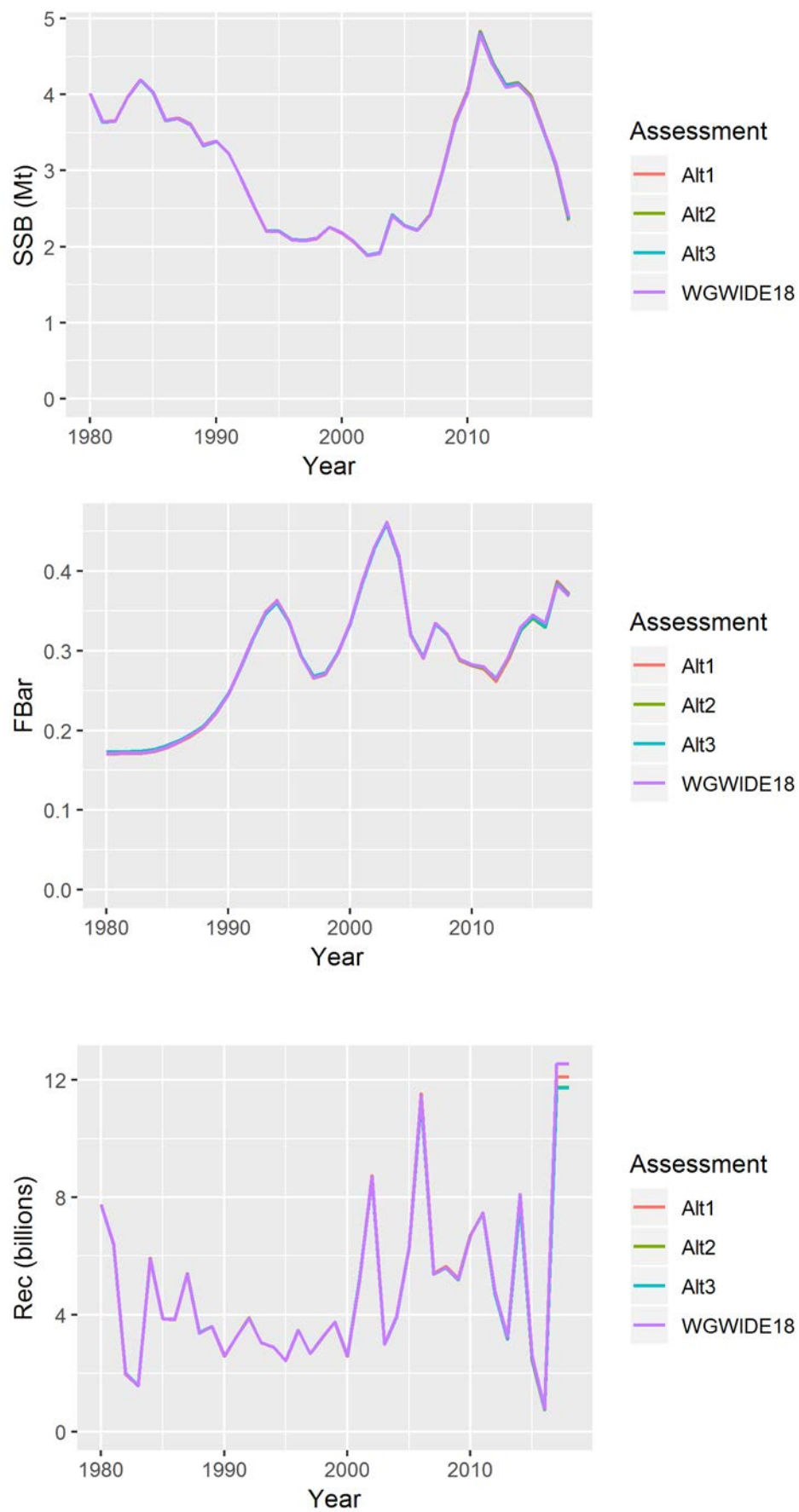


Figure 9: Assessment stock estimates comparison.

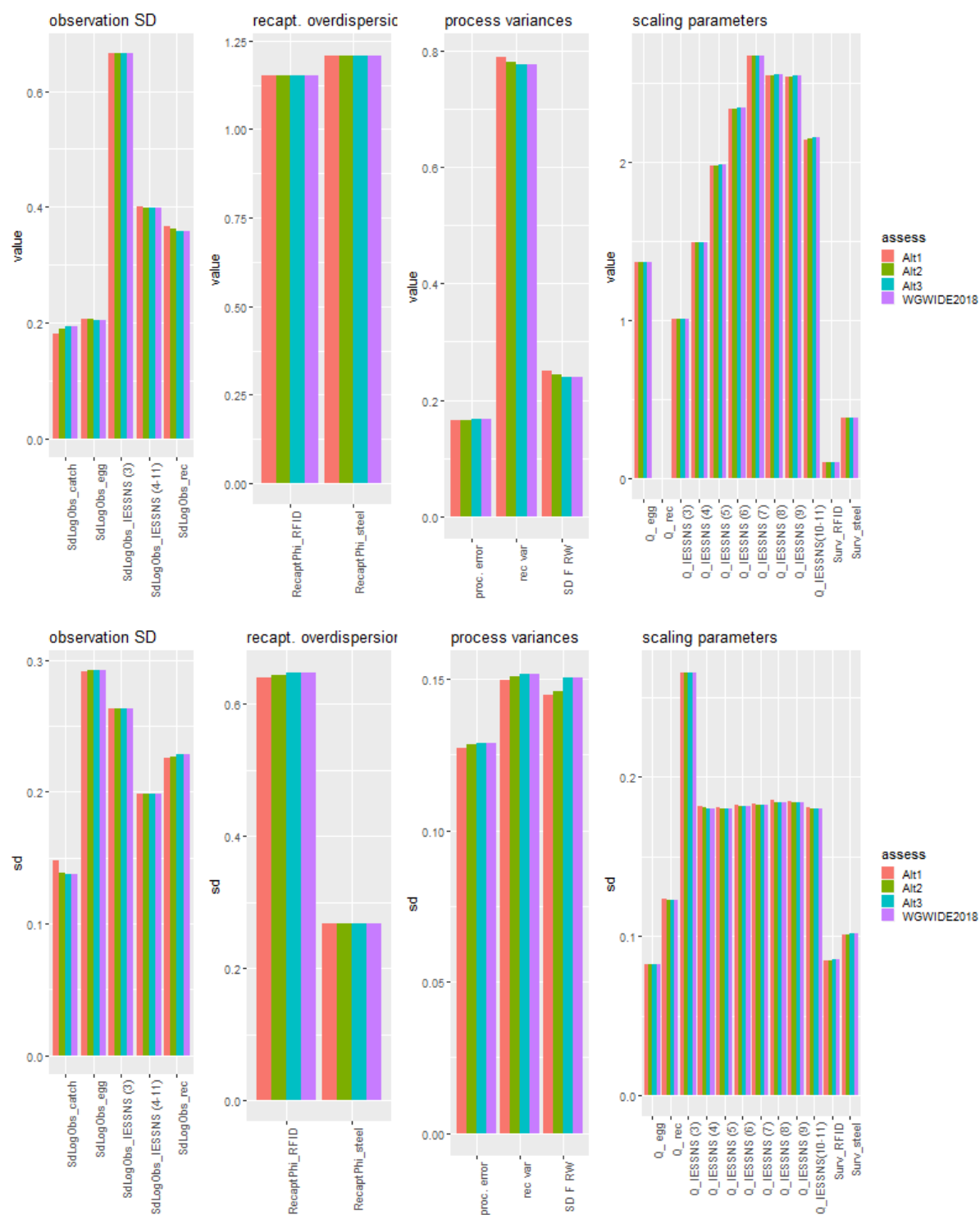


Figure 10: Parameter Values and SD Estimates

Discussion

During preparations for the 2019 inter-benchmark assessment of NEA Mackerel, differences in the age profile of the catch at age data submitted annually to the WG were noted. In particular, data for Norway, Scotland and Ireland from the division 4a, fourth quarter fishery in each of 2015-2017 diverge to a greater extent than in previous years. Both Irish and Norwegian data indicate a relatively weak 2013 year class with a low proportion of age 3 fish in 2016 and age 4 in 2017. Previous assessments also indicate that 2013 is below average. This signal is not seen in the Scottish data. Indeed, in 2017, age 4 fish are the most numerous. There is no indication in the length frequency profiles consistent with this. Catch reported by statistical rectangle indicates that the national fleets have similar fishing pattern. A potential conclusion is that the differences may be due to the ageing of the sampled catch by the national labs. Exchanges and workshop exercises involving age readers indicates that agreement between has declined between comparisons in 2014 and 2018 and is generally poor for fish older than 6.

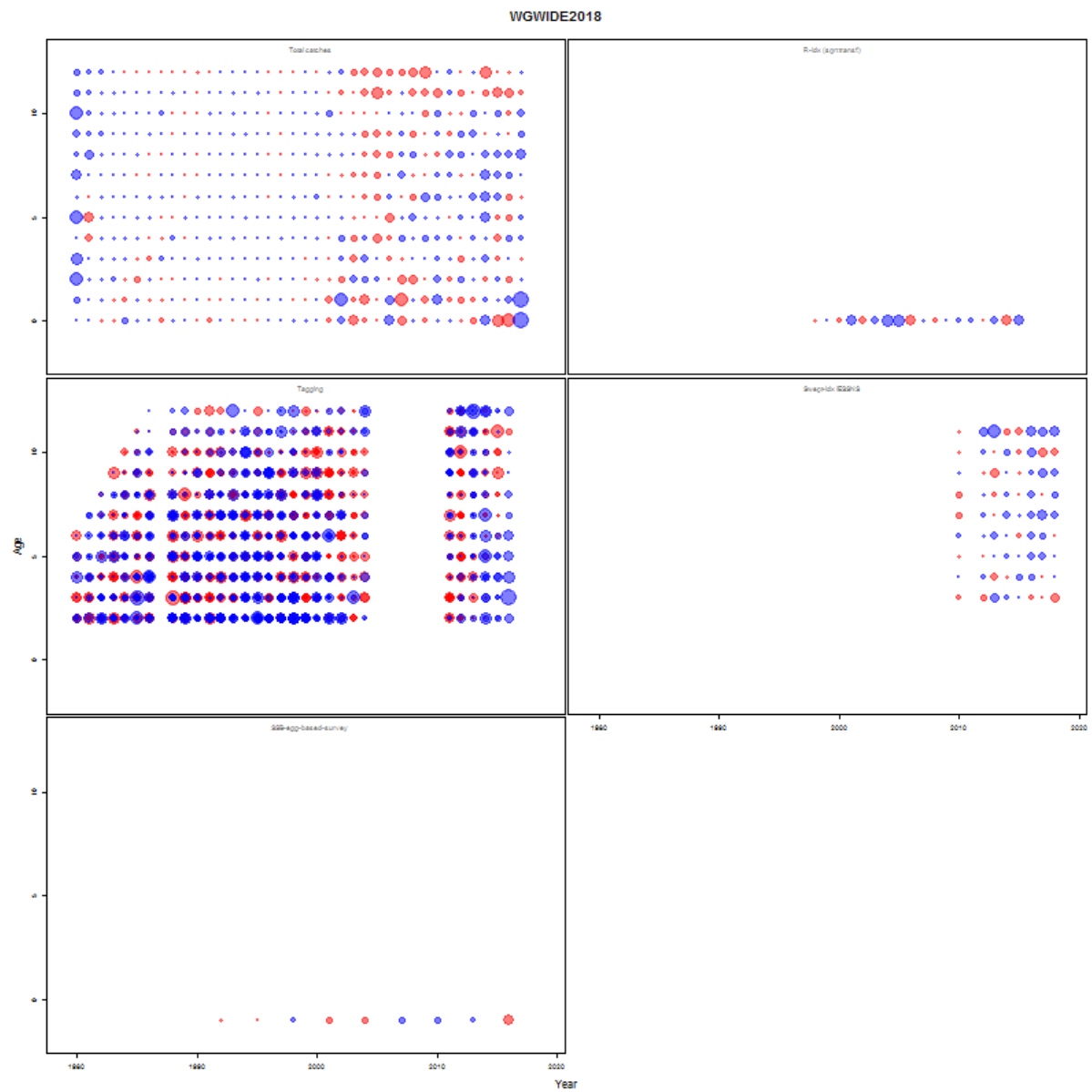
A graphical comparison of average weight and length at age data submitted to the WG does not indicate potential issues. However, on close inspection it can be seen that for the 2017 data submission (for 4a, Q4) Norwegian average length and weight at age is consistently lower than that reported for Scottish and Irish data. Scottish age 4 and 5 fish have almost identical average weight (6g difference) and length (0.2cm). Typically, a weight difference of the order of 30g is seen between ages 4 and 5.

A comparison of the national SOPs and methodologies for calculation of the catch at age estimates should be carried out. There appear to be systematic differences in the reporting of fish length, average length/weight at age, the full impact of which on the assessment is not clear.

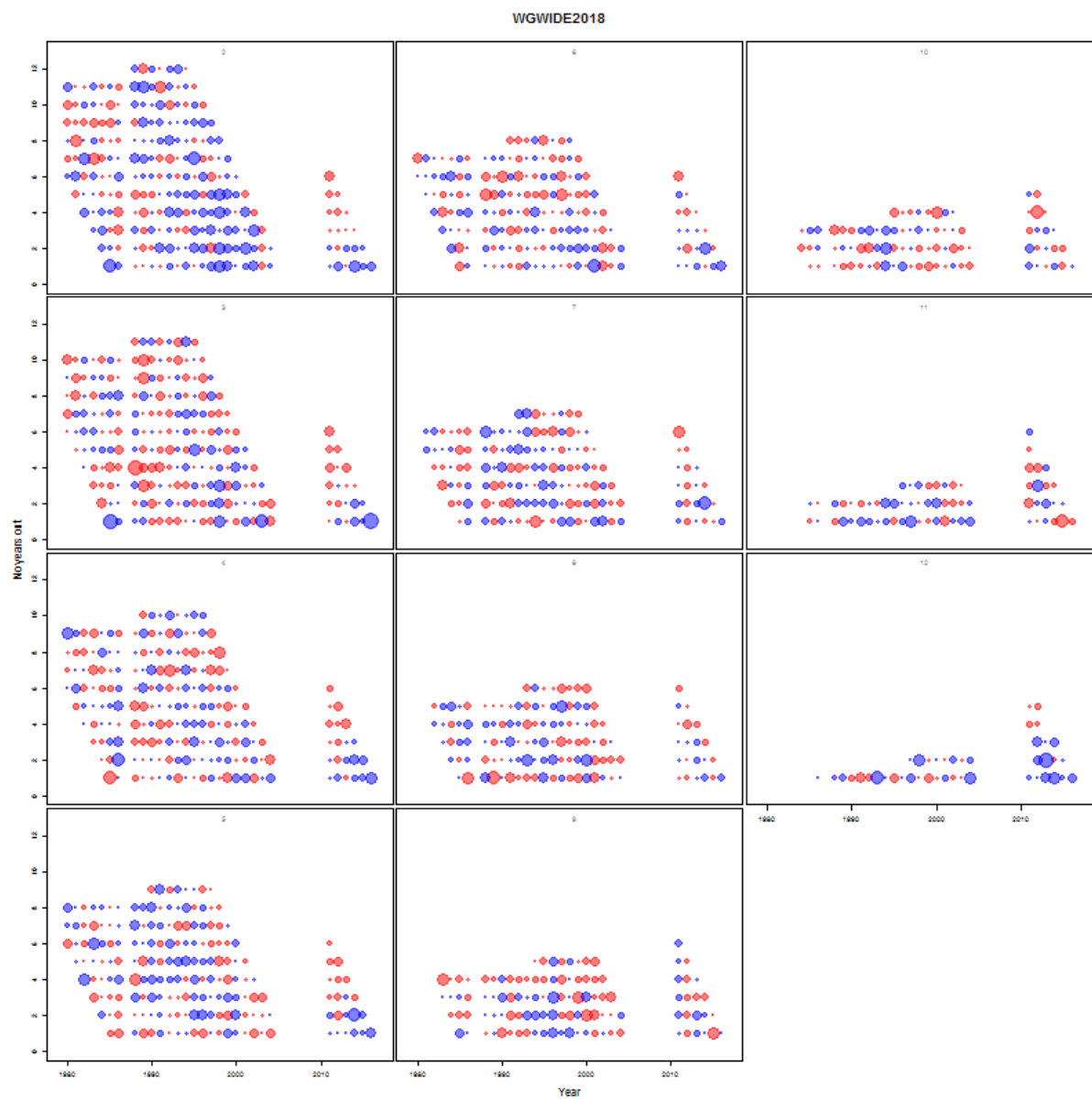
Reworking the catch abundance and weight at age inputs for the assessment by replacing Scottish sampling information with that from the similar Irish fishery indicates that the current assessment is relatively insensitive to this particular data although a more extensive exercise is warranted.

Appendix – Assessment Residuals

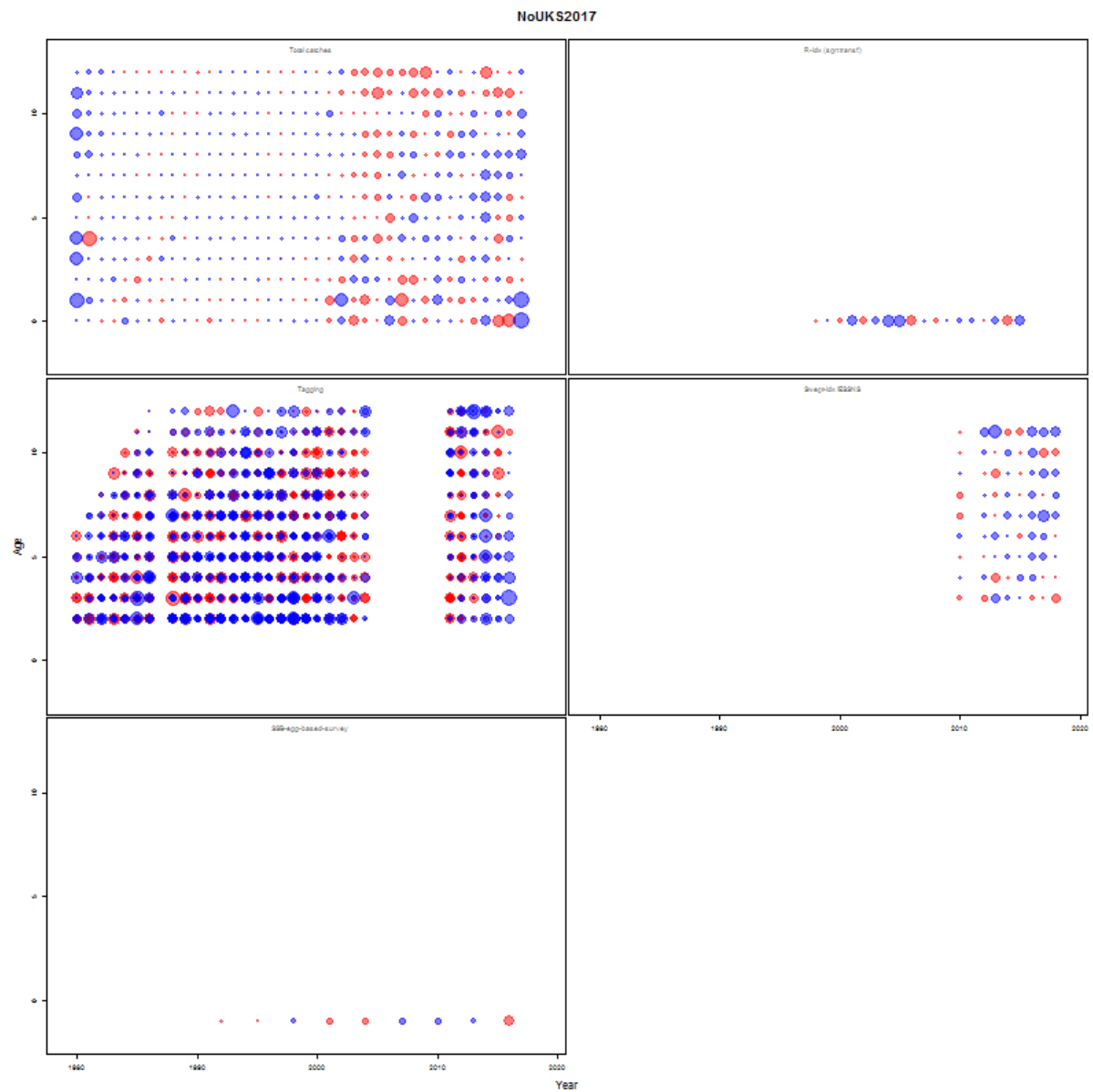
Baseline Assessment (WGWIDE2018 Final Assessment)



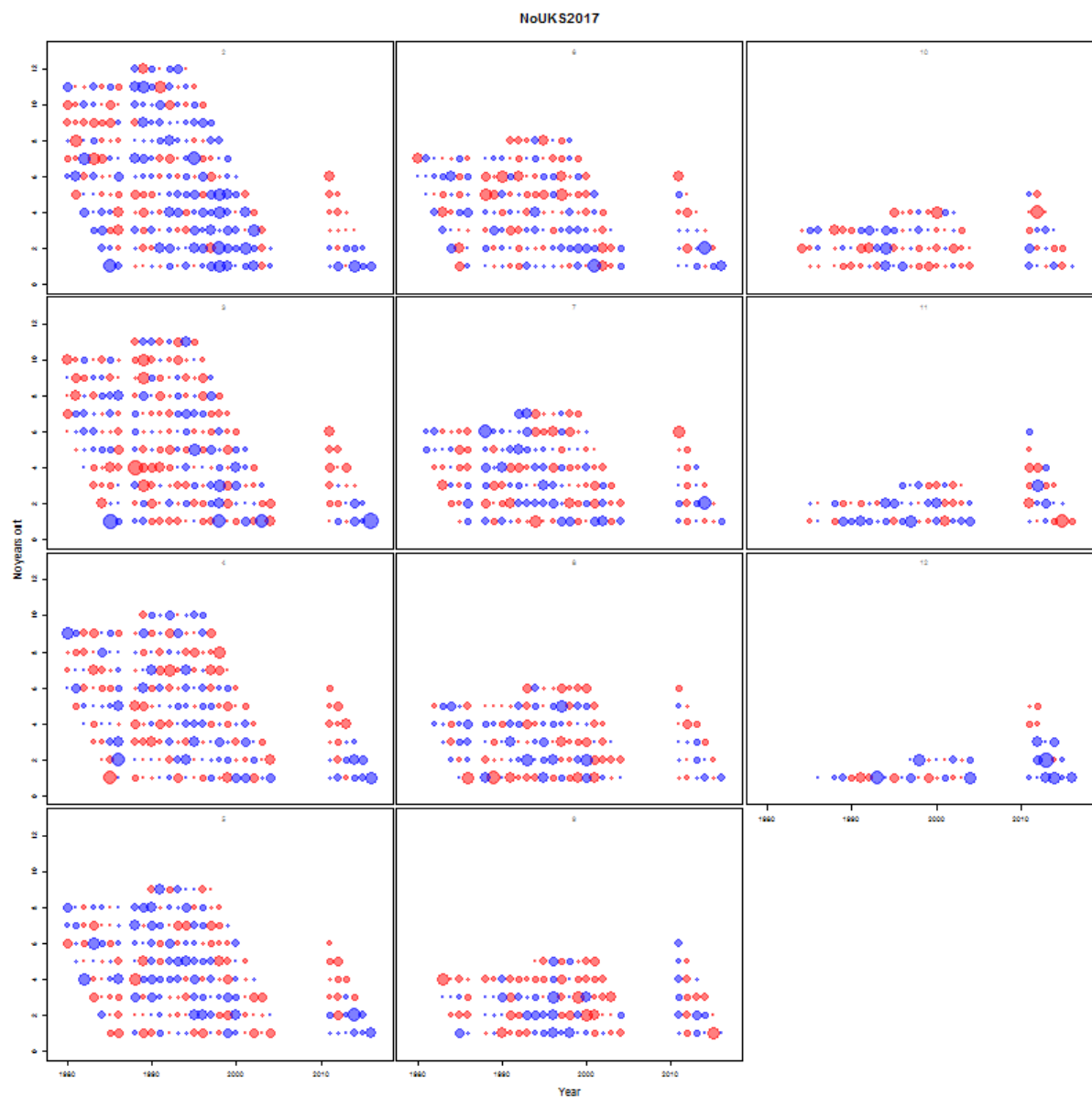
Baseline Assessment (WGWIDE2018 Final Assessment) – tag residuals



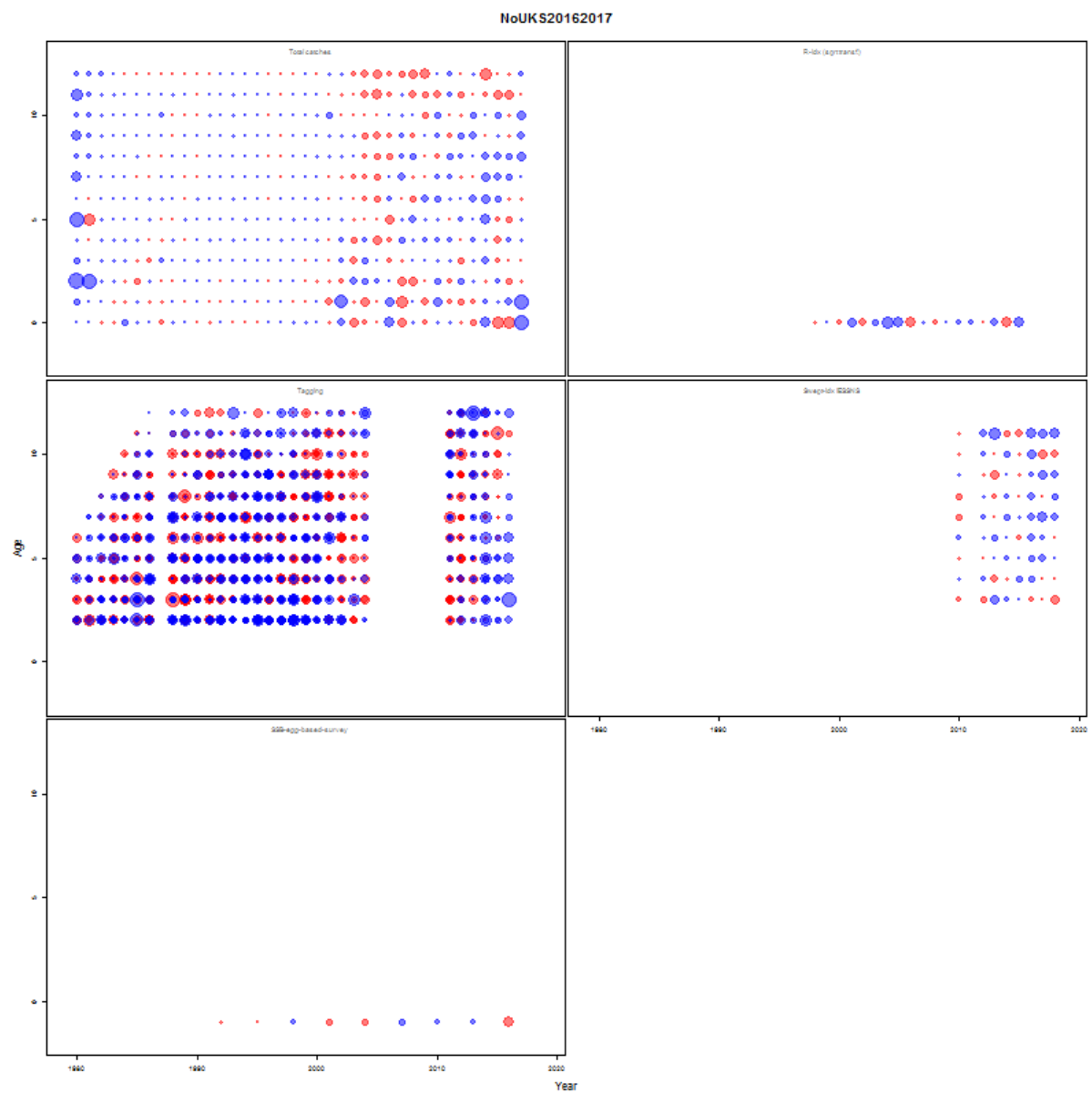
Alt1 Assessment (IE Sampling used for UKS in 2017)



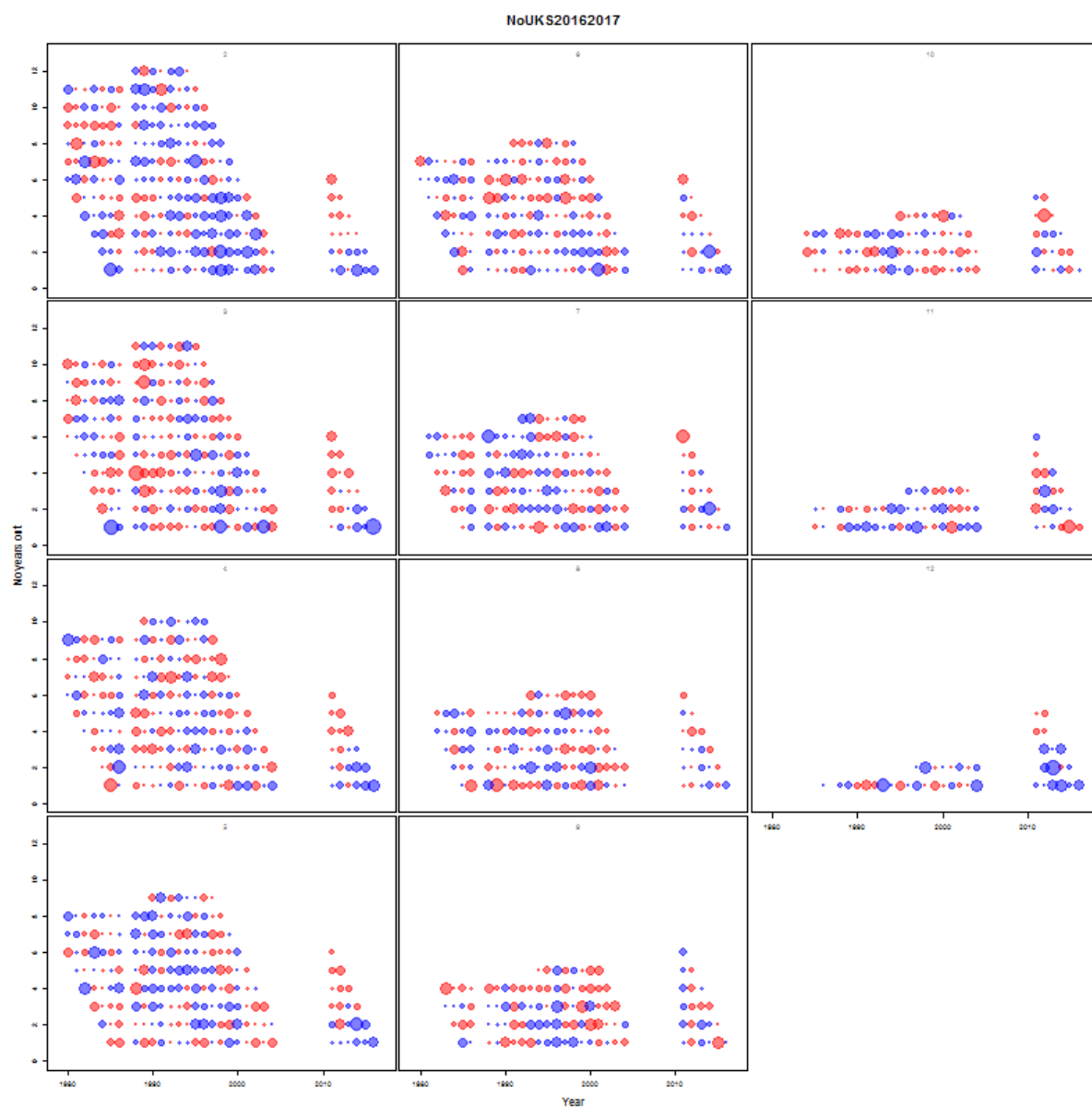
Alt1 Assessment (IE Sampling used for UKS in 2017) – tag residuals



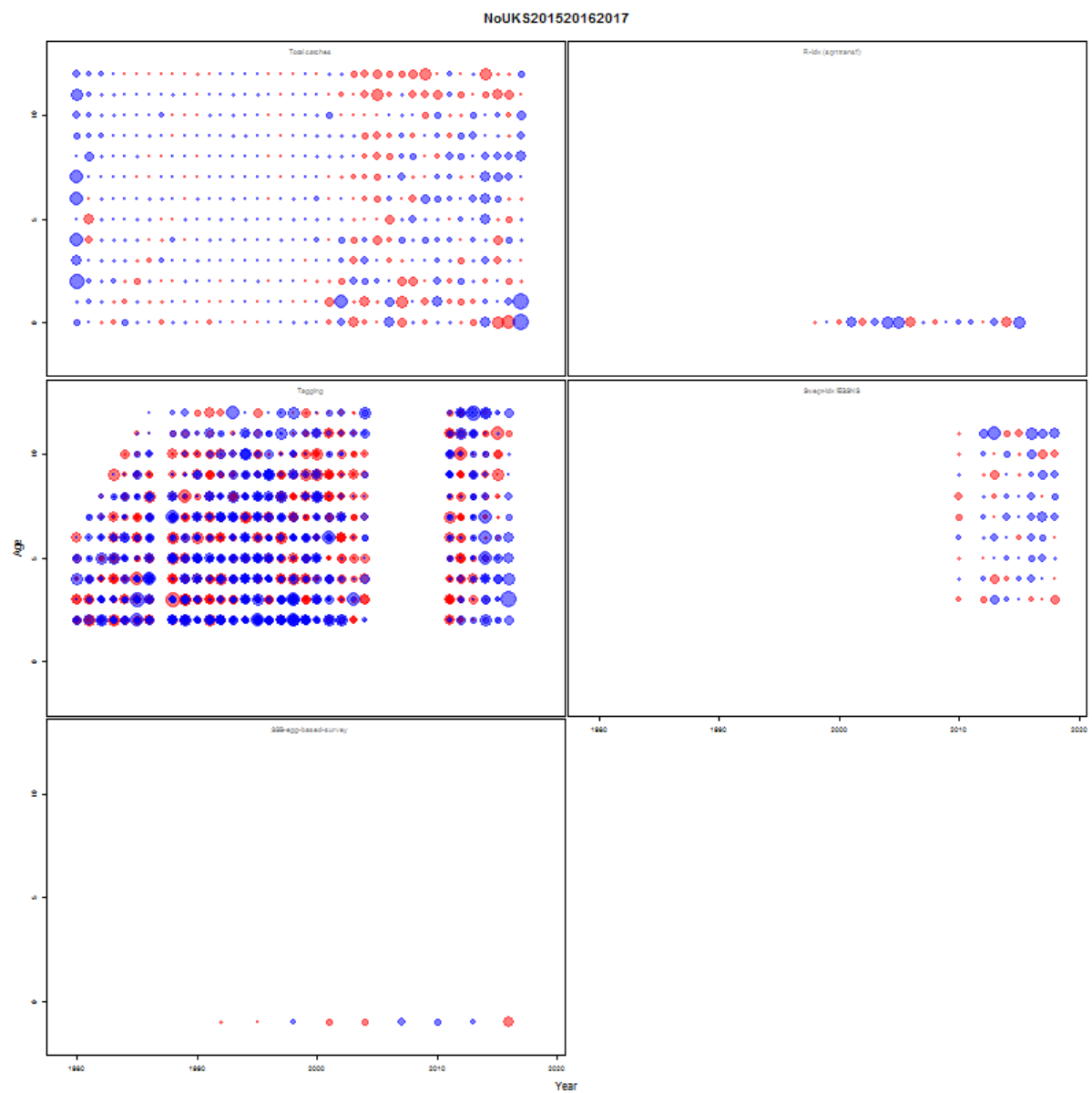
Alt2 Assessment (IE Sampling used for UKS in 2016 and 2017)



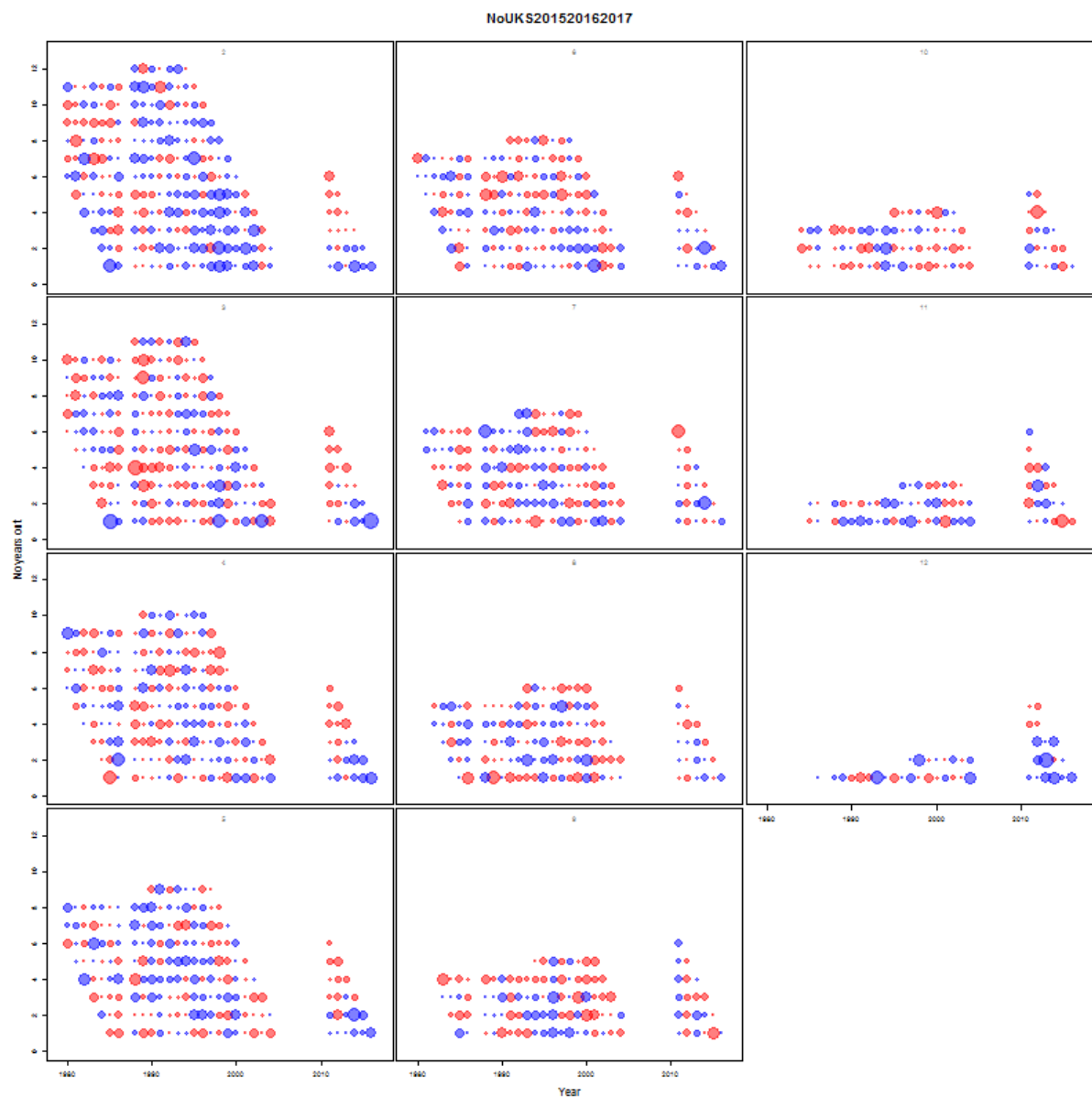
Alt2 Assessment (IE Sampling used for UKS in 2016 and 2017) – tag residuals



Alt3 Assessment (IE Sampling used for UKS in 2015, 2016 and 2017)



Alt3 Assessment (IE Sampling used for UKS in 2015, 2016 and 2017) – tag residuals



Assessment sensitivity to different criteria for selecting subsets of the RFID dataset

Thomas Brunel (Wageningen Marine Research)

Working document to the Interbenchmark Workshop on the assessment of northeast Atlantic mackerel

During IBPNEAMAC2019, the RFID data was carefully scrutinised, and potential sources of bias were identified. While some of these sources of bias could potentially be accounted for in the model (this was investigated during the meeting), a more pragmatic approach consisted in selecting a subset of the data that would not be prone to these sources of bias.

This document 1) presents the criteria used to select a subset of the data, 2) shows the sensitivity of the assessment to the subset selected and 3) reflects on the influence of this data quality v.s. data amount on the weight it has on the SAM assessment

rational for taking subset of the tagging data

- Selection of the data years to be used

A RFID dataset in which recaptures are disaggregated by large geographical areas has been prepared for the 2018 WGWIDE meeting. The SAM model was run with a configuration estimating separate post release survival for each recapture area (ICES WGWIDE 2018). The estimated parameters showed some differences between areas : higher in the Icelandic area, lower in the Norwegian Sea area, intermediate in the North Sea and west of Ireland areas. These differences are not likely to be explained by actual differences in survival immediately after release, but could potentially indicate mixing/spatial component issues. For example, if only a specific part of the population is tagged (because tagging occurs on a somewhat limited spatial and temporal window), and if this specific component does not completely follow the migration behaviour of the overall population, tag concentration may end up being higher in the areas where the tagged part of the population mainly moves (e.g. potentially explaining the higher survival rate for the Icelandic area), and lower in areas where it moves less (e.g. Norwegian Sea).

As long as the contribution of these different areas does not change through time, this spatial differences in recapture rate are not a source of bias for the assessment. However, in the development of the RFID program, both the amount of catches scanned, but also the contribution of the different areas have changed in time (figure 1). In the first 2 years of recapture, the catches scanned were at low number, mainly came from the North Sea, and none came from the Icelandic area. Since 2017 numbers are larger and the contribution of the 4 areas is more stable.

In order to avoid introducing a bias in the assessment, it seemed justify to use only the period during which there was some consistency in the geographic origin of the catches scanned, i.e. to exclude the first two years of recapture (2012-2013)

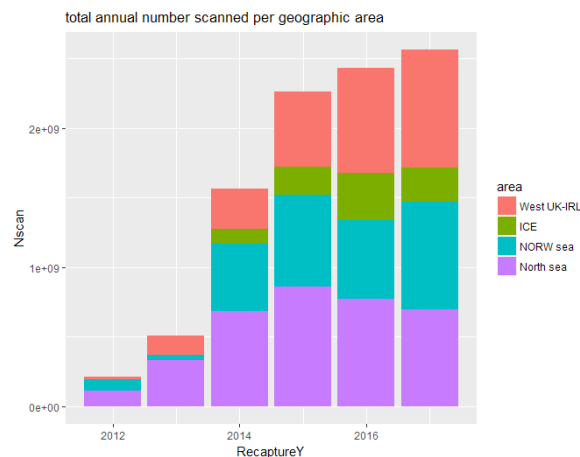


Figure 1 : changes in the number of mackerel scanned since the start of the RFID program, disaggregated by geographical areas.

- Ages at release

Analyses of the age composition in the released tagged fish, the recaptures and in the scanned catches were presented at IBPMAC2019. For the steel tag period, there was a good correspondence in the age compositions, with high proportion of young fish tagged (2/3 year olds), recaptured 1 year after, and in the scanned catches. Proportions then decreased as ages increase. For the RFID period, proportion of tagged and recaptured fish of age 2 and 3 are low, increasing with age up to age 7-8 and decreasing again for older ages. For the scanned catches however, proportions of young (2-3 year olds) fish are high, increasing until 6 year olds, and decrease at older ages. Comparison with IESSNS survey tag show that 2-3 year olds were present in large numbers in the survey area. Although age 2-3 are well represented in the catches scanned, they are tagged in very low numbers. The concentration of tags are therefore very low for the young age at release.

In addition to the low tag concentration (suspected to make recapture numbers more uncertain, because very small), low tagging intensity at young age is suspected to be a source of bias: as very few young fish are present on the tagging area, it is suspected that the fraction of the young age-groups that are tagged may not be representative of the bulk of the age-group. If that is the case, the small component of the population of 2-3 year olds present in the tagging area may not be distributed in the whole area from which catches are scanned. This would imply the recapture rates for these tagged fish is not representative only of original cohort size, but also of the representativeness of the component tagged. This might be the case for all ages, but the problem is much more acute for young fish since number tagged are very low and young fish do not migrate as far as the older ones.

On the basis of the low number tagged at age 2,3 and 4, the exclusion of these ages at release was also investigated

- Years at liberty

There is increasing concern that recapture rates from a given release year tend to decrease as the number of years at liberty increases. This is illustrated on figure 2 showing the recaptures expressed as abundance at time at release, as a function of the number of years at liberty (YO): the perception of e.g. abundance at age 5 in the stock in 2011 increases between the recaptures from 2012 to the recapture from 2016. Such a decrease in recapture rate over the years for a given tagging experiment can possibly be due to tag loss (which is known to occur on salmonids with RFID tags) or to tag malfunctioning.

The dataset is the asymmetrical: many recapture years/data points for old releases, a single one for the latest release year. If old recaptures systematically give a larger perception of the abundance at release than recaptures just after release, then there would automatically be a tendency for early release years

(e.g 2011) to show higher abundances than recent release years (e.g 2016). Tag loss would thereby introduce a spurious declining trend in the abundances at age through time. Inspection of figure 2 suggest that this may be a more serious problem for young ages than for older ages (i.e ages 2:6 from release years 2011-12).

This effect of tag loss- tag malfunctioning can in principle be incorporated in the model, but this requires some knowledge of how this process happens in order to adopt an appropriate model formulation. The more pragmatic way of avoiding this bias is to use only a dataset that includes the same number of data points for each release year. Using only the recapture 1 year after tagging provides the longest time series.

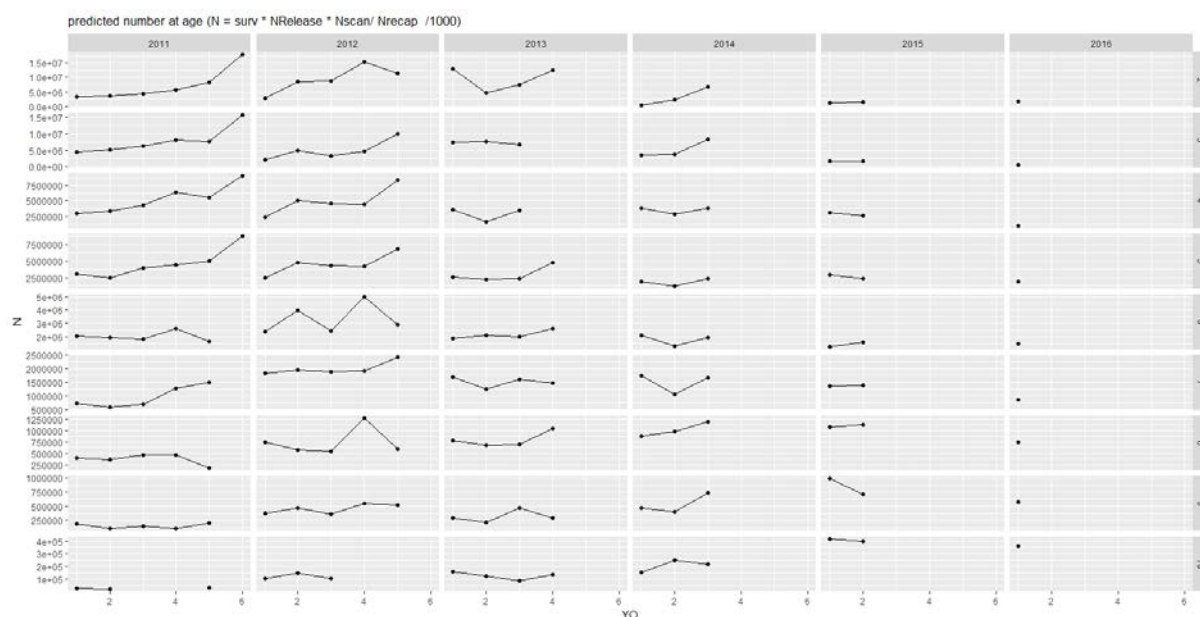


Figure 2 : Abundance-at-age (in row) per year (in columns) estimated from the recapture data as a function of the number of years at liberty.

Evaluation of the information provided by the RFID tags

For a typical age structured survey, looking at the ability of the index to follow cohort signals from one age-group to the next is a common way of judging of the quality of the index. A similar approach was adopted here for the RFID data. The data was considered separately for each number of years at liberty.

The internal consistency of the information provided by tags recapture 1 year after release (figure 3) is moderate, with reasonably good correlation between some combination of ages (correlation above 0.5 for 3-4, 4-5, 9-10, 10-11, 11-12) but also with low correlations (<0.3) for the rest of the ages. The internal consistency is lower for the information derived from tags recovered after 2 years at liberty (figure 4), with all correlations (except 1) under 0.5, and most for them under 0.3. The picture becomes even worse for tags recaptured after 3 years at liberty (figure 5), with even some negative correlations.

The consistency of the data can also be investigated comparing the time series for different numbers of years at liberty (i.e. correlation of the time series of abundance at age A seen from the recapture occurring after X vs. Y years at liberty). This analysis (results not shown here) indicates a complete absence of correlation (between number of age at liberty 1 to 4) in the data for the younger ages (ages at release 2 to 6) and a good correlation for ages 7 to 12.

Although the validity of this analysis is somewhat limited by the short length of the time series (sometimes only 4 points for comparisons), this suggest that the recaptures after one year at liberty represent the most informative part of the RFID data set.

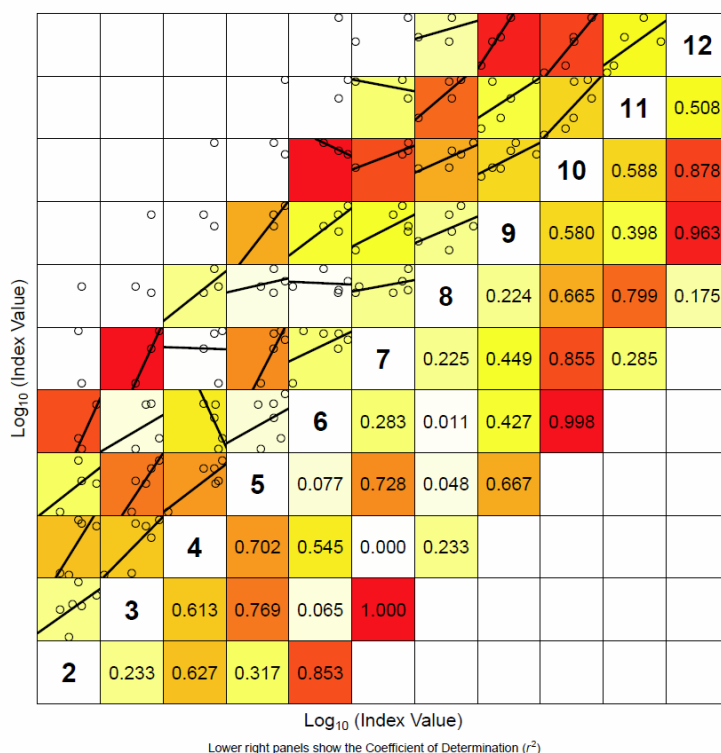


Figure 3 : internal consistency plot for the abundance at age estimates derived from recaptures occurring after 1 year at liberty

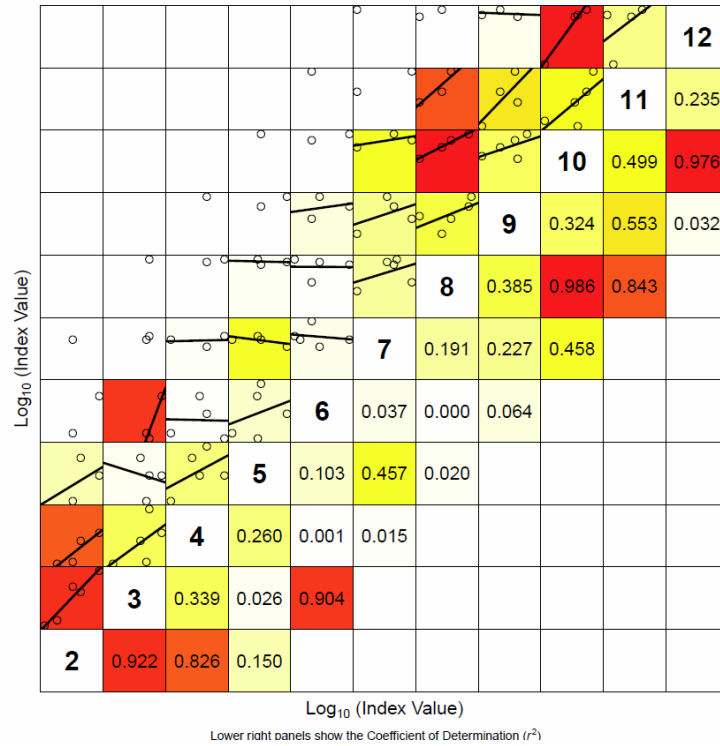


Figure 4 : internal consistency plot for the abundance at age estimates derived from recaptures occurring after 2 years at liberty

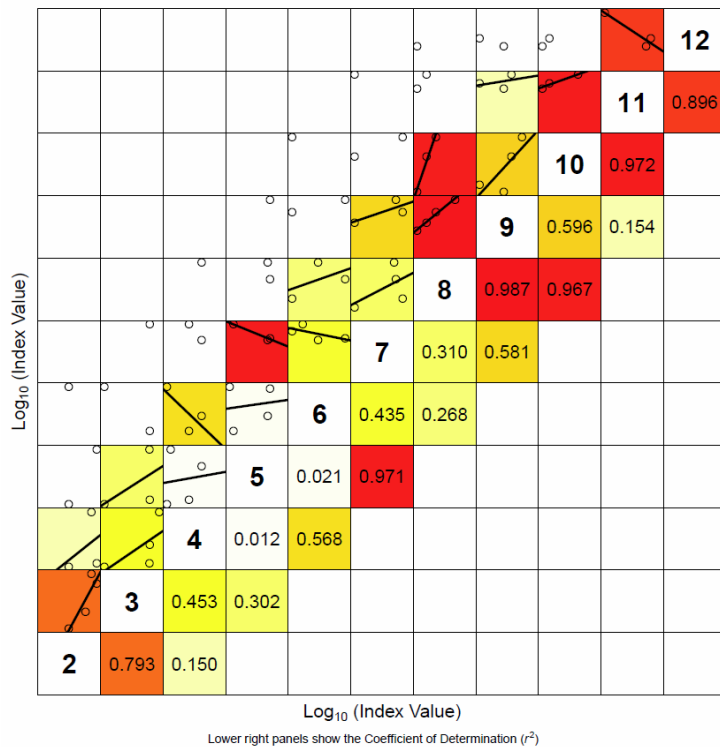


Figure 5 : internal consistency plot for the abundance at age estimates derived from recaptures occurring after 3 years at liberty

Sensitivity to the different potential subsets of the RFID data

The assessment was insensitive to removing the first 2 years of recapture data (figure 6). Removing in addition ages 2 and 3 also did not have a strong effect on the assessment output. Using only recapture after 1 year at liberty had the largest effect, with a strong upward change in SSB for the recent period and a corresponding downward change in Fbar.

Using only the recapture after 1 year at liberty lead to a change in the parameters (figure 7) related to variance of the errors on the observations, with namely an increasing overdispersion for the RFID tags (indicating a poorer fit), a decreased variance for the IESSNS, and increased variance for the egg survey. These changes in the observation variances imply a different weight of the different data sources on the assessment, namely a reduction of the weight of the RFID data and an increasing weight of the IESSNS.

One can notice that although the selection of recaptures after 1 year at liberty alone should in principle result in a higher overall quality of the RFID data used (based on the better consistency plot), the model put less trust in this data source (high overdispersion).

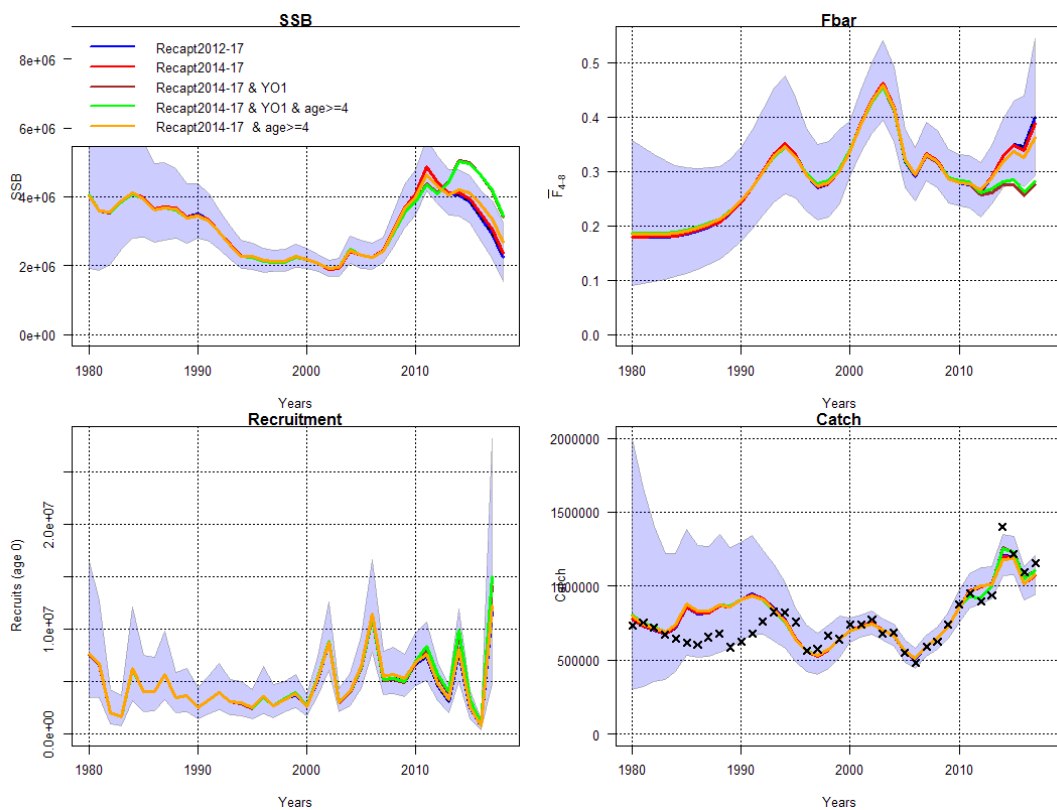


Figure 6 : stock trajectories for a series of runs on different subsets of the RFID dataset

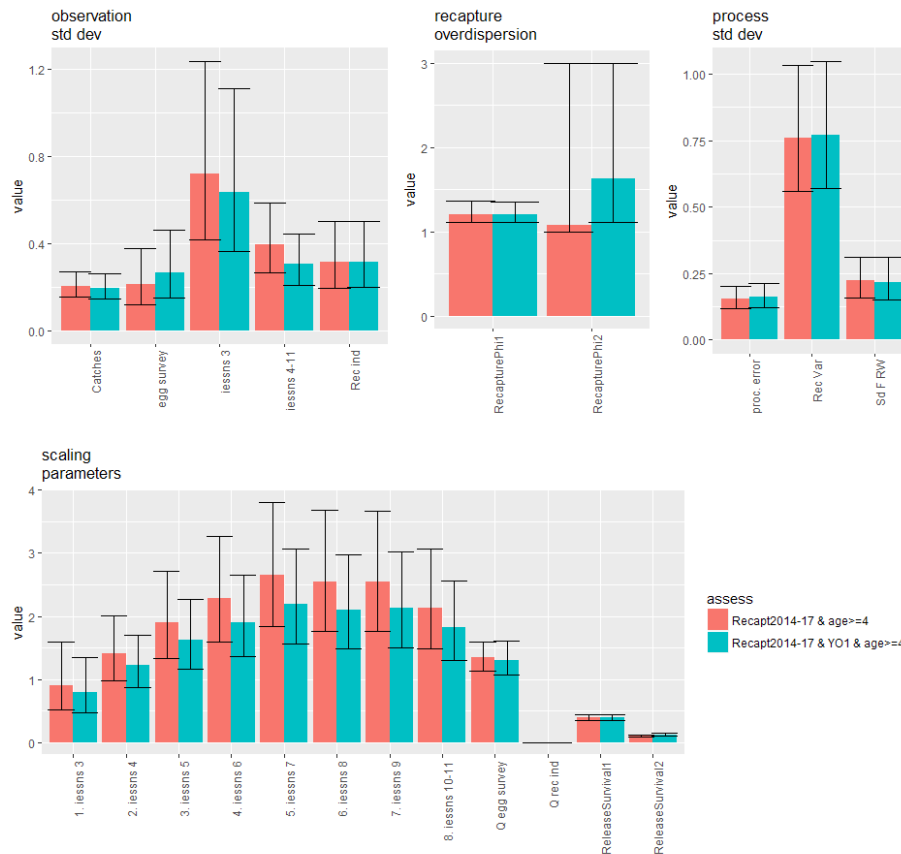


Figure 7 : comparison of the estimated parameters for the runs using all numbers of year at liberty v.s only 1 year at liberty (and excluding recapture years 2011 and 2012, and ages at release 2 and 3).

Interpretation of the sensitivity observed

Why does the recent perception of the stock change so much when only recaptures after 1 year at liberty are used? Two potential explanations were investigated here. First, the information on the development of the abundances at age provided by the subset of the RFID data might be different from the information in the whole dataset. Second, by taking this subset of the data, the number of data point is substantially reduced, which decreases the weight of the tagging data on the assessment, giving thereby more weight to the other data sources.

Influence of the information provided by the RFID data

The information provided by the subset of the data with recaptures 1 year after tagging is different from the information in the whole RFID dataset (figure 8).

In the whole dataset, the information on the abundance-at age of young ages (mean across all recapture years for a given release) indicates a decline in abundance-at-age 2-5 (and to some extent 6) for the period covered by the data. No particular trend is observed for age 7, and the trend is increasing for age 8 and above. In comparison, the data from recaptures after 1 year show less (or no) trend for younger ages, and similar trends as the whole data set for older ages.

This difference of perception for the age-classes 4-6, which represent a large part of the SSB, could explain the change in the assessment output.

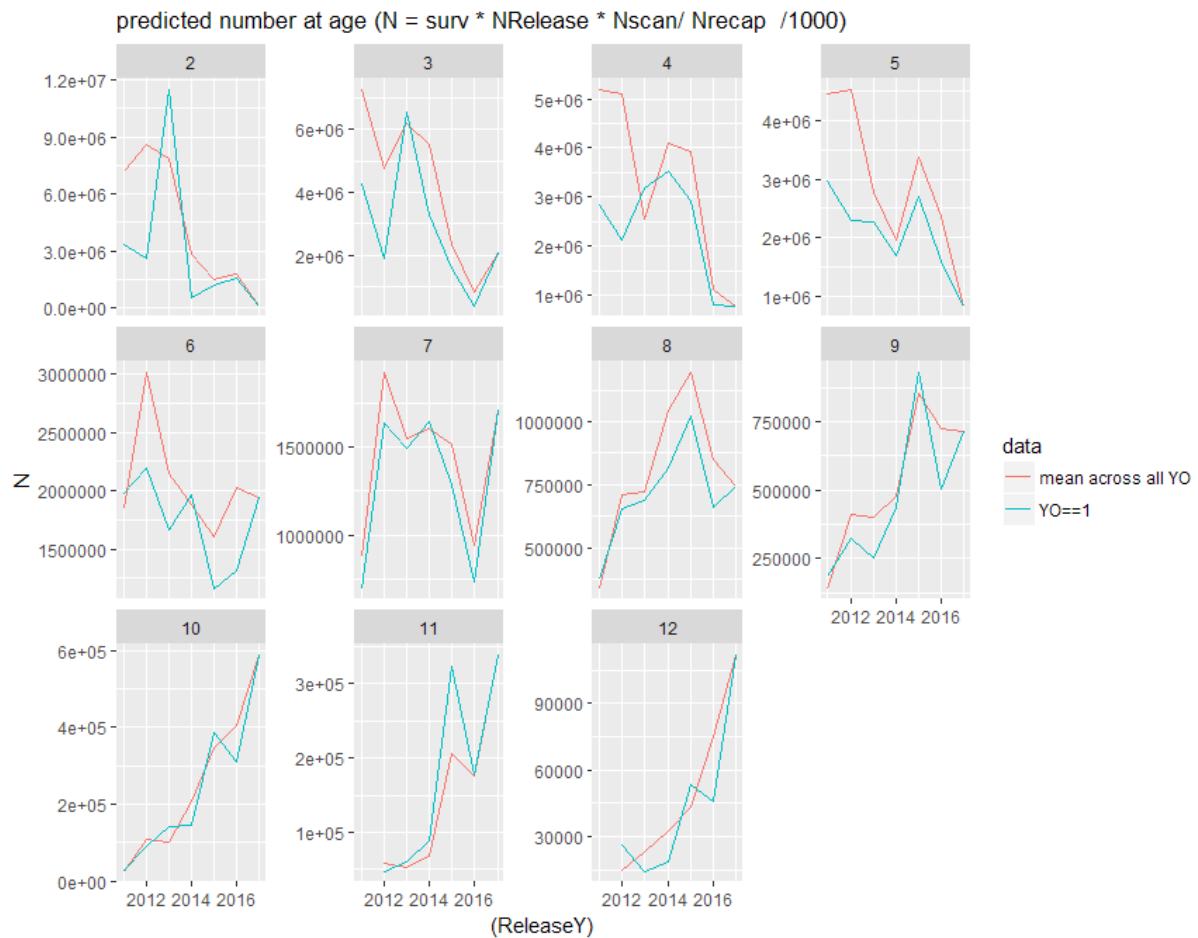


Figure 8 : information from the recaptures of RFID tags (expressed as abundance at age) for the whole data set (mean across recapture years for a given release year) and for the dataset with recapture after 1 year at liberty.

In order to investigate whether the change in the assessment is due to the perception of young age-classes not or less declining in the subset of the RFID data, the assessment was fitted again using the recaptures corresponding to tags having spent 2 years at liberty only.

The information in this other subset of the data is different from the subset corresponding to 1 year at liberty (figure 9). For the 4 years old, the data for tags with 2 years at liberty indicate a strong declining trend (as the whole dataset), but not for ages 5-6 for which there is no specific trend, but large variations. For ages 7+ the trends are similar in the 2 subsets.

Despite of this difference in the information of the 2 RFID subsets, the assessments are very similar. Parameter estimates (including the overdispersion for the RFID tags) are very close for both runs (not shown). Stock trajectories are also very close (figure 10).

Since the information of the RFID from 1 year at liberty and 2 years at liberty is different, and as the model gives a similar output for both subsets, this suggests that the revision of the stock perception when using only a subset of the data is not caused by the difference of information provided by the specific subset of the dataset.

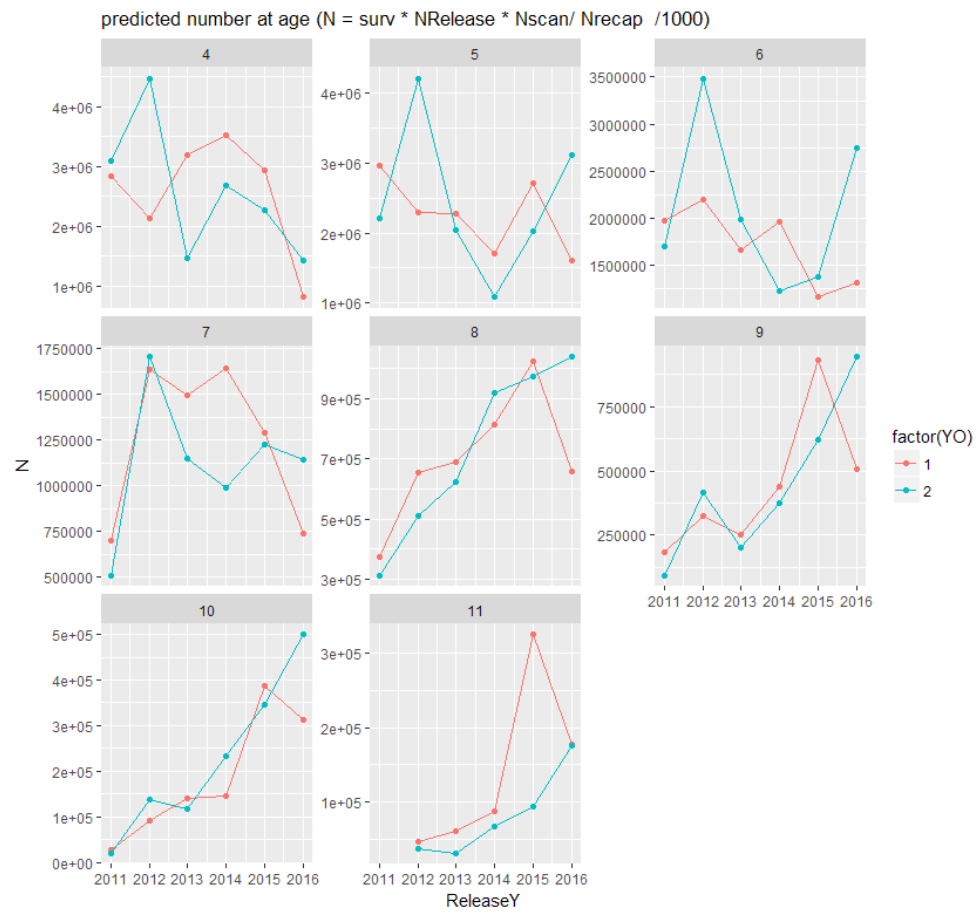


Figure 9 : information from recaptures after 1 or 2 years at liberty

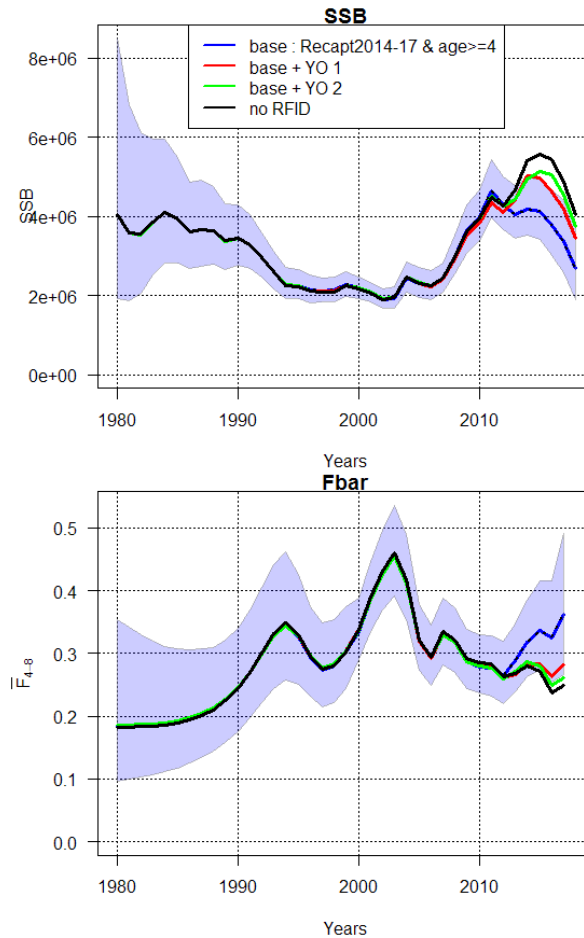


Figure 10 : comparison of the stock trajectories for assessment run with the whole RFID data set, with recaptures after one year at liberty only (YO=1), 2 years at liberty only (YO=2) and no RFID data.

Link between weighting of the data sources and number of data points

In order to understand the link between the weight of a data source on the assessment and the number of data points in this data source, 2 experiments were conducted.

- Using random subsets of the RFID data

The model was run using subsets of the RFID data of different sizes, corresponding to a proportion of the total dataset ranging from 0 to 100% (by increment of 10%). For each percentage X, 30 data sets were generated by sampling randomly X% for the data from the whole RFID dataset, and the assessment was then run for each of these 30 subsets.

The stock trajectories of these 11 x 30 runs are shown on figure 11. The assessment with 100% corresponds to the last accepted assessment at WGWISE (2018), and the assessment with 0 shows the sensitivity of this assessment to the inclusion of the whole RFID data.

One could imagine that as long as enough RFID data is included (say e.g. down to 50%), the information about the development of abundances at age contained in any subset would be similar, and therefore the assessment should be similar to the base case. This is not what is observed, as there is an almost linear progression between the two extremes (0 and 100%) as the percentage data used increases.

Furthermore, for small percentages (10-30%), all randomly drawn subsets likely contain quite different information. Despite of that, the estimated stock trajectories are quite consistent among the 30 runs (narrow envelopes).

This suggest the weight of the RFID tagging data has on the assessment does not only depend on the signal it gives on abundance-at-age through the years, but also strongly depends on the number of data points in the dataset, which is very high compared to other data sources.

The assessments presented above, using only data from recaptures occurring either after 1 year at liberty or after 2 years at liberty use respectively 30% and 25% of the number of data points in the RFID data set. The stock trajectory for the assessment using YO=1 is sensibly different from the median of the 30 runs with random samples of 30% of the dataset (lower SSB and higher Fbar for 2016 to 2018). This suggests that the subset YO=1 is different from a random sample of 30% of the whole dataset, and gives information that leads to a slightly earlier and steeper decrease of the stock.

This is not the case for the subset using YO=2, for which the assessment run is quite similar to the media of the runs using 20% of the data.

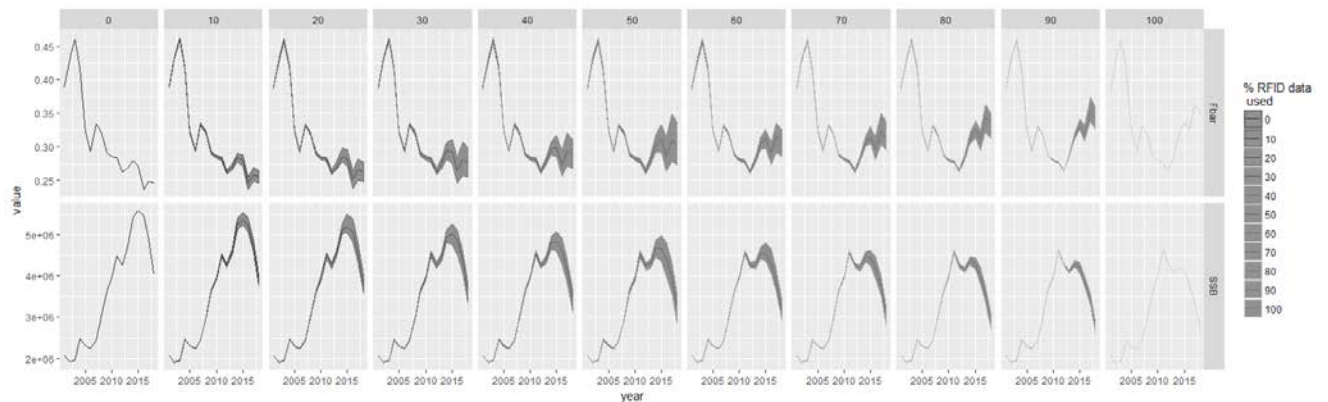


Figure 11 : stock trajectories from assessment runs using random samples of the RFID dataset. Sample size varies from 0 to 100% of the whole dataset. For each percentage, 30 samples are generated, and the assessment run for each of them (shaded areas representing the 90% envelop of these 30 runs).



Figure 12 : the ssb and Fbar (median across 30 runs) for subset of the RFID data of different sizes, and for the assessments using only recapture after 1 year at liberty (YO=1) or only recaptures after 2 years at liberty (YO=2).

- Using a survey index twice

The opposite type of experiment was also carried out. The IESSNS index was duplicated by using the same data twice (as 2 separate surveys, but with the same catchability and variance parameters). To achieve convergence, the model had to be run without correlated observation errors.

Repeating the information from the IESSNS had an effect on the estimated stock trajectories, with the assessment using the survey twice giving a higher stock and lower Fbar in the recent years (figure 13). This assessment also had lower observation variances for the IESSNS survey than the base case assessment (figure 14). In the conflict between the different data sources (IESSNS pulling the SSB upwards and egg survey and tags pulling it downwards), the weight of the different data sources therefore appears to be to some extent depending on the amount of data points of each survey.

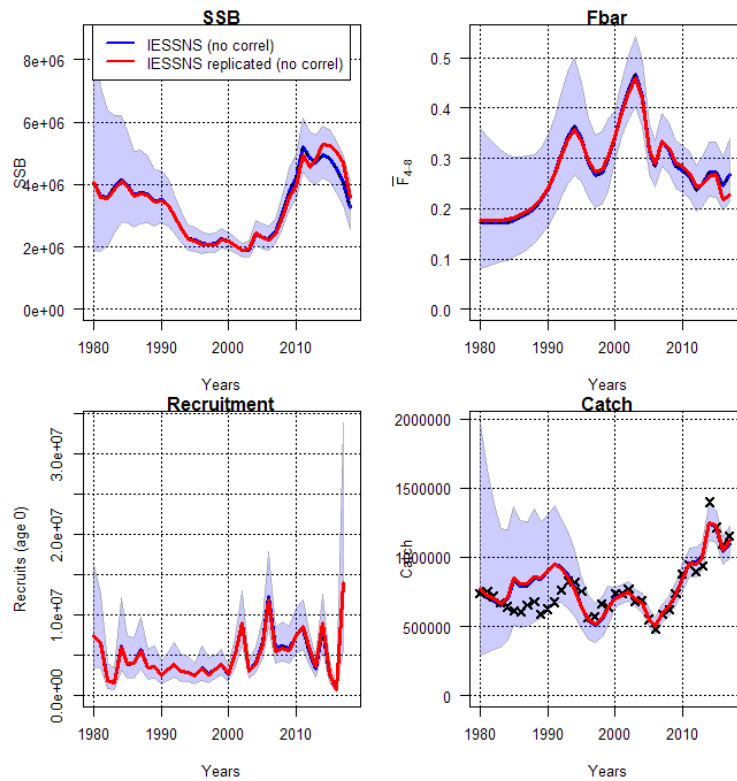


Figure 13 : effect on the assessment of using the IESSNS twice

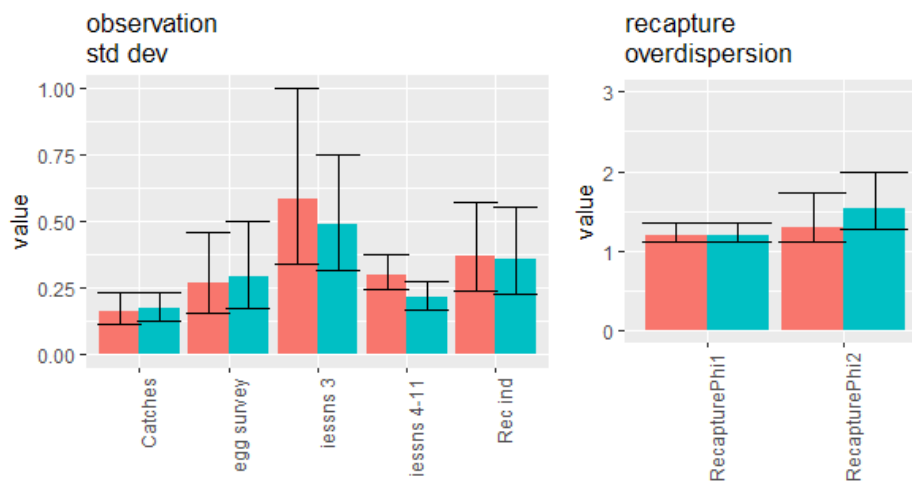


Figure 14 : effect on the estimated parameters of using the IESSNS twice in the assessment

Conclusions

This series of analyses and exploratory runs lead us to the following conclusions :

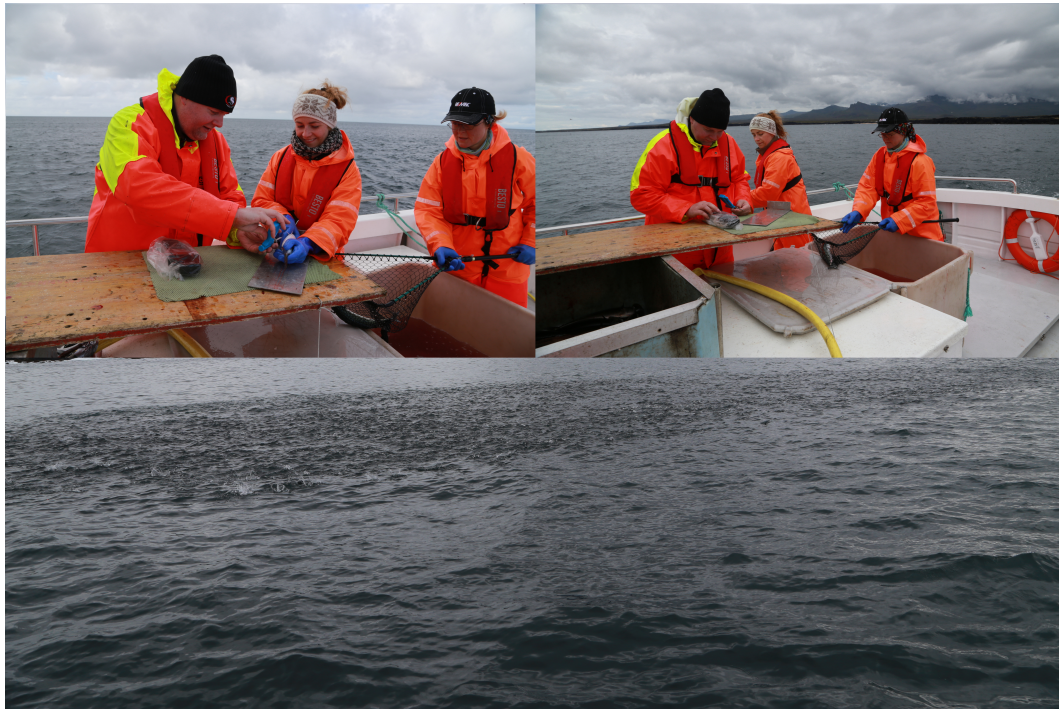
- There are potential sources of bias in the RFID data. A potentially serious one is the effect of tag loss, causing a spurious declining trend for abundance at age in the RFID data.
- While tag loss could be dealt with by modifying the model formulation a more pragmatic approach consists in using only a subset of the data including data that have been subject to the same amount of tag loss, i.e. having spent the same number of years at liberty
- Analysis of the RFID data show that the dataset with one year at liberty has the highest quality compared to higher numbers of years at liberty (which are less consistent both internally and among each other)
- The weighting of each data sources in the assessment does not only depend in the information provided by the data (and how it conflicts or agrees with the rest of the data) but also on the number of data points. As the RFID data is still no mature, the number of data points added each year continues to increase, thereby increasing the weight of the data on the assessment

Using a subset of the RFID data as suggested in this working document appear to be a way to remove the sources of bias identified, and control the weight of the RFID data on the assessment. Selecting YO=1 or YO=1 and 2 ensures that the number of data points added at each update assessment is the same for each year. It also removes the less informative (based on the consistency analyses) data.

However, one drawback of using the selection proposed here is that the RFID time series is shortened by 2 years (first release year in 2013), which may imply that the corresponding model parameters (survival rate and over dispersion) will be accurately estimated only in a couple of years, when more data will be available.

NEA Mackerel

Analysis of assessment based on tagging data



Höskuldur Björnsson MRI Iceland

March 2nd 2019

1 Introduction

As known, considerable discrepancy exists between data sources in the Mackerel assessment where that tagging data seem to have rather large effect, at least leaving out the tagging data changes the stock size in recent years.

To look at sensitivity of the model the input data were changed and the effect on the results investigated. The first test was to double all indices in the pelagic survey 2018 , and check the model response. The result was rather surprising, the SSB in 2018 changed from 2354112 to 2332762 . Unchanged!!

Log likelihood in those 2 cases was.

Normal index

SAM model: log likelihood is -2662.876 Convergence OK

double index 2018

SAM model: log likelihood is -2666.232 Convergence OK

Introducing a high SSB index 2018 (4.8 million) leads to ssb of 2716021, some increase.

The problem must partly be how short the pelagic survey is . The tagging series is also short but somehow seems to take over. Model without tagging data responds to this artificial doubling of the last pelagic index (as expected) .

Tagging data were also added to the muppet Model that was used for the HCR evaluations of mackerel in 2017. The tagging data were modelled in the same way as in the SAM model and the result was the same, not sensitive to doubling of the most recent pelagic survey. Two other tagging data alternatives were added to the Muppet model, both based on using more aggregated measures than used in SAM today. Also tagloss was added to both the Muppet and the SAM model as there were indications that observed recaptures were systematically below predicted recaptures for tags that had been out for a long time. (figure 1)

The tags in the Muppet model are modelled in the following way, mostly the same as in the SAM model.

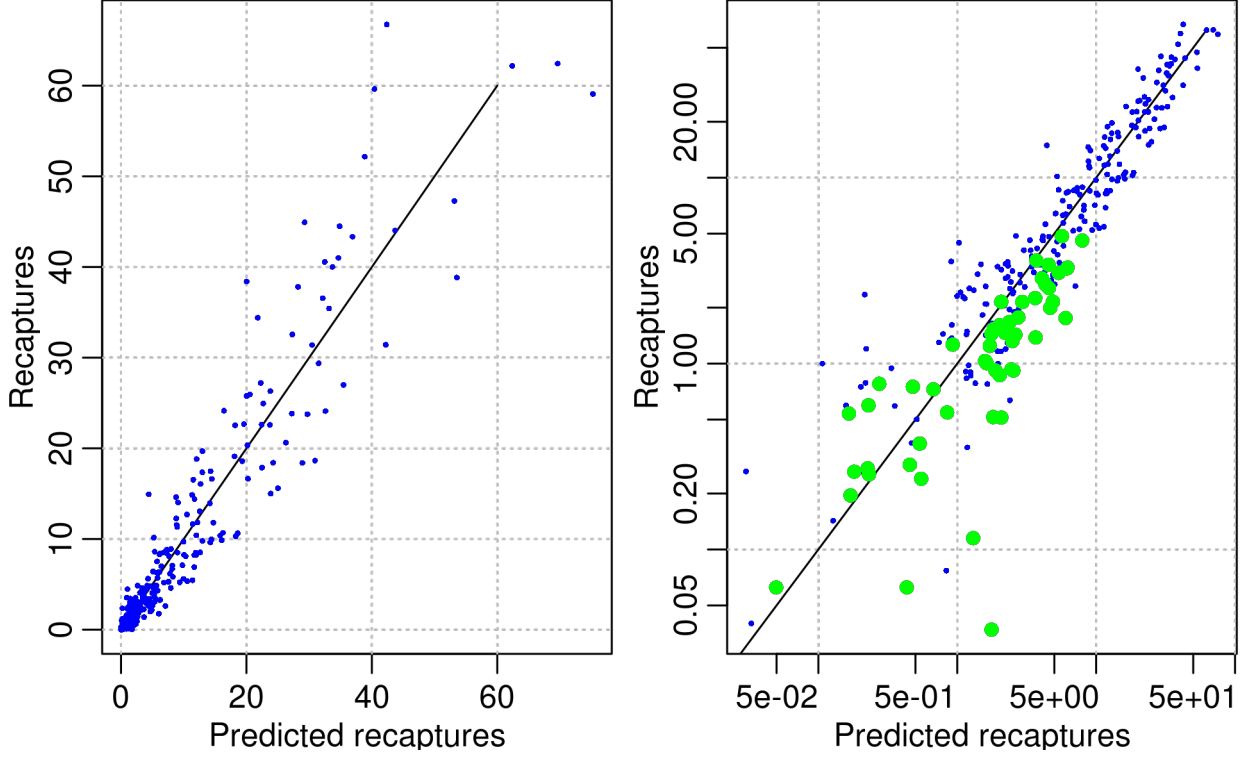


Figure 1: Observed vs predicted recaptures. The green points are tags that have been out for more than 5 years. Tagloss not included in the model behind this figure.

	1	2	3	4	5	6	7
predRecap < Recap	51	45	25	14	13	6	4
predRecap > Recap	23	19	29	30	20	16	7

Table 1: Comparison of predicted recaptures and recaptures as function of time out. Tagloss not included in model behind this table.

$$r_{tY, rY, Yc} = \frac{R_{tY, Yc} N_{scan, rY, Yc} e^{-tM} e^{-tL(rY - tY)}}{N_{tY, Yc}} \quad (1)$$

$$\ell_- \sum \log(NB(r_{tY, rY, Yc}, r_{tY, rY, Yc}, \Gamma)) \quad (2)$$

where tY is tagging year rY is recapture year, Yc yearclass, tM tagging mortality, tL annual tagloss $r_{tY, rY, Yc}$ predicted recaptures, $R_{tY, Yc}$ number tagged of yearclass Yc in year tY and Γ the dispersion parameter in the negative binomial distribution NB .

Additional option is to use more aggregated data.

$$r_{tY, rY} = \sum_{Yc} r_{tY, rY, Yc} \quad (3)$$

$$r_{tY, rY} = \sum_{Yc} r_{tY, rY, Yc} \quad (4)$$

$$\ell = \sum \log(N(\log(\frac{r_{tY, rY}}{r_{tY, rY}}), \sigma_T^2)) \quad (5)$$

σ_T is an estimated parameter and N means Normal distribution. Additional possibility is to use negative binomial distribution instead of lognormal on the aggregated data. This option will give more weight to the high values compared to the lognormal option.

$$\ell_- \sum \log(NB(r_{tY, rY, Yc}, r_{tY, rY, Yc}, \Gamma)) \quad (6)$$

Advantages and disadvantages of aggregated vs more disaggregated data are.

- The tagging data do get less weight. Might be too low for the aggregated data but too high for the disaggregated data.
- Loss of information. Age composition gives useful information.
- Not dependent on the double use of age - length keys that I think is somewhat questionable (See appendix A).

Tag loss could be an important parameter as shown in figure 1 and table 1. Including tag loss leads to rather high value ≈ 0.17 per year and the change in the objective function is -27.58 for one parameter based on type 1 likelihood and -12.76 for type 2. Both highly significant change but not enough to let the pattern with "time out" disappear. The weight of different data points is of course not even so a complete balanced ones not appear. The estimated tag loss is similar in the aggregated models and the improvement in the objective function is similar there.

Figure 2 shows some result testing different configurations. The models are always run with the data used for the assessment 2018 with the addition of tag data 2018 and the new recruitment index. The new recruitment index was not just the old index with an additional year but introduced some changes in older values. The only difference in results caused by the new recruitment index relates to the 2017 yearclass that is estimated to be relatively large and affects the spawning stock in 2019 at age 2. In figure 2 treatment of tags is referred to as type 1, type 2 and type 3, referring to equations 2-6

The use of more aggregated tagdata leads to very similar results (type 2 and 3) but give quite different results from type 1 when tag loss is not modelled. With tag loss included all the runs lead to similar results in terms of SSB and similar to what is obtained when not including the tags. As described earlier inclusion of tagloss leads to considerable improvement in fit both in SAM and Muppet.

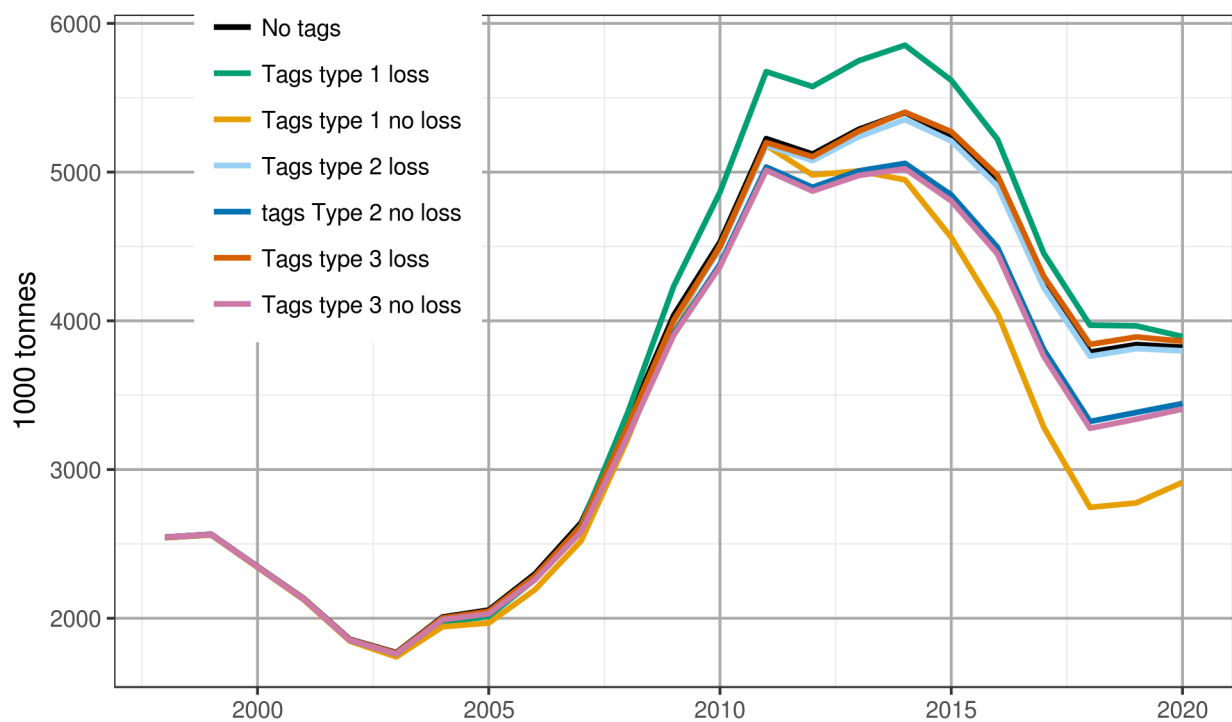


Figure 2: Development of spawning stock from a number of different runs

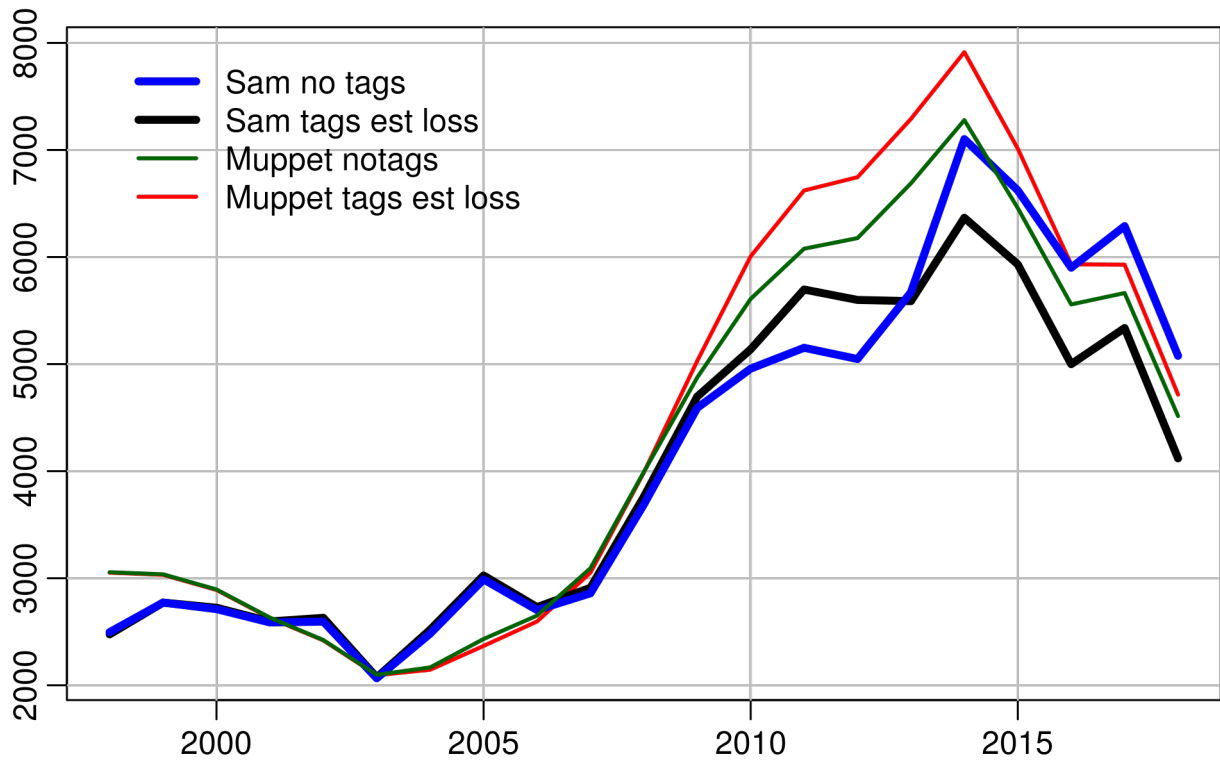


Figure 3: Development of biomass 3+ in the beginning of the year from a number of different runs

One difference between the Muppet and SAM model is treatment of the catch data in 1980-1998. In SAM the catch data are seriously down weighted but in Muppet a multiplier on the catches before 1999 is estimated/or fixed but the catch at age data used. The value of the multiplier has no effect on stock trends after 1999. It does not affect SSB in the assessment year but can affect SSB in the prediction year through age 2 in the prediction year that affects the SSB. Advice will not be affected by the multiplier when $SSB > B_{pa}$. The estimated value of the multiplier is small (1.05-1.25 depending on the run)

The main features of the Muppet results is similar to what is obtained from SAM both in terms of Tag/No tag difference and the effect of "Tag loss". SSB from the models is not exactly comparable as proportion of M and F before spawning are not the same and maturity and weight at age in 2018 are not exactly the same. Looking at B3+ compiled with the same weights some difference can be seen, SAM does probably have more flexibility in deviating from "converged stock assessment" (figure 3)

Introducing tagloss in SAM increases in loglikelihood from -2777 to -2747 for practically parameter as the tagloss for the steeltags was -3.64 (0.026 per year) but -1.69 (0.17 per year) for the RFID tags. Most of the change in likelihood does therefore come from the RFID tags as the tagloss for the steel tags is estimated as very small. Spawning stock 2018 changes from 2.6 to 3.3 million tonnes with the introduction of this one parameter. (this is based on a run using 2018 recapture)

There is a problem with the SAM run that it does not converge. "SAM model: log likelihood is -2747.593 Convergence failed". There were similar problem with the Muppet model (written in ADMB) where the model had to be assisted in some cases to a solution. I am not sure how to do it in SAM but multiple local optima might be a problem (overparameterisation).

What is cause for high tag loss for the RFID tags compared to the steeltags? Tagging data from demersal fishes (internal tags) do often indicate considerable tagloss so this should not be a surprise. But there is a comparison with the same specie, using internal tags. Is there a possibility that the tags stay in the fish but deteriorate somehow with time so they can not be identified???

Estimated parameters from SAM with tagloss are shown below. (In ADMB I would never get standard error on solution that did not converge)

141	logTagmort	2.1489e+00	1.2291e-02
142	logTagdispersion	3.7333e+00	3.8509e-01
143	logTagloss	nn -1.6920e+00	1.5007e-01

One thing that needs to be checked in the effect of adding the 2018 tagging data that leads to considerable increase in stock size when tag loss is estimated and type 1 likelihood function is used. Looking at the muppet model adding 2018 recaptures and 2017 tagdata increases SSB_{2018} from 2805 to 3970 (similar development in

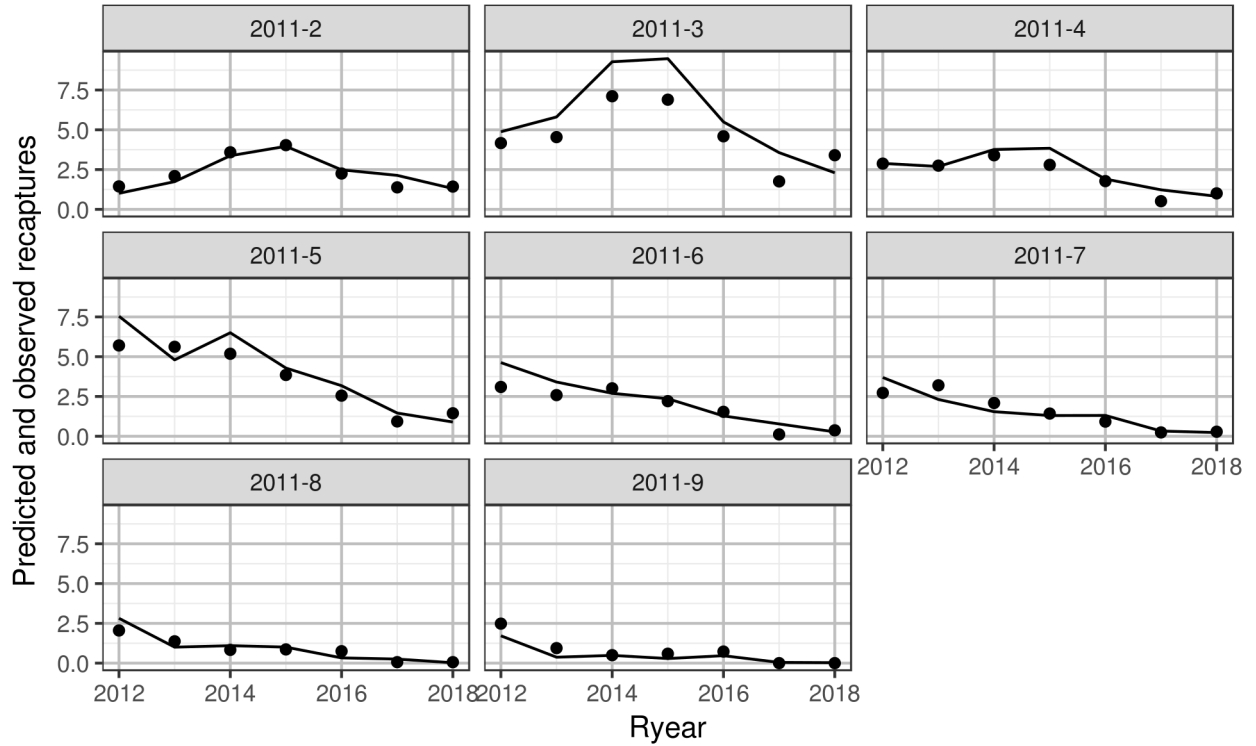


Figure 4: Predicted and observed recaptures from the 2011 tagging.

SAM). Looking at the composition of the objective function shows that the solution adding the 2018 data fits the pelagic survey better but the catch composition not as well. Using type 2 tag likelihood then change in SSB in 2018 is from 3762 to 3476, of course tags have less weight here.

To summarize introducing tagloss does seem like a logical change in the assessment model, even though we do not know what causes the tagloss. Understanding the high tagging mortality in the RFID tags is of course even larger problem.

What happens when tag loss is introduced is that tagging survival does increase. The increase is though not very large or from 0.110 to 0.117 in the muppet model using aggregated tags (the change in SAM is similar). This leads to increased number in stock according to the most recent tagging experiments but decreased number in stock from the older tagging experiments. ??? The total effect does though seem to be considerable.

The most logical way to proceed is though to use the more aggregated measures, i.e 1 value per tagging year and recapture year. Ratio of observed and predicted recaptures look reasonable, perhaps an indication that

2 Appendix. Problems with aging

There might be a problem with treatment of the age data in the tags after we have stopped having the possibility of ageing the recaptured fish. The problem is best described by a simple example. We tag a 35cm fish in 2012, that gets an agedistribution between 4 and 9 years. We recover it 5 years later with an age distribution of between 9-14 years but we scan extremely few 14 years old fishes. Tests have shown a bias in age reading of old mackerel (ref??) and introducing old fish this way can potentially affect the results.

This leads to rather low stock size of 9 year old fish in the tagging year as we scan very few or no 14 years old fish 5 years later but recapture some 14 years old fish. $N = \frac{N_{scan}R}{r}$. What do catch in numbers data tell us about Z of old fish. They must be available by age even though 12+ is used in the assessment.

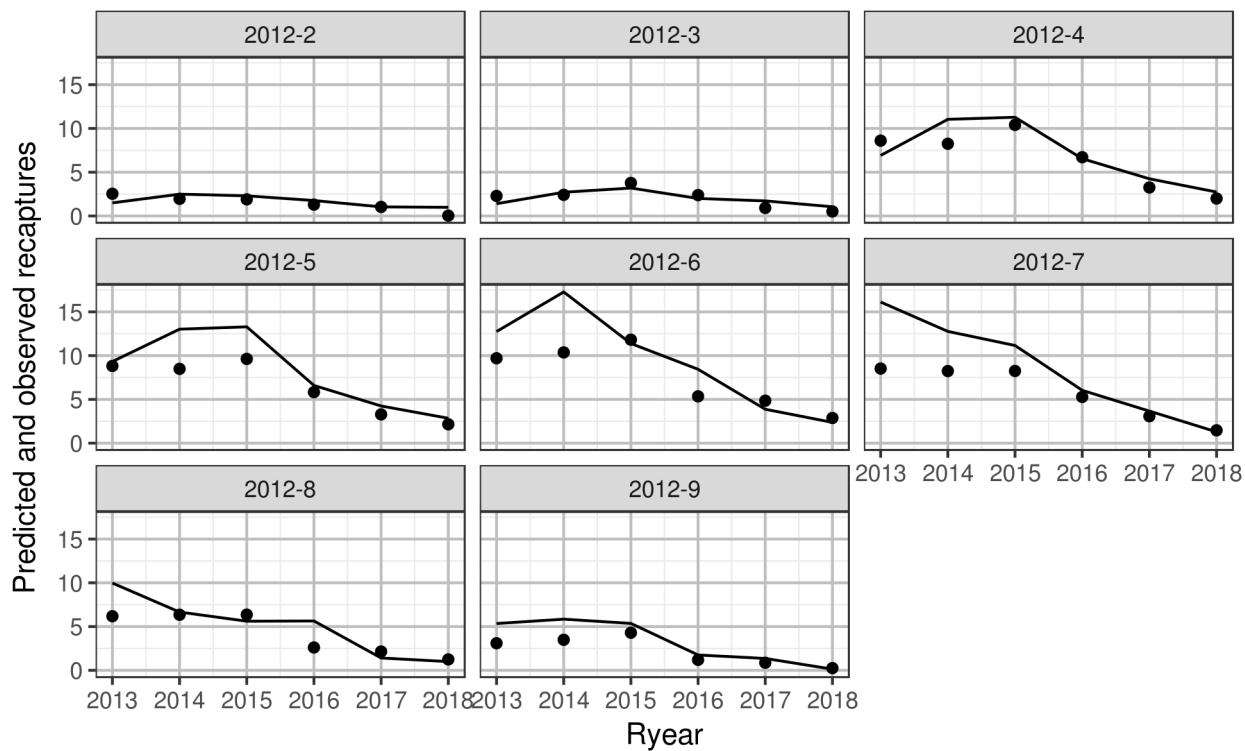


Figure 5: Predicted and observed recaptures from the 2012 tagging.

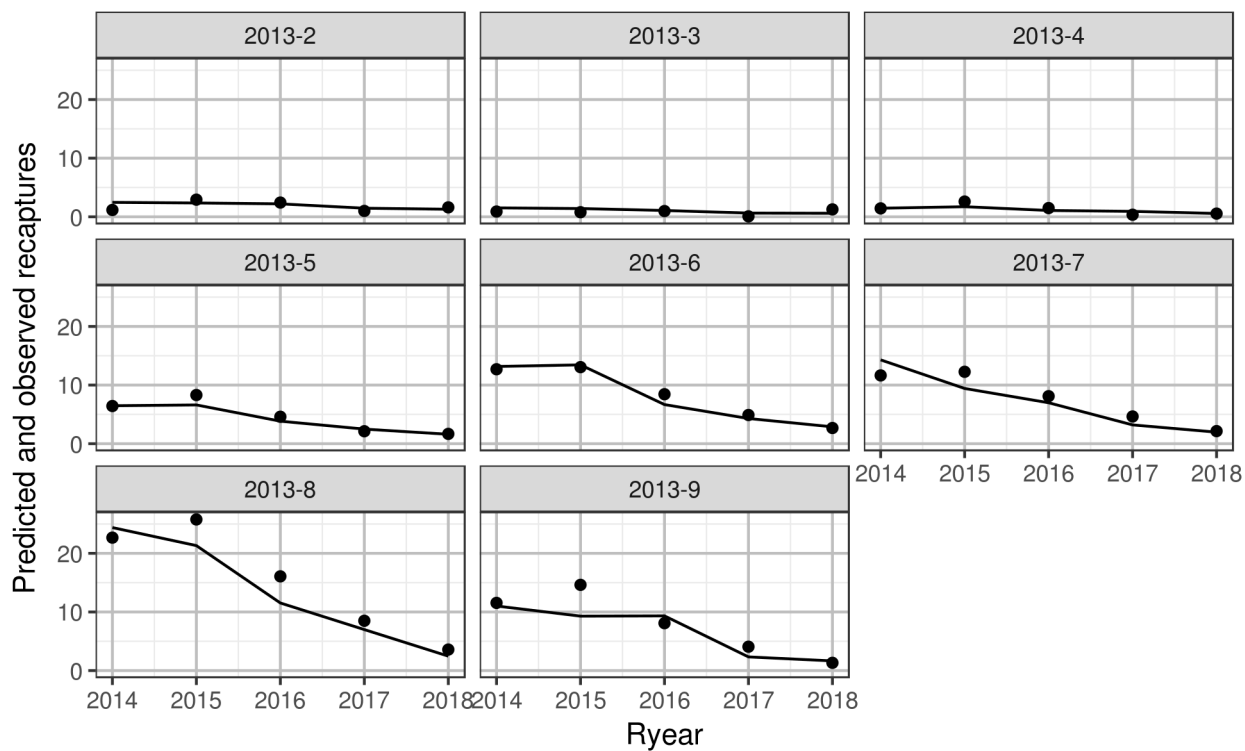


Figure 6: Predicted and observed recaptures from the 2013 tagging.

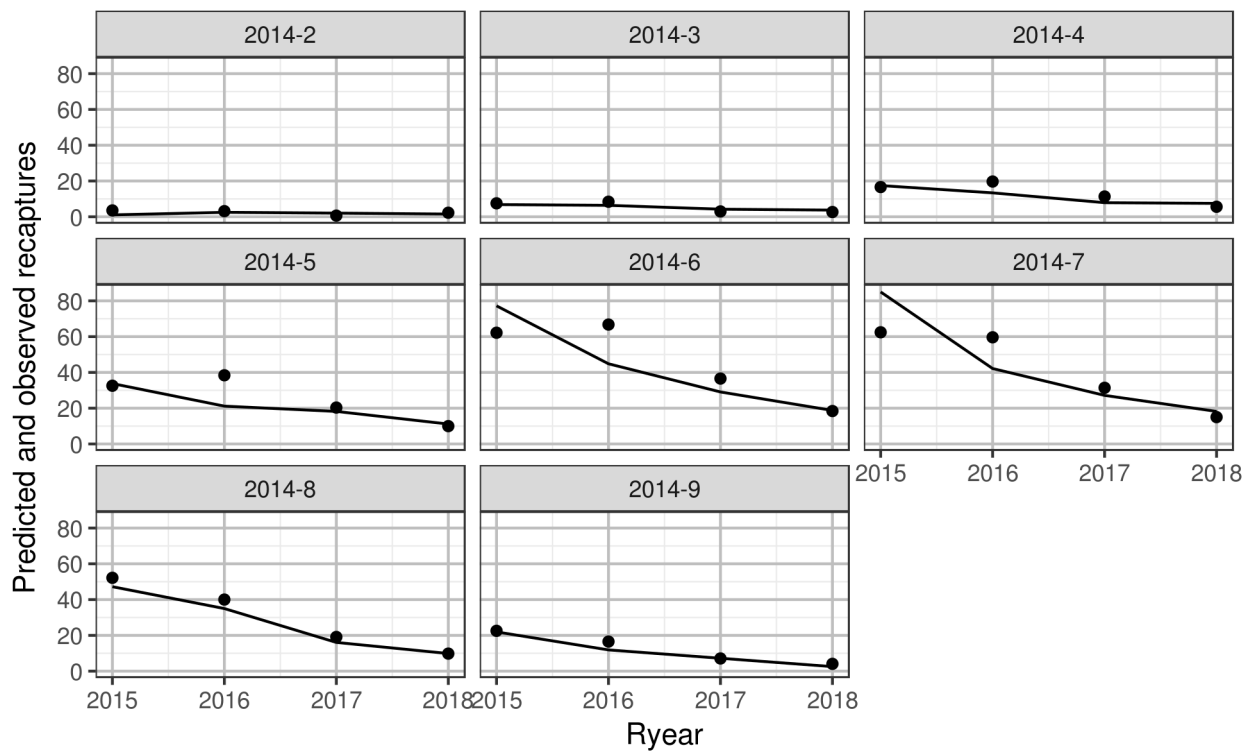


Figure 7: Predicted and observed recaptures from the 2014 tagging.

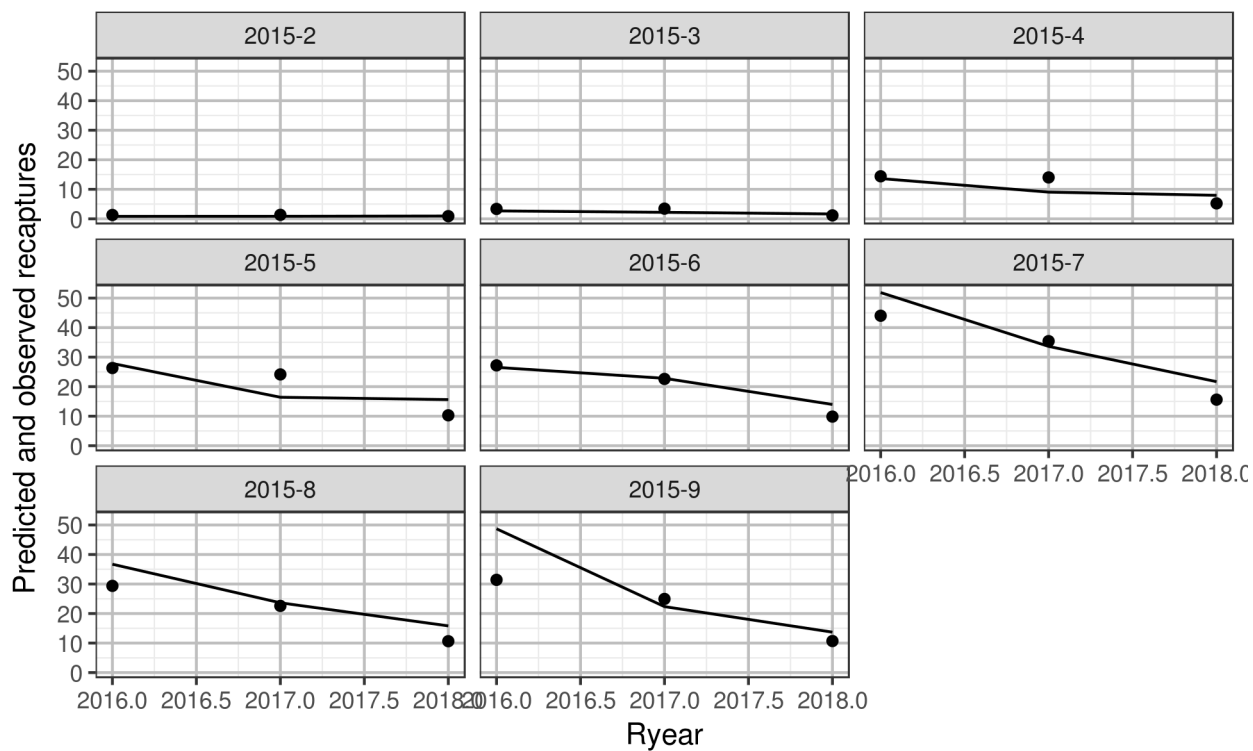


Figure 8: Predicted and observed recaptures from the 2015 tagging.

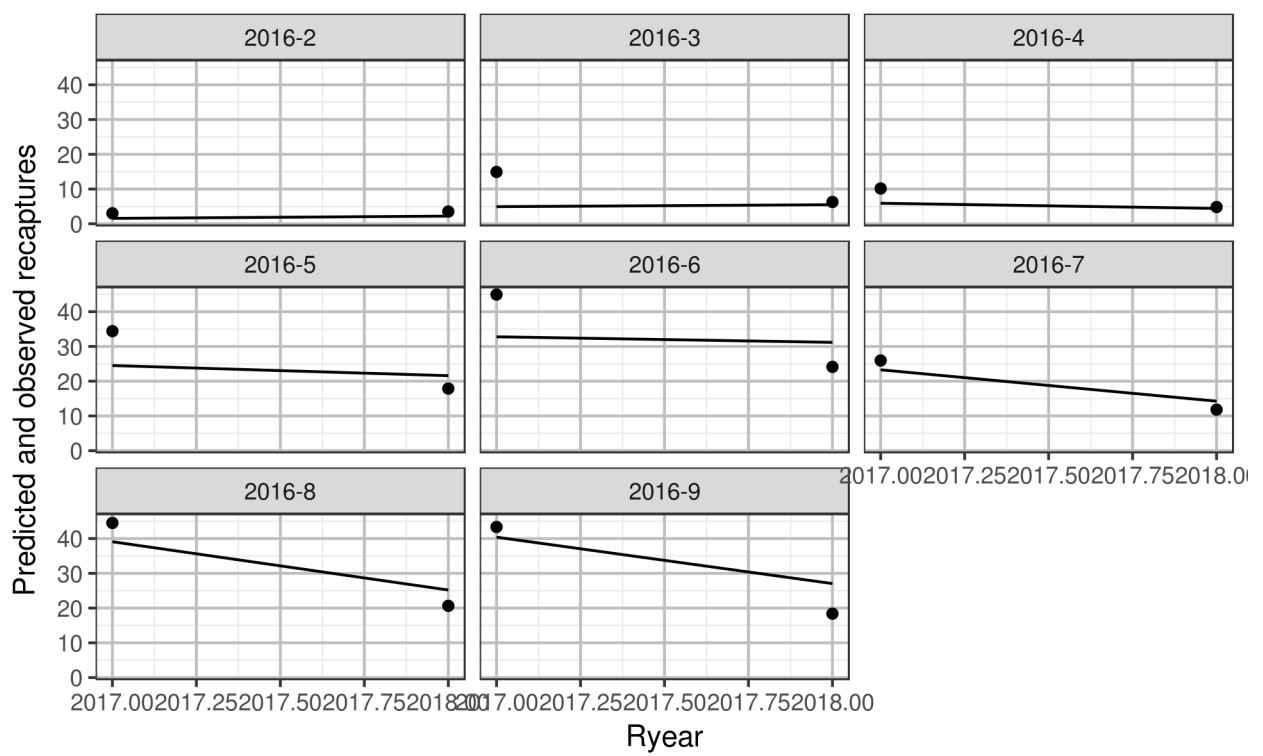


Figure 9: Predicted and observed recaptures from the 2016 tagging.

Changes in catch at age distributions of mackerel 2005-2016

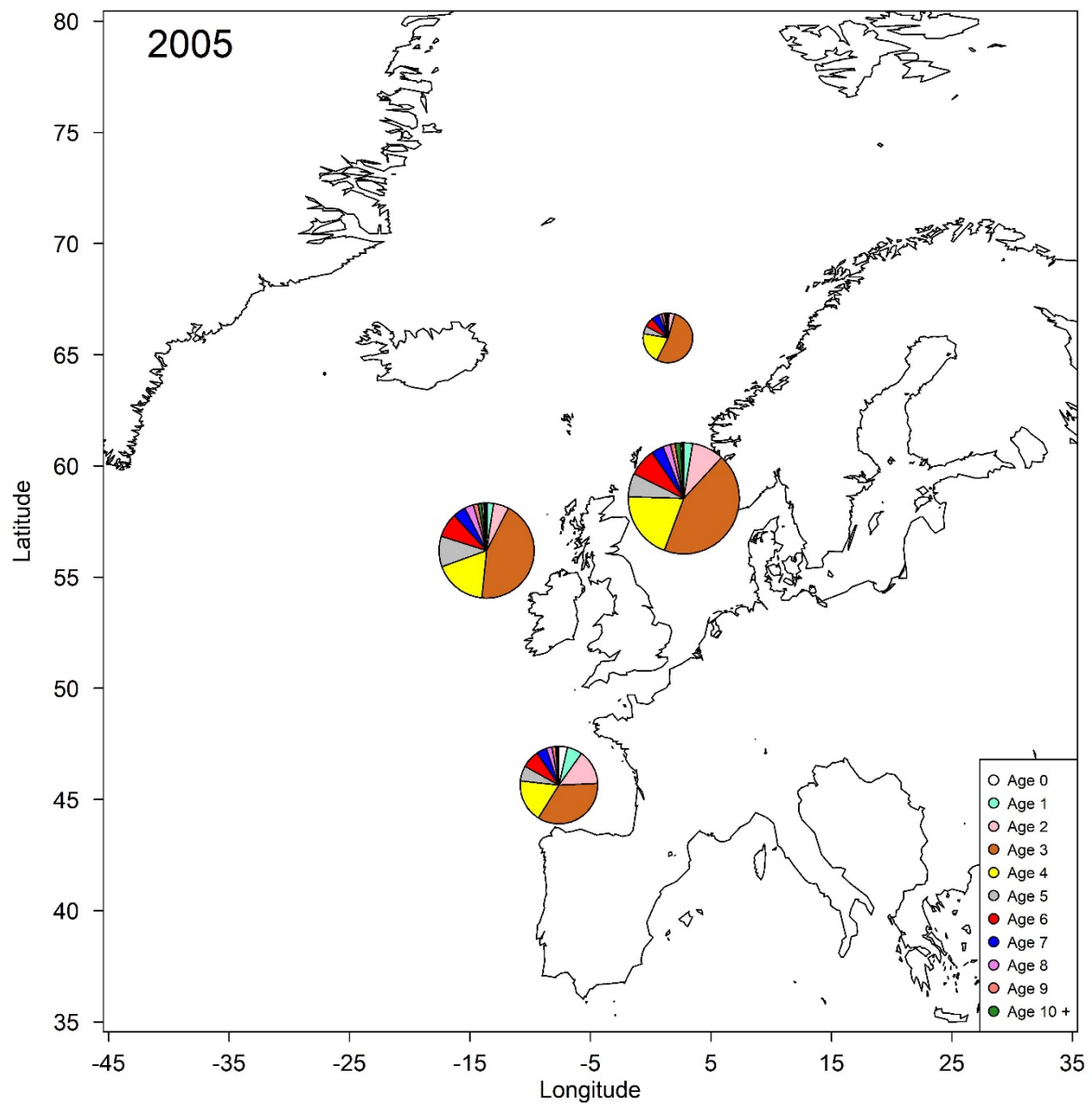
Thassya C. dos Santos Schmidt and Aril Slotte

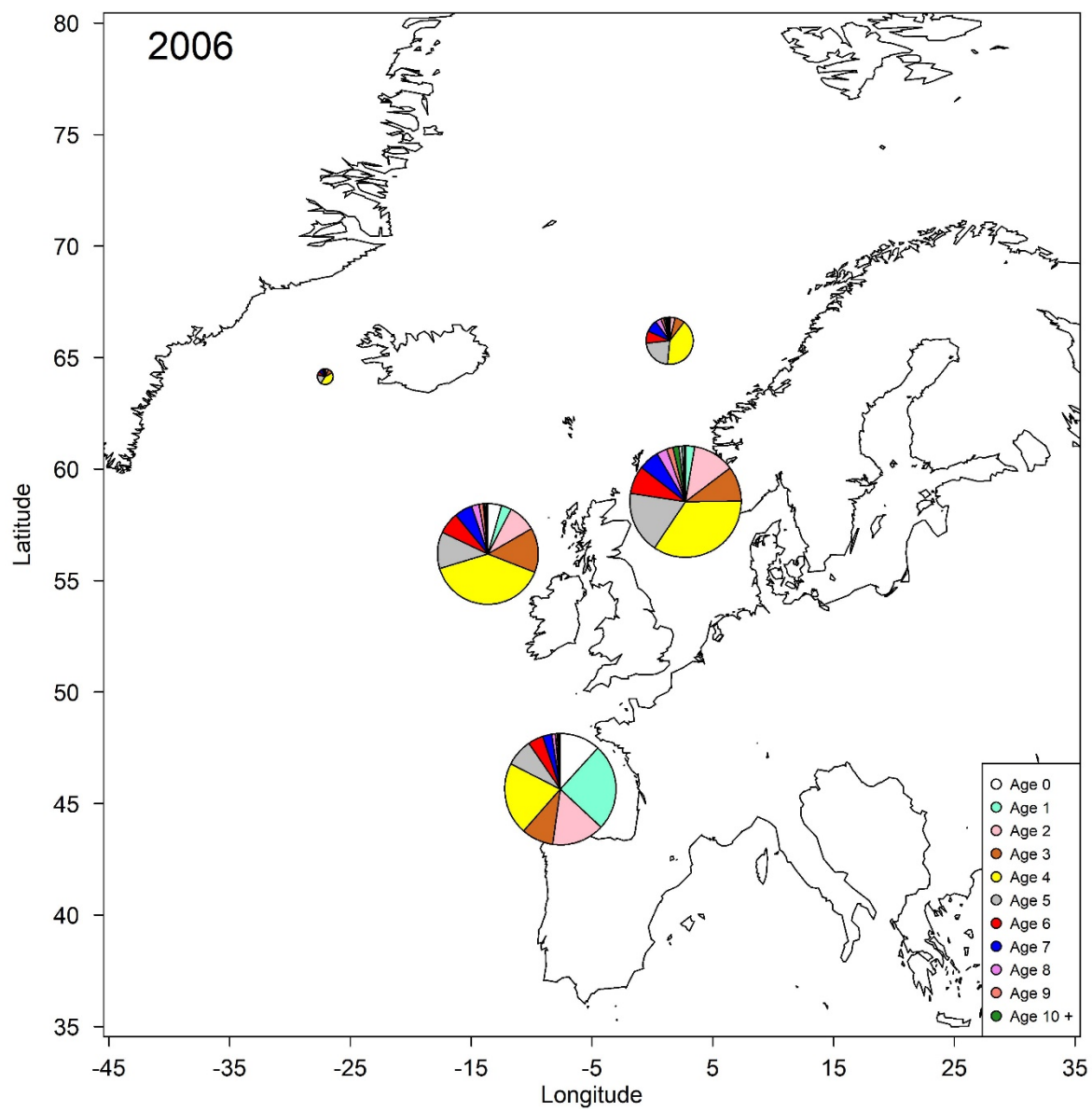
With the aim to visualize the change in catch by age distributions over the period 2005-2016, we have made annual maps with cake diagrams for 5 main catch areas: Iceland (ICES areas Va, Vb, Vb1, XII, XIVa, and XIVb), Norwegian Sea (ICES areas IIa and IIb), North Sea (ICES areas IIIa, IIIb, IIIc, IIId, IVa, IVb, and IVc), West of Scotland-Ireland (ICES areas VIa, VIb, VIIa, VIIb, VIIc, VIIf, VIIf, and VIIk), West of France-North of Spain-Portugal (ICES areas VIId, VIIe, VIIf, VIIh, VIIa, VIIb, VIIc, VIIc.E, VIIc.W, VIId, VIIe, IXa, IXa.CN, IXa.N, and IXa.S). For each year the size of the cake diagram represents relative size of total catch within the year, whereas catch at age in within the area is shown as proportions in the cake diagram.

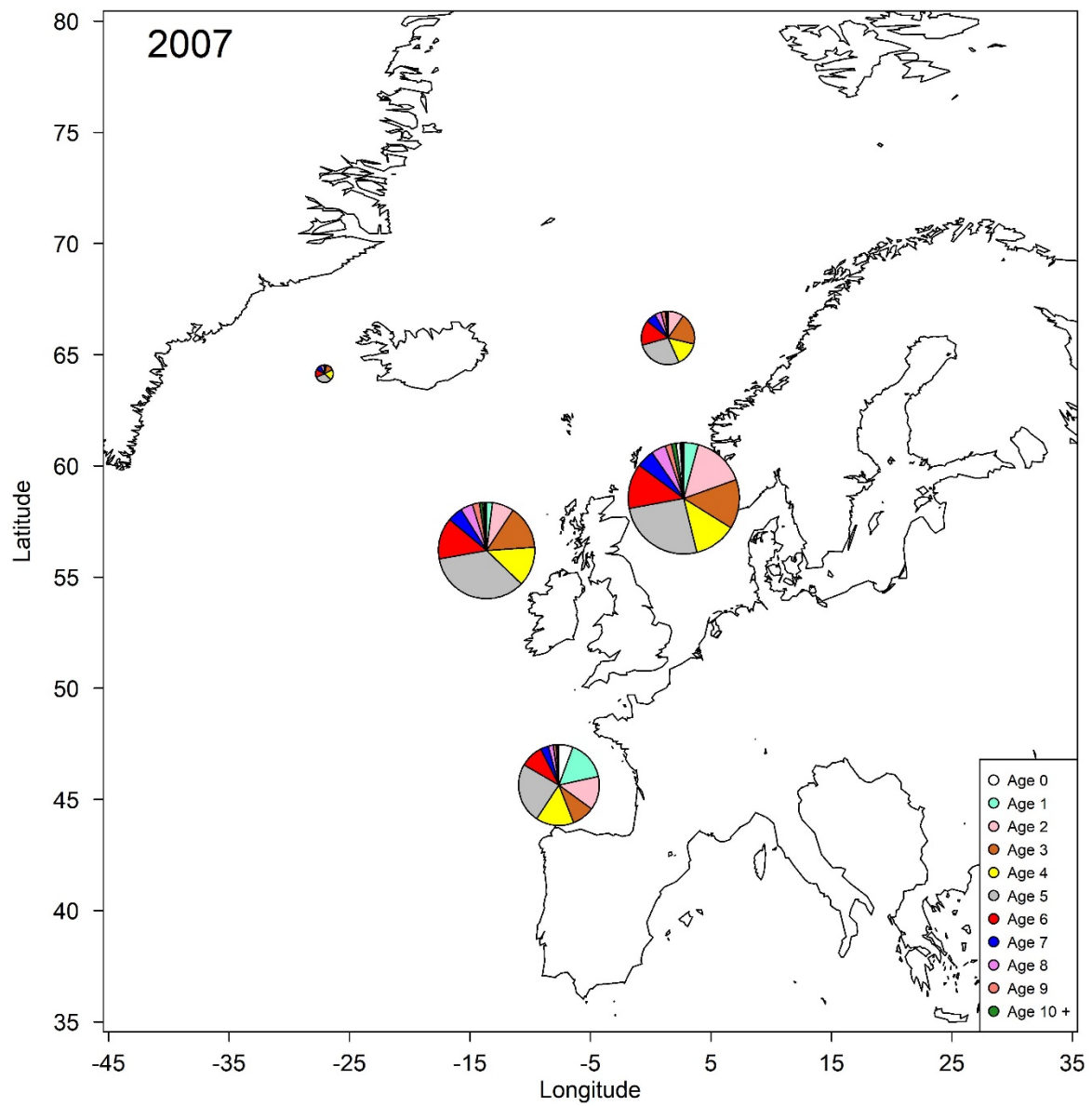
The annual cake diagrams demonstrate the northward shift in main fisheries from 2005 to 2016. At the same time, it shows that the fishery on younger fish (age 3-5) has been variable, with a tendency to be higher in the first part of the data set. In addition there is a tendency that the fishery of young fish of more recent large year classes such as 2010-2011 has been happening more in the North Sea, Norwegian Sea and Iceland than at spawning grounds along Scotland, Ireland and southwards, as demonstrated by the difference between catch areas in proportions of age 3-5 in 2013-2014.

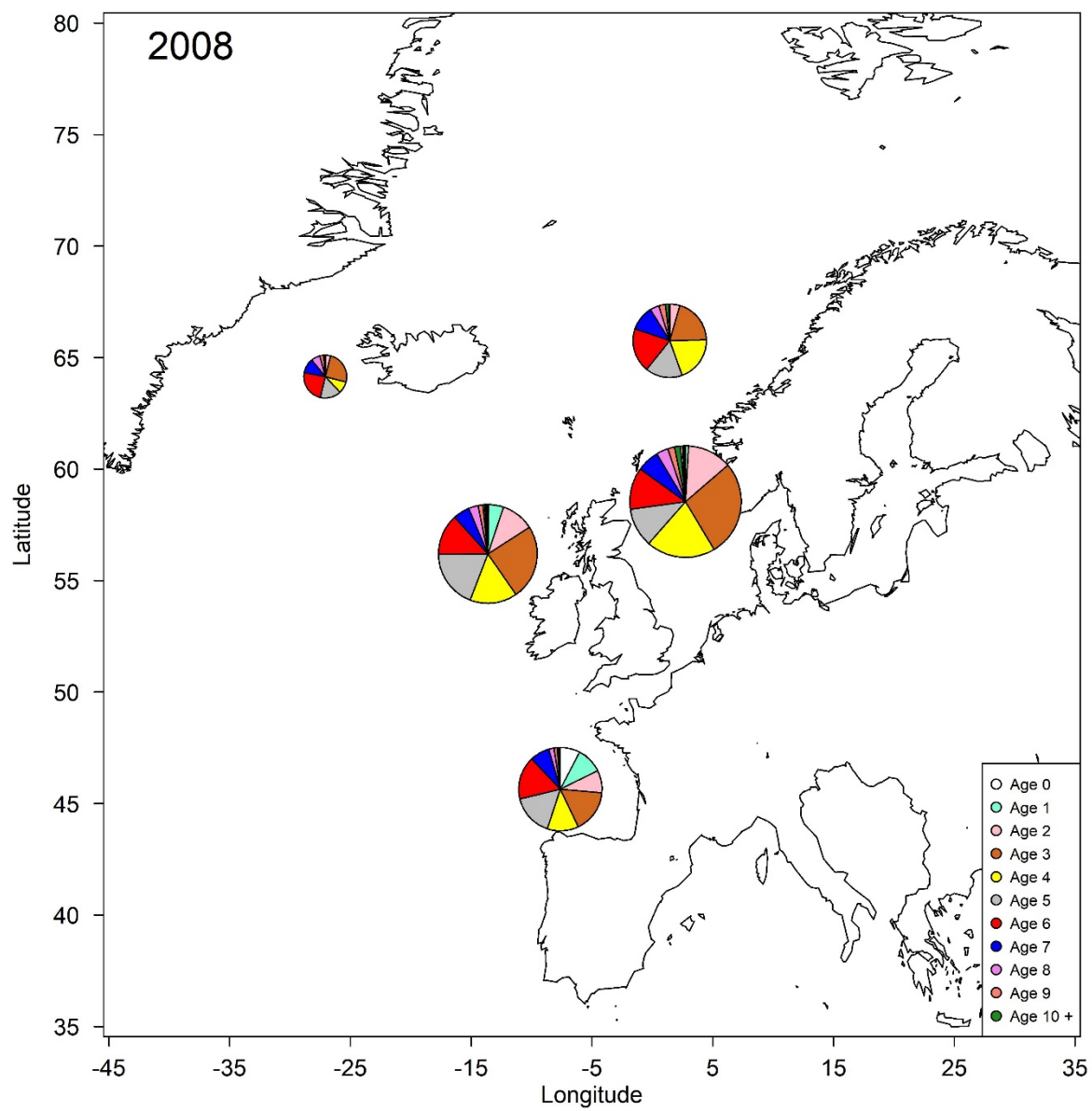
These changes in catch at age distributions could be indicative of a northward shift of distribution, both spawning, nursery and feeding, and it could imply that indexes used in the mackerel assessment, such as the egg survey, recruitment index, IESSNS survey and tagging data, could have been influenced, perhaps with biases in observed trends in the data.

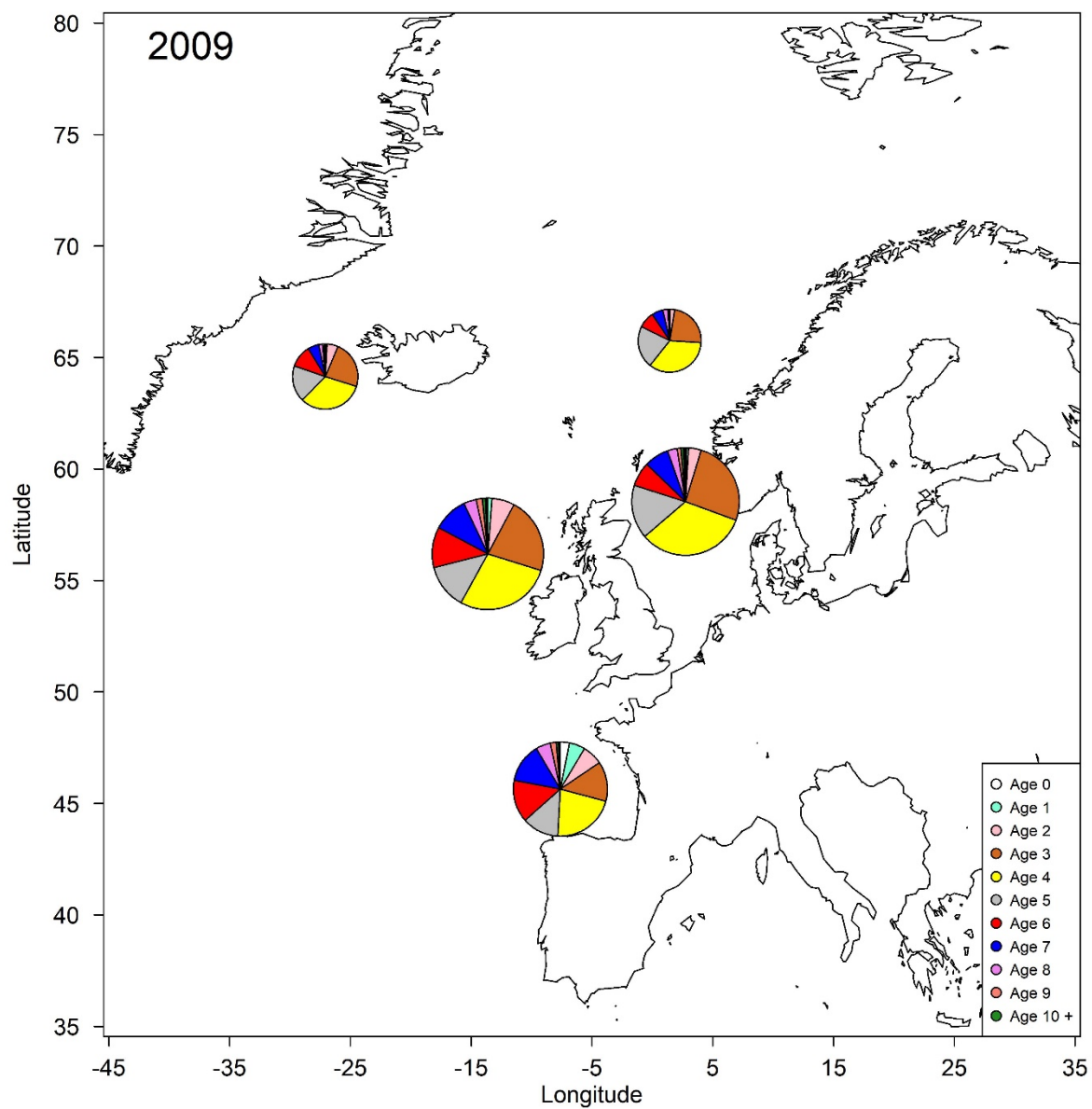
Catch at age data for this analysis were supplied by Andrew Campbell.

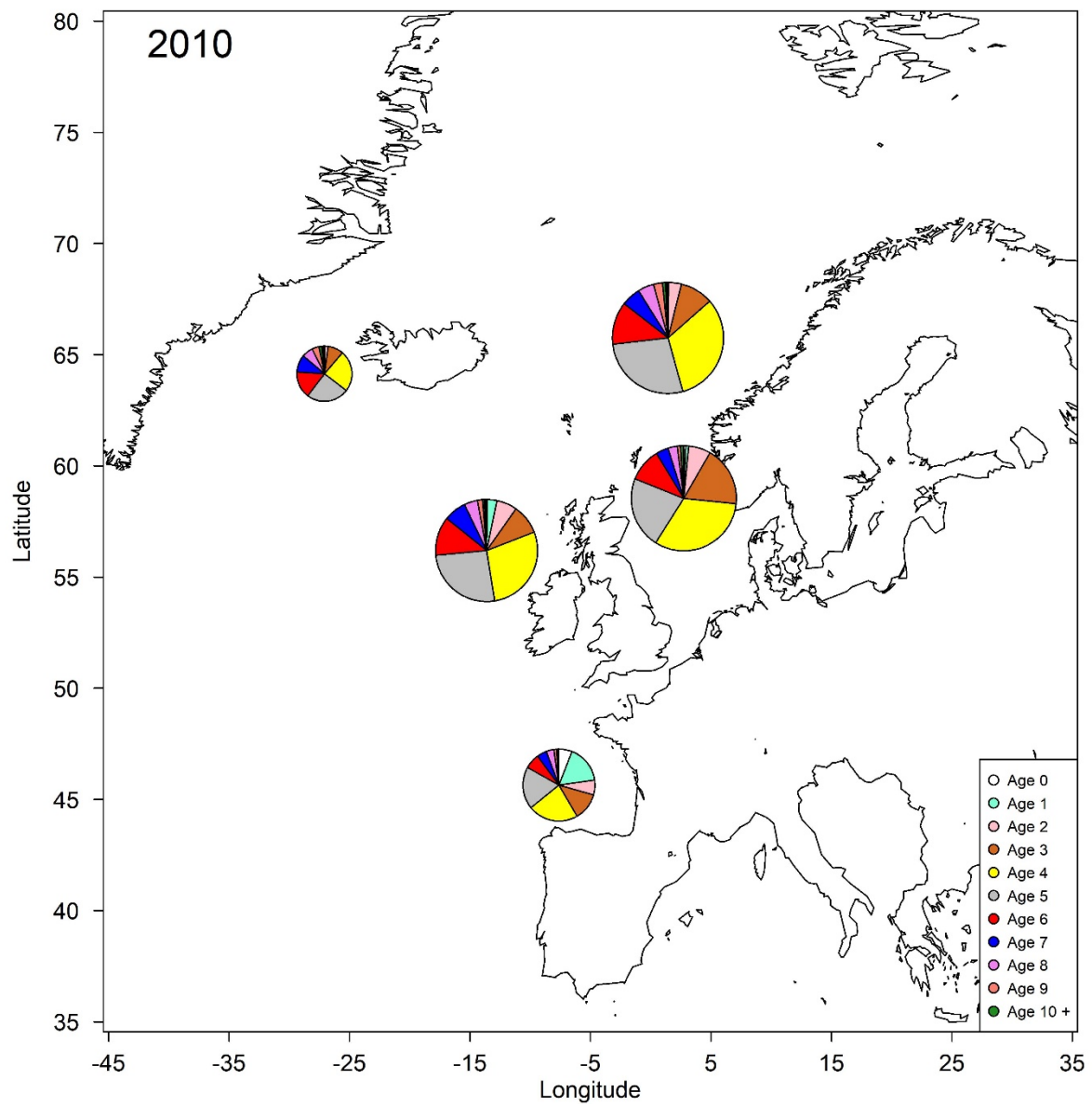


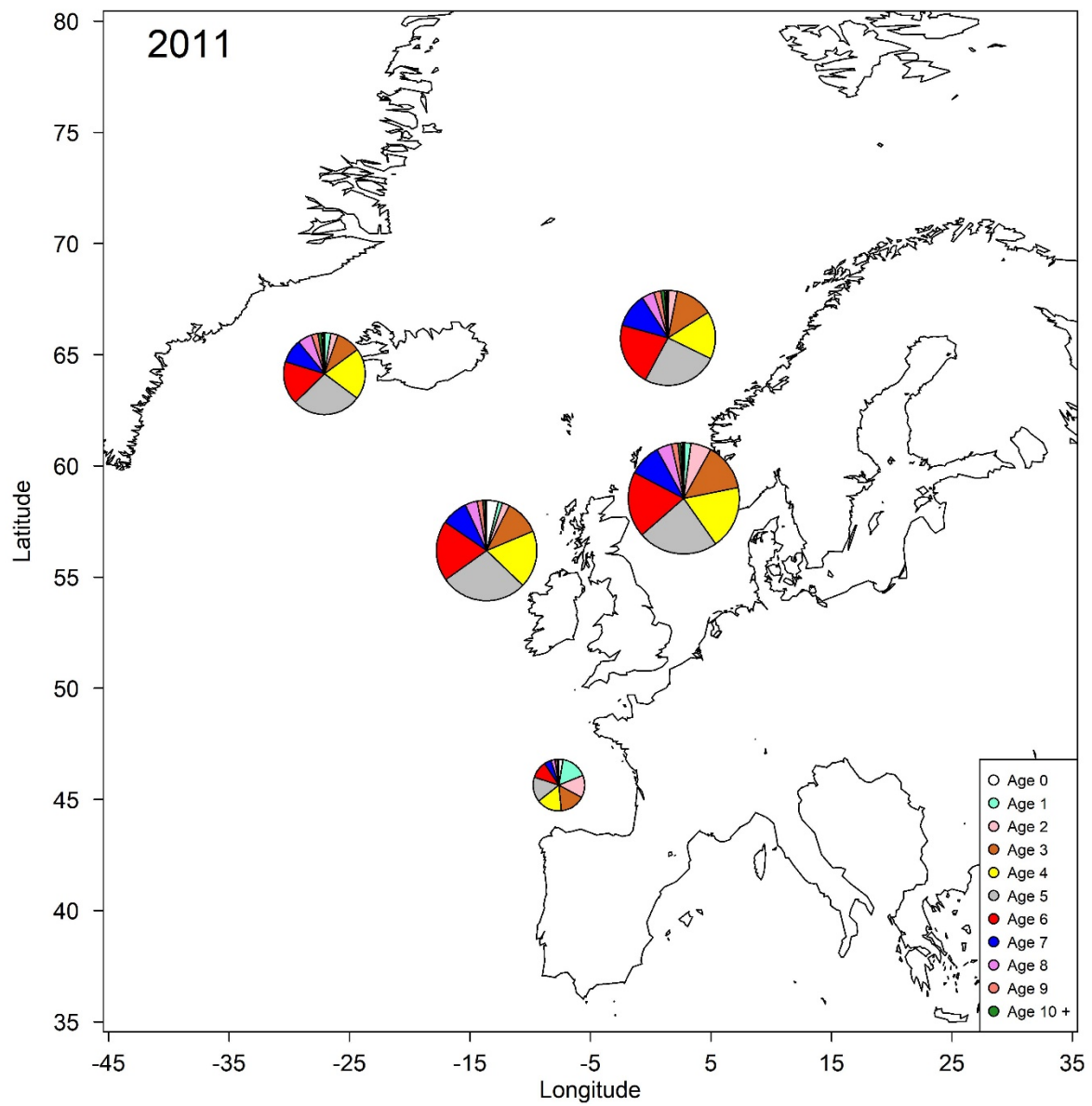


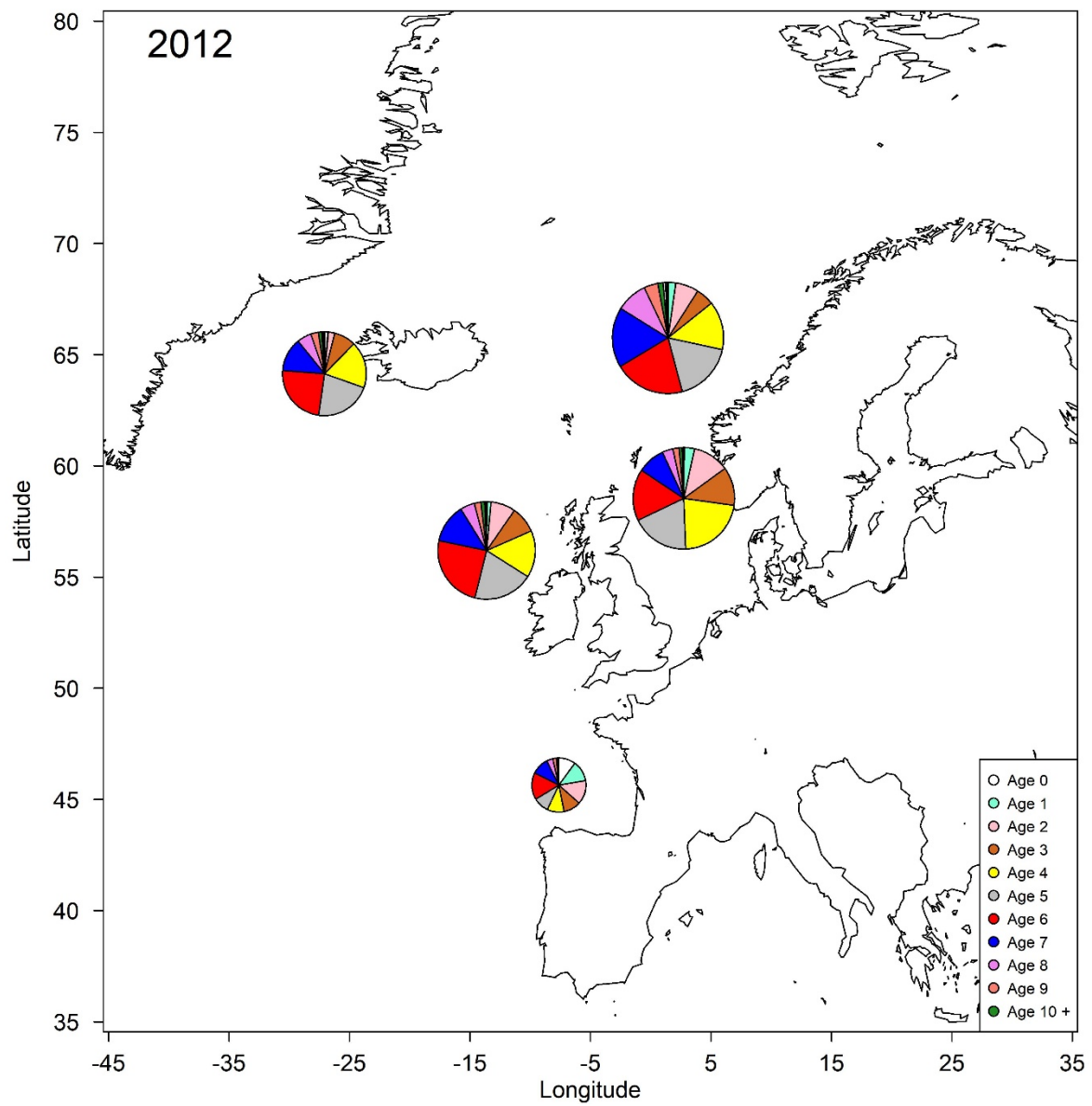


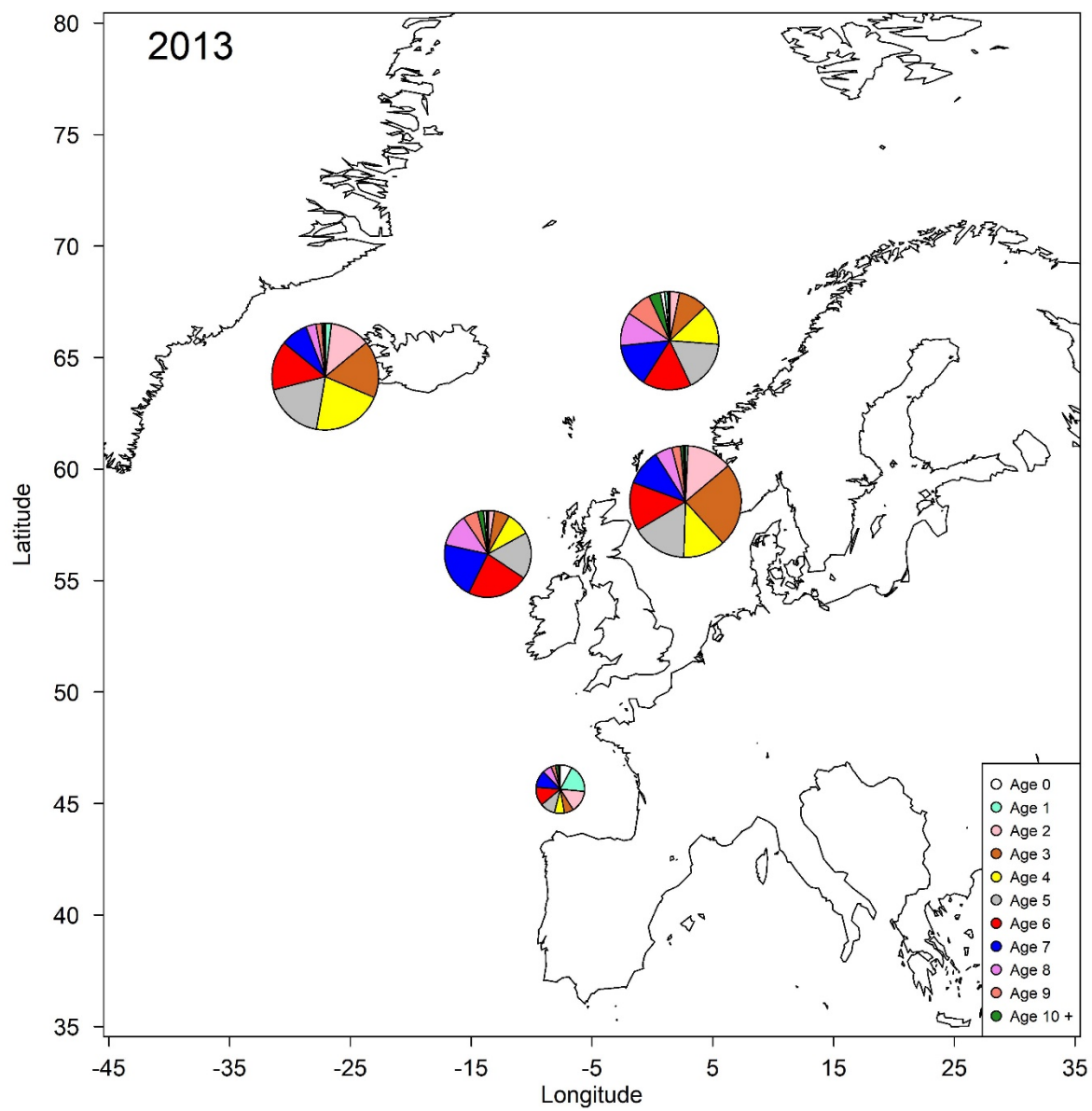


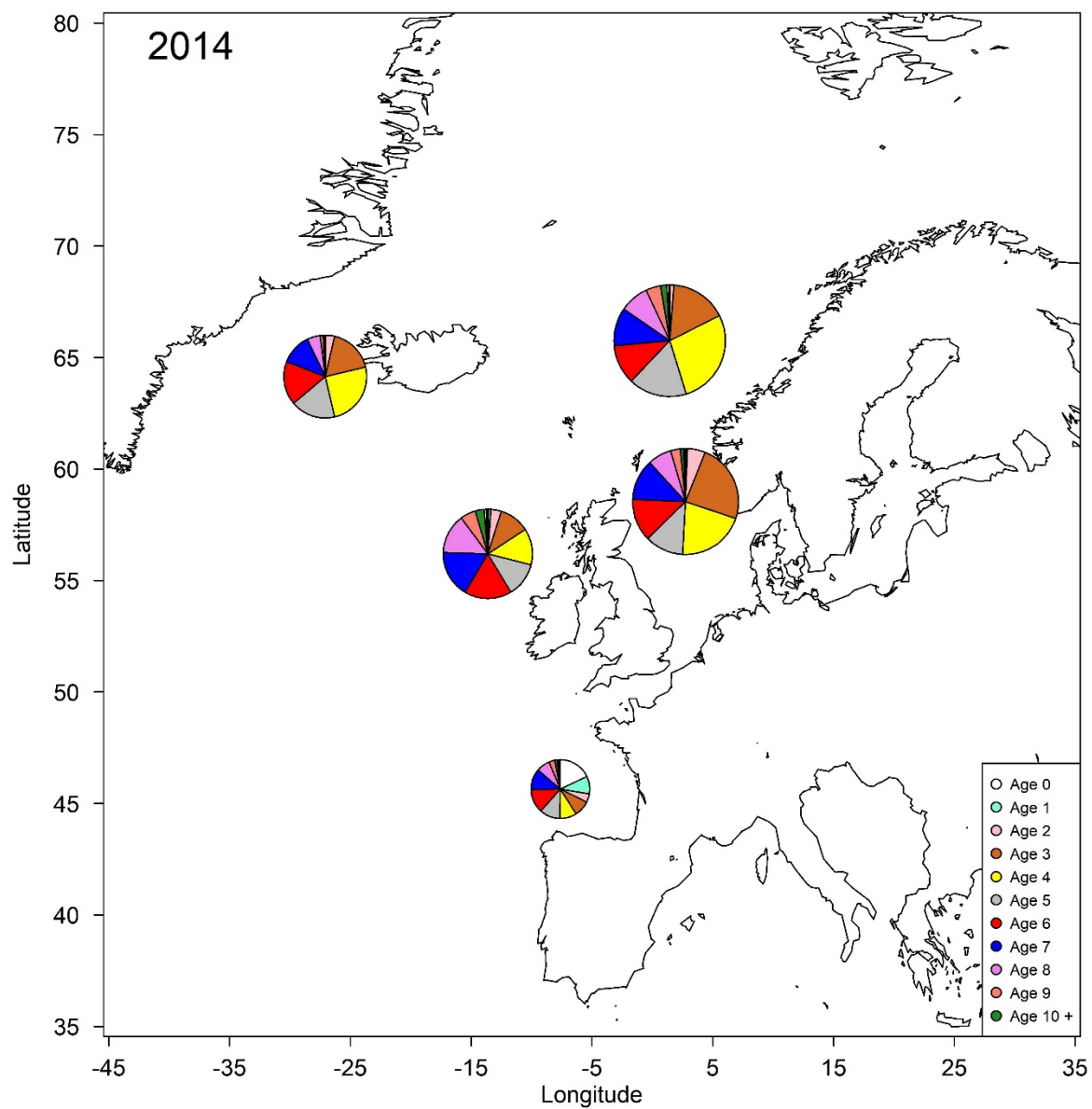


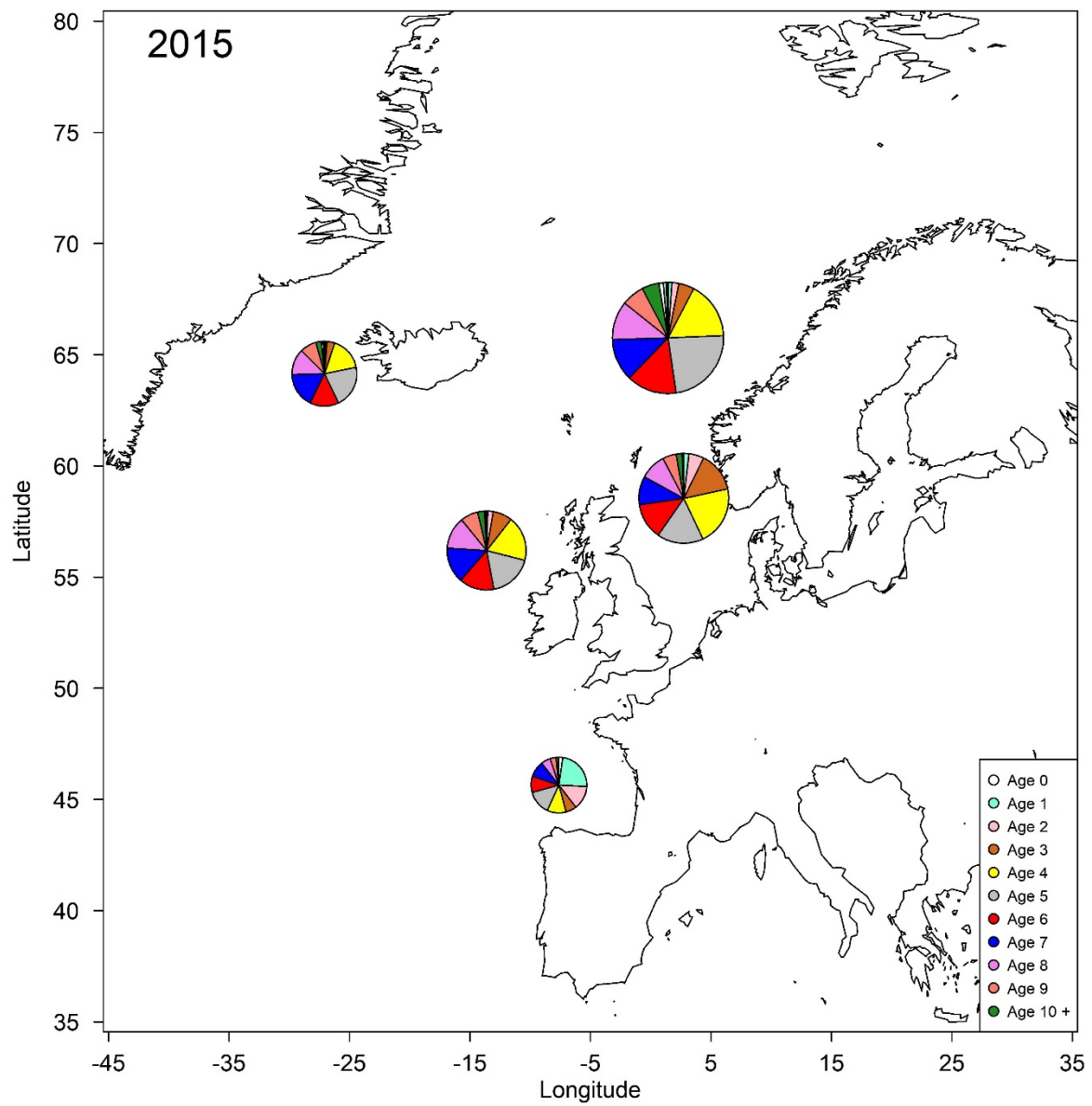


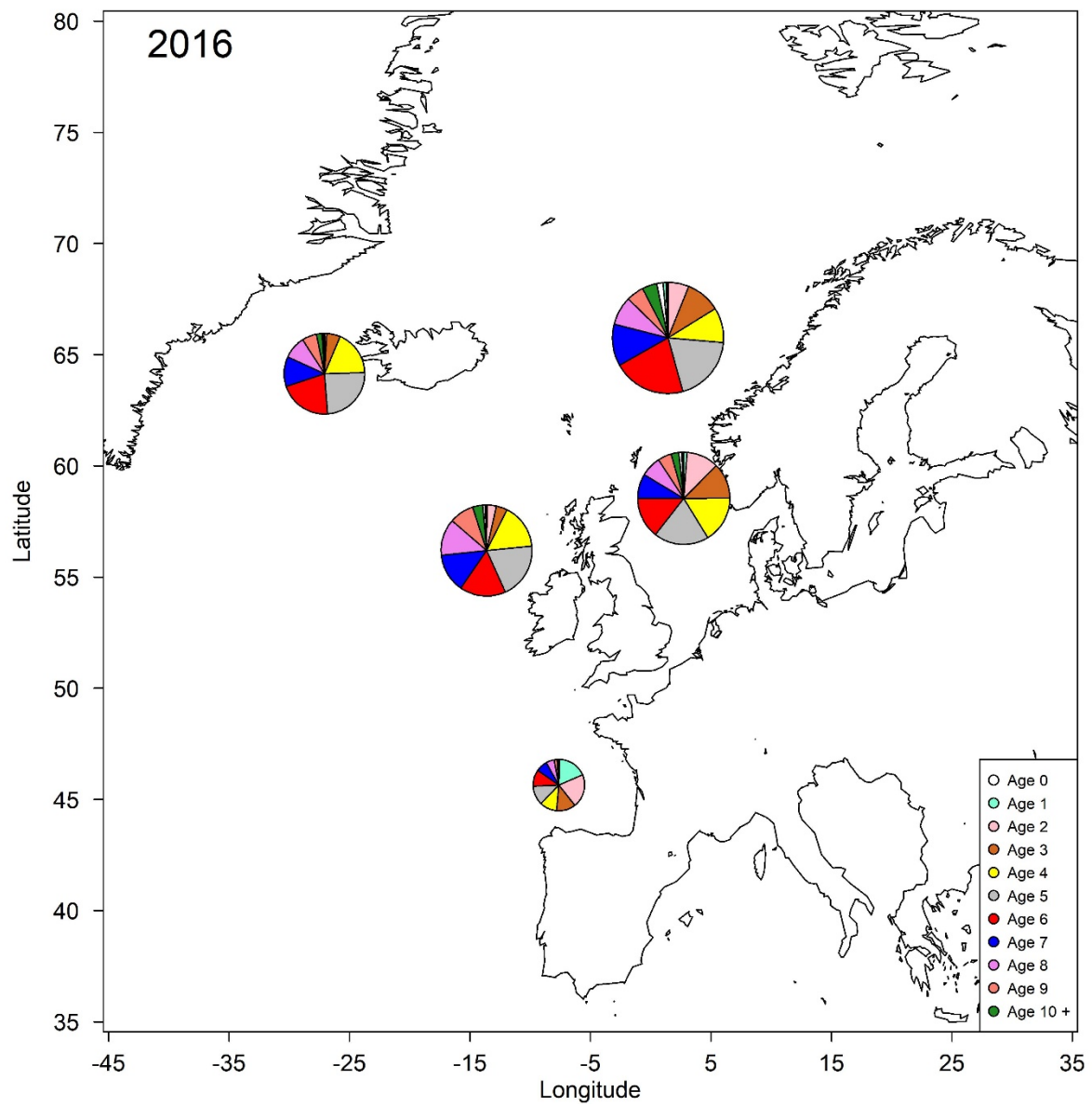












Pelagic AC Input Inter benchmark NEA mackerel

January 2019

Introduction

The Pelagic AC welcomes the opportunity to participate in the inter-benchmark on NEA mackerel organised by ICES on March 4-7th 2019 in IJmuiden.

In the following two sections, we present an overview of the information we have supplied to ICES over the last five years (2014-2018) for the stakeholder information section in the ICES mackerel advice, as well as an overview of the advice on mackerel that the PELAC has submitted to the European Commission and Member States during the same time period. Please consider this information as our contribution to the benchmark process.

Overview stakeholder information ICES mackerel advice (2014-2018)

- **2018**

Over the last ten years the pelagic industry has encountered large shoals of mackerel over the entire distribution area. Based upon these observations the industry believes the stock size has increased. This increase in the stock is not confined to one area or observed by only one fleet. The industry has noted signs of good recruitment to the fishery (above average numbers of 2–3 year old fish) in recent years, particularly in 2014 and 2015, and again in 2017 and at the start of 2018. Mackerel is also caught in substantial amounts outside of the directed mackerel fishery and in places where it has not been caught in recent years (e.g. during the herring fishery in the North Sea). The main spawning period in 2018 was found to be consistent with last year at the end of January. In 2018, the IESSNS survey was expanded into the North Sea using a Danish commercial vessel.

- **2017**

Over the last nine years the pelagic industry has encountered large shoals of mackerel over the entire distribution area which has expanded both south and north. Based upon these observations the industry believes the stock size has greatly increased. This increase in the stock is not confined to one area or observed by only one fleet. The industry has noted signs of good (above average) recruitment to the fishery (ages 2–3) in recent years, particularly in 2014 and 2015. The same signs were not evident during 2016. Mackerel is also caught in substantial amounts outside of the directed mackerel fishery and in places where it has not been caught in recent years (e.g. during the herring fishery in the North Sea). Danish fishers have reported catches of spawning mackerel in the sandeel fishery. In 2017, the main spawning migration (of the western component) began at the end of January, roughly three weeks later than normal.

- **2016**

Over the last eight years the pelagic industry has encountered large shoals of mackerel over the entire distribution area which has expanded both south and north. Based upon these observations the industry believes the stock size has greatly increased. This increase in the stock is not confined to one area or observed by only one fleet. The industry has noted signs of good recruitment (above average) in recent years, particularly in 2014 and 2015. Mackerel is also caught in substantial amounts outside of the directed mackerel fishery and in places where it has not been caught in recent years (e.g. during the herring fishery in the North Sea). Danish fishers have reported catches of spawning mackerel in the sandeel fishery.

- **2015**

Over the last seven years the pelagic industry has encountered large shoals of mackerel over the entire distribution area which has expanded both south and north. Based upon this observation the industry believes the stock size has greatly increased. This increase in the stock is not confined to one area or observed by one fleet. The industry also sees signs of good recruitment (above average) over the last number of years. Mackerel is also caught in substantial amounts outside of the directed mackerel fishery and in places where it has not been caught in recent years (e.g. during the herring fishery in the North Sea). Danish fishers have reported catches of spawning mackerel in the sandeel fishery.

- **2014**

Over the last six years the pelagic industry has encountered large shoals of mackerel over the entire distribution area which has expanded both south and north. Based upon this observation the industry believes the stock size has greatly increased. This increase in the stock is not confined to one area or one fleet. The industry also sees signs of good recruitment (above average) over the last number of years, particularly in 2009, 2010, 2011, and 2012. The widespread distribution of the stock over the entire area sometimes creates problems with unwanted bycatches of mackerel for fleets targeting demersal species. Stakeholders are actively seeking mechanisms that would improve data input into the survey indices and are involved in a number of pilot projects in this regard. Industry has scaled up its participation in the mackerel RFID tagging project: processing plants in Denmark, Iceland, Ireland, Faroe Islands, and Scotland are now equipped to read mackerel tags, in addition to the existing tag-reading facilities in Norway.

Overview PELAC recommendations NEA mackerel (2014-2018)

- **Northeast Atlantic Mackerel: PELAC 2018 advice**

The PELAC requests that the inclusion and influence of RFID tagging data used in the North East Atlantic mackerel assessment, as well as other data and considerations, be reviewed as a matter of urgency at an Inter-Benchmark meeting.

All the Industry stakeholders in the PELAC have voiced a strong opinion on and their lack of confidence in the assessment and advice provided for this stock. The PELAC requests that Coastal States ask ICES to re-evaluate the mackerel assessment and that once this re-evaluation is completed the TAC is set in accordance with the Three Parties Management Strategy as set out in the *Agreed Record of conclusions*

of fisheries consultations between Norway, the European Union and the Faroe Islands on the management of mackerel in the Northeast Atlantic for 2018, done in London, 11 October 2017.

The PELAC alerts the Commission and the Coastal States to the fact that the scientific advice for this stock has been considerably exceeded in recent years. If this practice continues, it is unlikely that the downward trajectory for the stock will be reversed. The PELAC supports the mackerel egg survey in 2019 and seeks the inclusion of the North Sea within the international egg survey scheduled for 2019.

- **Northeast Atlantic Mackerel: PELAC 2017 advice**

The PELAC recommends that the long-term management strategy developed by the EU, Norway and the Faroe Islands, to be revised on 10 -12 October 2017, will be followed, if it meets the criteria of the precautionary approach and article 2.2 of the CFP. If this is not the case, the PELAC recommends that the ICES MSY approach will be followed and that the TAC will be set at 550 948 tons in 2018.

Furthermore, the PELAC strongly recommends that technical issues impacting the stock assessment will be resolved as soon as possible. This includes the differences observed in the egg survey compared to the IESSNS survey, as well as problems with the RFI tagging program.

- **Northeast Atlantic mackerel: PELAC 2016 advice**

The Pelagic AC recommends that the management strategy proposed by the EU, Norway and the Faroe Islands will be submitted for evaluation to STECF as soon as possible. If STECF deems the management strategy precautionary and in line with the CFP before the Council meeting on 12 December 2016, then the Pelagic AC recommends that the management strategy will be followed and that the TAC in 2017 will be set accordingly at 1 020 996 tonnes. If the management strategy is not deemed precautionary by STECF, the Pelagic AC recommends that the ICES MSY approach will be followed and that the TAC will be set at 944 302 tonnes in 2017.

In addition the Pelagic AC recommends that a number of technical issues will be addressed at the upcoming benchmark including the contradictory signals provided by the egg survey versus the IESSNS survey, the collection of RFI tagging data and the issue of density dependent growth.

- **Northeast Atlantic mackerel: PELAC 2015 advice**

As a matter of urgency the Pelagic AC strongly encourages Coastal States to adopt a long-term management strategy for this stock. Furthermore, the Pelagic AC, through its industry partners, is fully committed to support the international egg survey in 2016 and to fill the void created by Norway's decision not to participate in the upcoming egg survey.

A number of technical issues must also be pursued. This includes the unstable assessment which ICES considers a source of major concern and which might be tightly related to the IESSNS survey. Density dependent growth is another issue that requires more attention by the scientific community. A workshop held in Bergen, Norway on 13-14 August 2015 addressing density dependent growth in mackerel and possible implications for management strategy evaluations is considered a step in the right direction. Another promising avenue leading to a better assessment of the Northeast Atlantic mackerel stock is radio-frequency identification (RFID) tagging currently carried out by Norway. The

EU pelagic industry participates in this project through tag reading equipment fitted in pelagic processing factories.

- **Northeast Atlantic mackerel: PELAC 2014 advice**

Given that the multiannual management plan agreed by the Coastal States in 2008 is only considered precautionary if the TAC constraint in clause 4 of the plan is not applied and given that the management plan does not maximize yield anymore, the Pelagic AC advises to follow MSY in 2015 and to revise the existing management plan as soon as possible. Following the MSY approach implies setting the TAC at 1.017 Mio tonnes. If a precautionary multiannual management plan has been agreed in 2016 the Pelagic AC will advise following this plan in 2016. In the absence of a revised management plan following MSY should be continued in 2016.

The Pelagic AC would also like to emphasize that the scientific information underlying the stock assessment and in particular the egg survey has to be improved significantly. It was confirmed by the ICES representative that the precision of the egg survey is very low while uncertainty is very high. Therefore the pelagic industry is currently considering carrying out a pre-survey in order to identify the actual start of the mackerel egg production period. Furthermore the Pelagic AC would like to emphasize the need for adequate control and enforcement and that the same measures applying to EU fishermen in European waters must also apply to third country fishermen in European waters.

Exploring trends in the age distributions within the tag-recapture data

Sindre Vatnehol and Aril Slotte (22.1.2019)

Purpose:

The main question raised in this WD, is if we are tagging the same population that are scanned, i.e. if numbers recaptured correspond to numbers tagged of the same age. The purpose is to describe age distributions of tagged, scanned and recaptured mackerel relative to each other over the time series 1986 to 2017. The document is not intended to give any conclusion of the data, but rather to add information to a further discussion.

Vocalization:

The data includes several information of the tag-recaptures. To avoid confusion, key variables (as used in the document) are defined underneath.

- Survey year is the year the fish has been caught and scanned.
- Release year is the year the fish was tagged.
- Year class is the year the fish was hatched.
- Age is the age of a year-class at a survey year (i.e. not the age at release year)
- r is the number of recapture fish with tags connected to a year-class and a survey year.
- N is the total number of scanned fish and is connected to a year-class and survey year.
- R is the number of fishes tagged and is connected to a year-class and a release year

Analysis

We have used data located at stockassessment.org. In the distribution's pattern, R , N , and r are compared with each other. For each survey year, an age distribution of N and r are made. The age distribution of R was made by using data where Survey Year – Release year = 1, i.e. when the tags have been at sea for 1 year.

Summary:

The age distribution was aggregated into three groups, survey year of 1986-1995 (steel-tag), 1996-2005(steel-tag) and 2012-2017 (RFID) (Figure 1). There was a clear trend that, in the earlier years of steel tag, the tagged and scanned population (year after tagging) had the same age structure, where young fish ages 3-4 predominated. In the last part of the steel tag time series, the tendency was towards tagging more old fish than what were scanned. This tendency developed even further for the RFID time series, where the scanned fish were more skewed towards younger fish than what is seen in the tagged population.

In Figure 2, the correlation between age distributions from r , N and R per survey year are shown. The correlation r vs N and R vs N show lowest correlation around survey year 2000 and increased correlation for older and newer survey year. The correlation for r vs R show a stabile correlation of 0.8

from 1985 to 2005. For 2011 to present date the correlation between r and R has increased to be almost 1.

The age distributions of N , R , and r are shown for each survey year in Figure 3. The tendencies showed for the periods are very much the same, yet more variable, in the annual comparisons, which emphasizes that there have been some changes in trends that may be worth discussing further in relation to the use of tagging data in the assessment

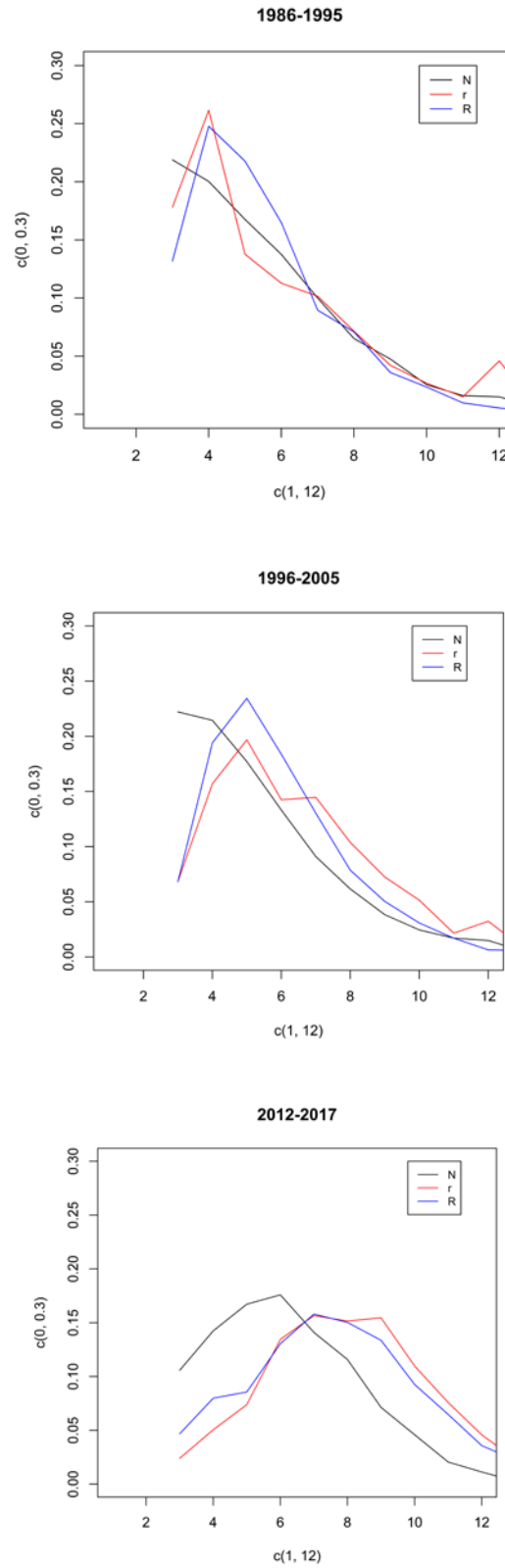


Figure 1. The relative age distribution of N , R and r , when aggregation the recapture data into three groups. The groups for survey year 1986-1995 is shown in the upper figure, the 1996-2005 is in the middle, and for the 2012-2017 group is in the lower figure. N is indicated with black line, R with blue line, and r with the read line. X-axis indicate the age, and y is the relative level of the distribution.

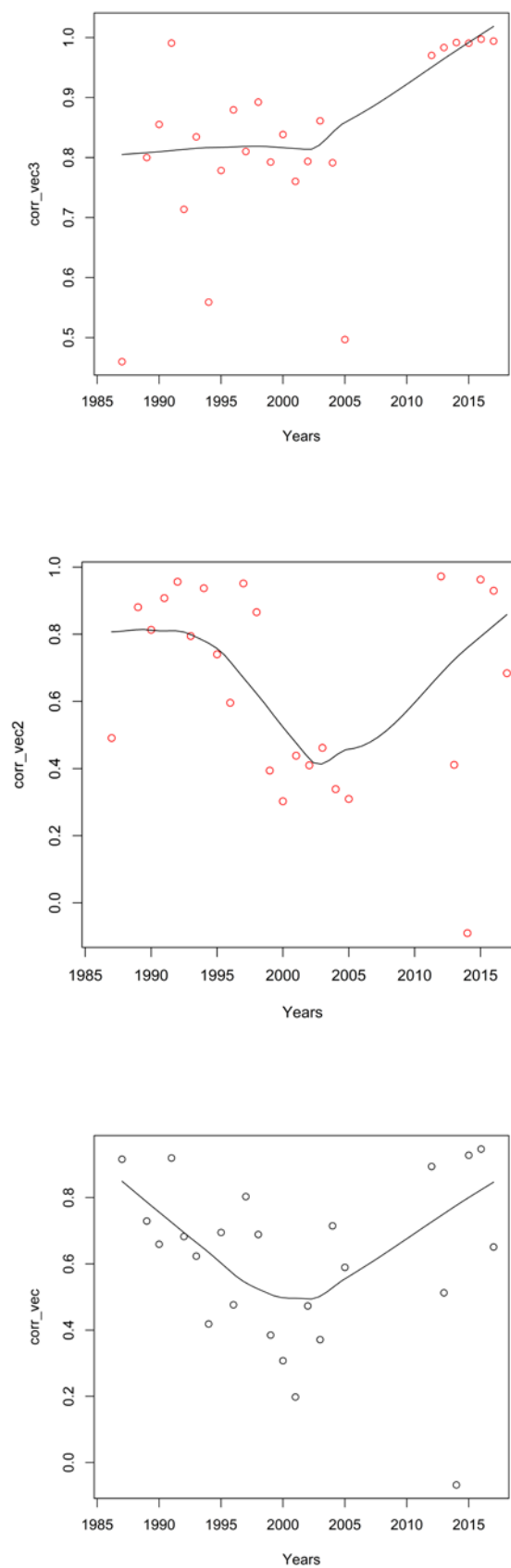


Figure 2. The correlation between age distribution between N , R , and r . In upper figure, correlation between R vs r is shown, R vs N is shown in the middle and r vs N is shown in the bottom figure. An overall trendline are shown as a black line.

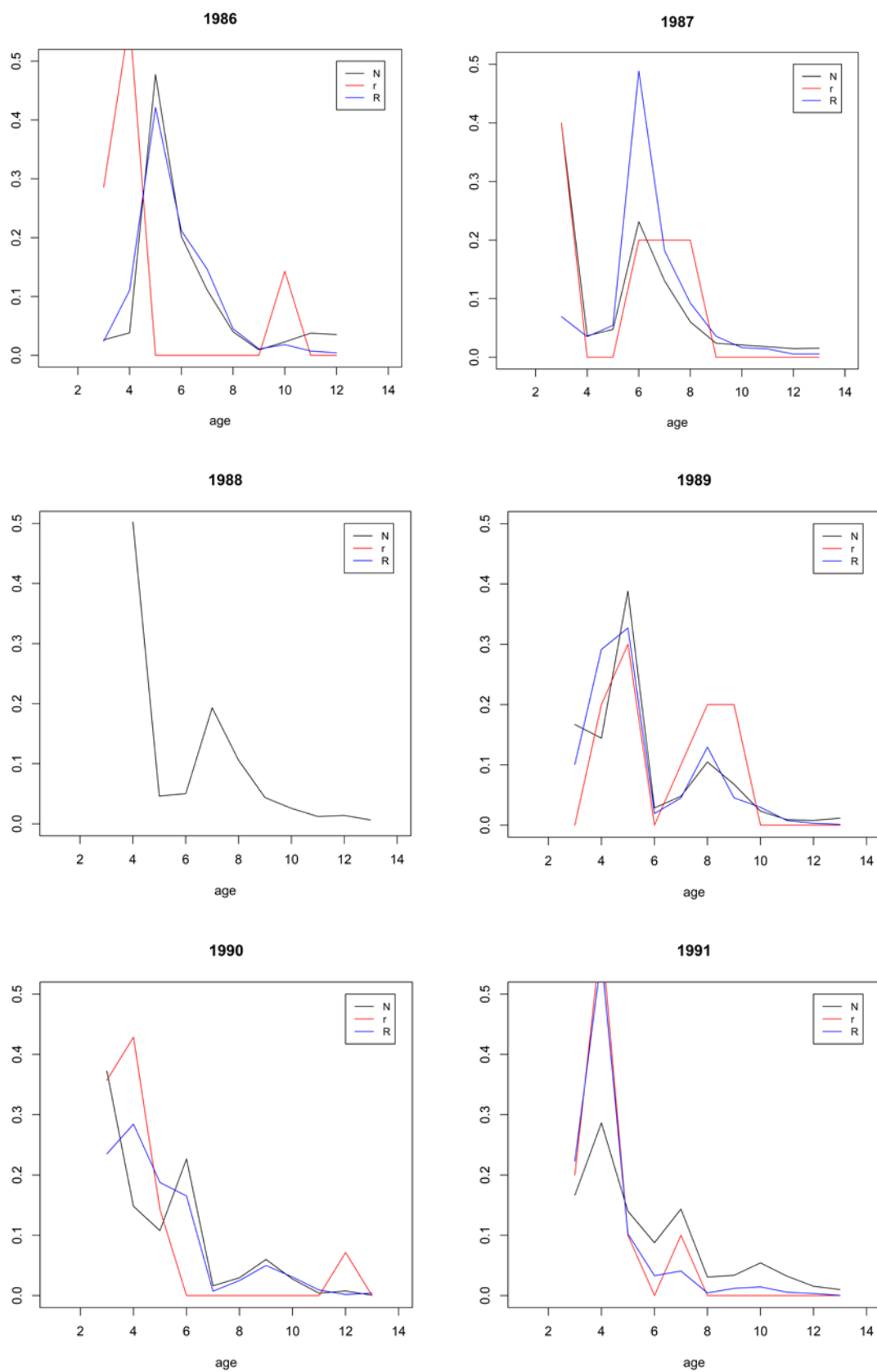


Figure 3. The relative age distribution of N (black line), R (blue line) and r (red line). The survey year is shown as the figure title.

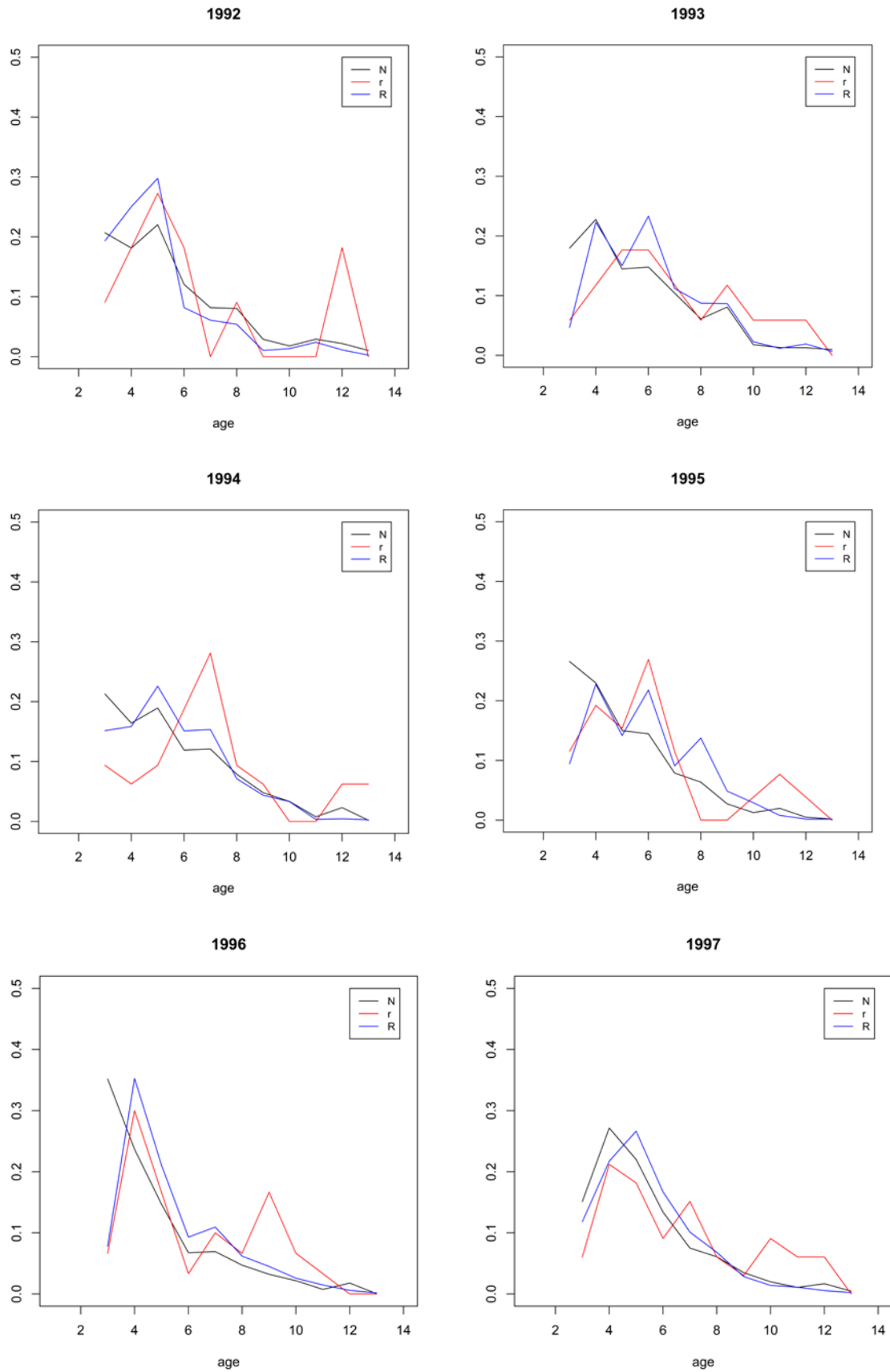


Figure 4. The relative age distribution of N (black line), R (blue line) and r (red line). The survey year is shown as the figure title.

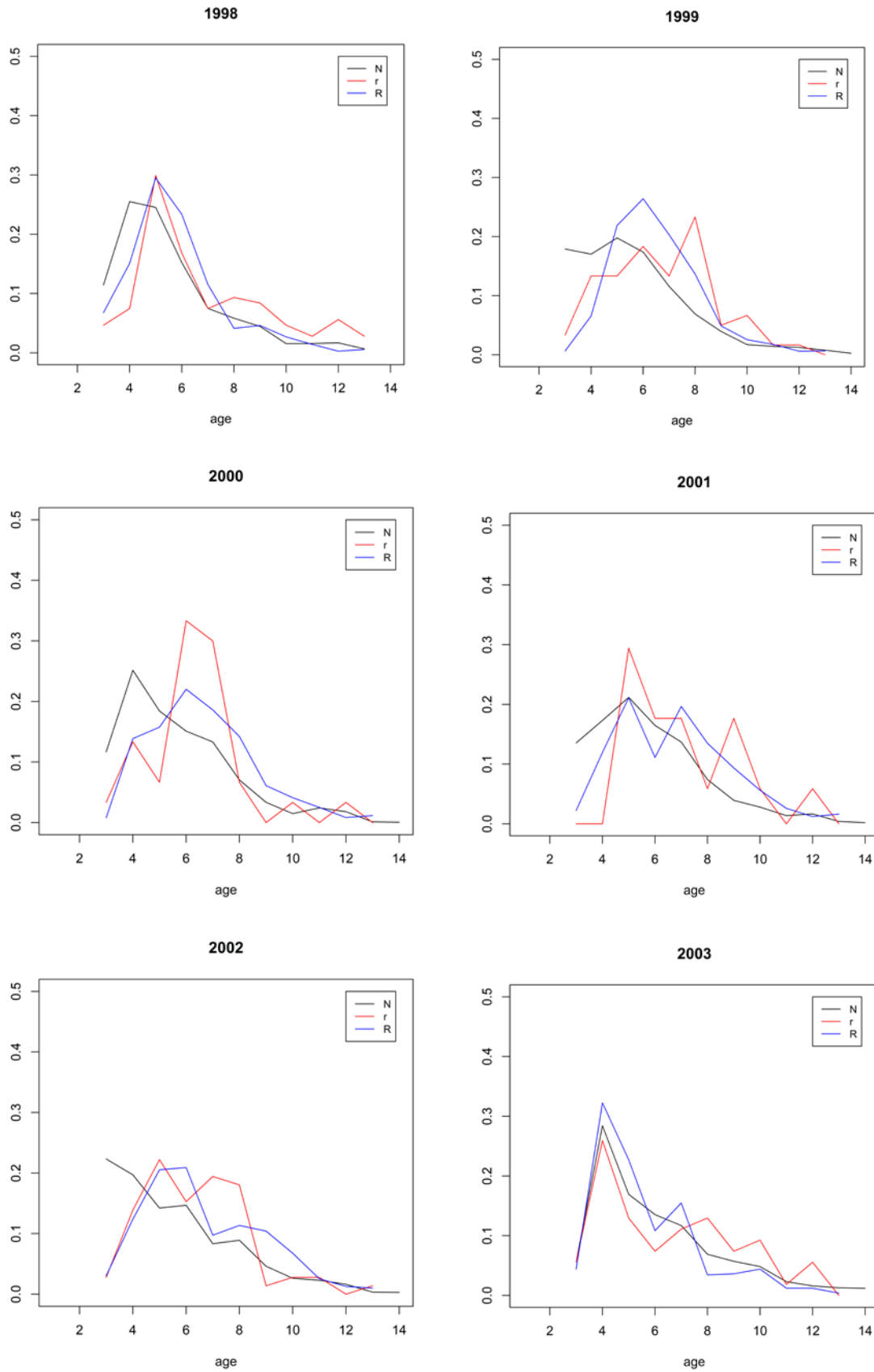


Figure 5. The relative age distribution of N (black line), R (blue line) and r (red line). The survey year is shown as the figure title.

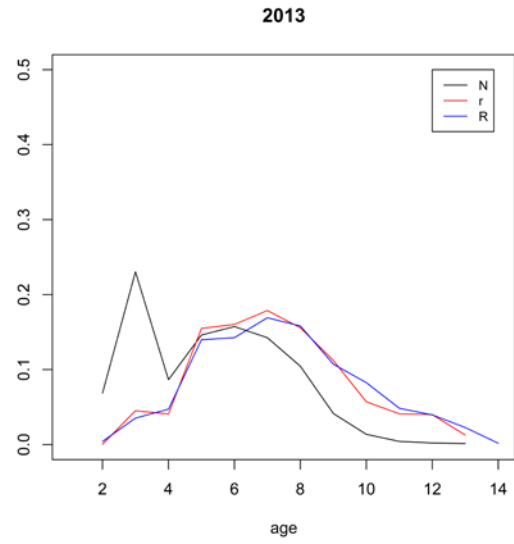
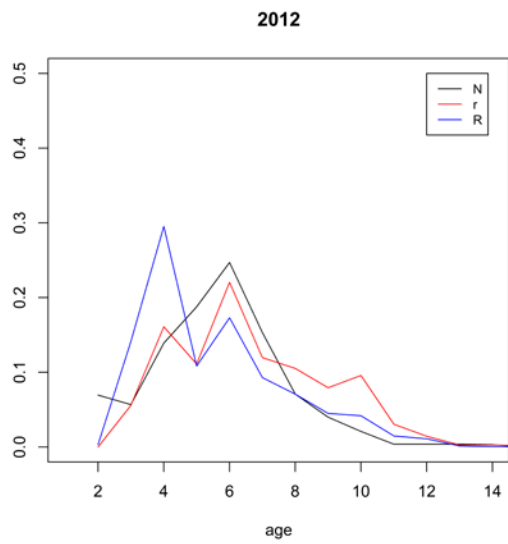
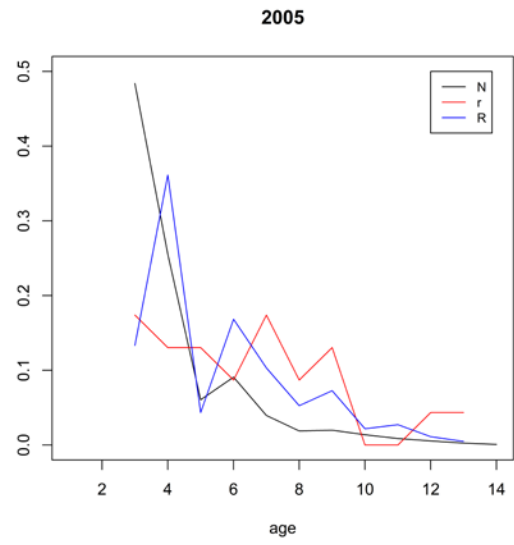
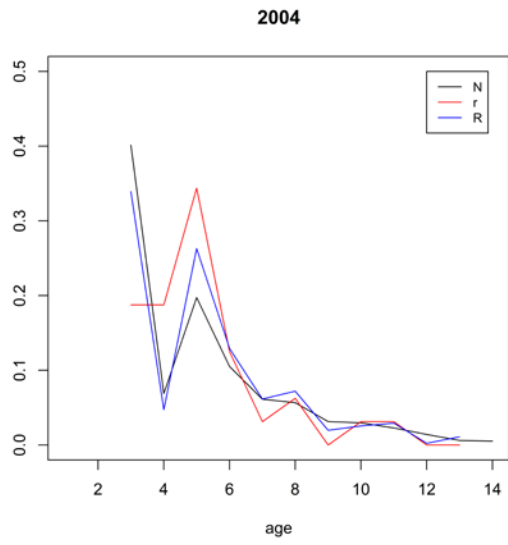


Figure 6. The relative age distribution of N (black line), R (blue line) and r (red line). The survey year is shown as the figure title.

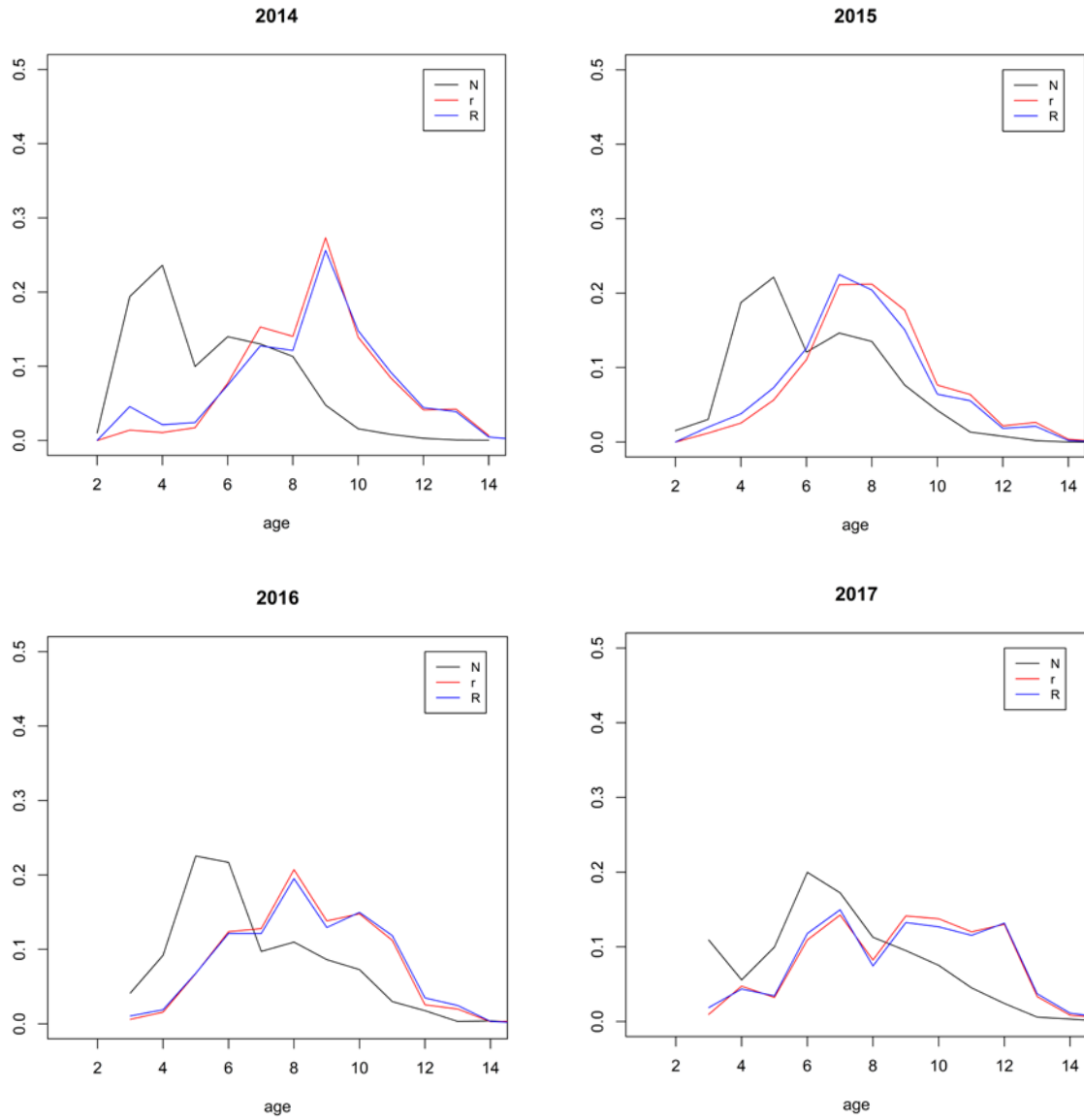


Figure 7. The relative age distribution of N (black line), R (blue line) and r (red line). The survey year is shown as the figure title.

Review of procedures and data in the Norwegian mackerel tag-recapture material

Dankert W. Skagen
Independent consultant.
dankert@dwsk.net
January 2019

1. Introduction

Norway has tagged mackerel with internal tags since 1969. Previously, the results were used as a supplement to analytic assessments, and occasionally to study migrations. Since the introduction of the SAM assessment method in 2013, tag recapture observations have been included in the objective function in the assessment. It has become apparent that the assessment is quite sensitive to these data, and that quite substantial changes appeared in the perceived stock abundance for each new year of recapture data (WD Claus m.fl). Also, as the tag recapture data were used as a relative measure of abundance, the scaling factor which represents survival in the tagging operation, became much smaller than one would expect.

A new benchmark is now being prepared for, which is a good opportunity to go critically through data and procedures regarding the tagging project. In this WD, we go through the various steps and examine potential sources of error or uncertainty. We then examine various usages of the data, and discuss each in the light of uncertainties and errors. Finally, we present some suggestions for how to make better use of the data.

The material analyzed here is the input file ('tag3.dat') to the SAM assessment in 2018. The file has numbers released, numbers recaptured and numbers screened (scanned) to discover the tags at recapture. The numbers are indexed by release year, recapture year and year class of the tagged fish.

(<https://www.stockassessment.org/index.php?dataFile=11&page=data>), supplemented with data from 2017. This file is only for the RFID tags, Steel tag data could be analyzed the same way, but that has so far not been completed.

1.1 Notation:

In the discussion here, we use the following notation:

yrel: Release year

yrec: recapture year

ycl: Year class.

yclbin: A collection of year classes.

$R(yrel,ycl)$: Number released

$r(yrel,yrec,ycl)$: Number recaptured.

$Nscan(yrec,ycl)$: Number scanned (screened) when tags are recaptured.

$N_explain$: Numbers as explained (assessed, estimated etc)

2. Discussion and examination of individual steps in the handling of tags

2.1. Release

2.1.1. Number released: That is recorded precisely. The steel tags were recorded on paper and transferred to a file as needed, In practice, the information about each individual tag was looked up if the tag was recovered, and recorded along with recovery data. The RFID tags are recorded directly to a database. The record for each tag includes the unique ID of the tag, date, location and length of the fish. No major problems suspected at this stage.

2.1.2. Length at release. Each tagged fish is length measured, in centimeters rounded downwards. There may be some inaccuracy in these measurements, perhaps an SD of 1 cm is reasonable.

2.1.3. Age of released. The fish cannot be aged at release, for obvious reasons. Instead, fish that are discarded for some reason, mostly because it is wounded in some way, are collected and aged, as well as length measured and weighed. From this material, an age length key is produced, normally one for each release cruise, i.e. one per year. At present (since ?????), the age length key is made length stratified (20 fish per cm), previously it was a random sample of about 500 - 800 individuals.

The age of tagged fish is obtained by applying the age-length key to the measure length of the fish. The result is an age distribution for each tag. With the steel tags, recaptured fish were often isolated and could be aged, and ALKs were only used as a fall-back. With the RFID tags, this is not possible.

The age distribution at each length is quite broad. Even though the number of fish in the ALK is large, there may be quite few fish in some length groups, and their ages may be spread, with holes in the distribution. Therefore, a valid objection is that following year classes may be misleading. The figure below gives one example.

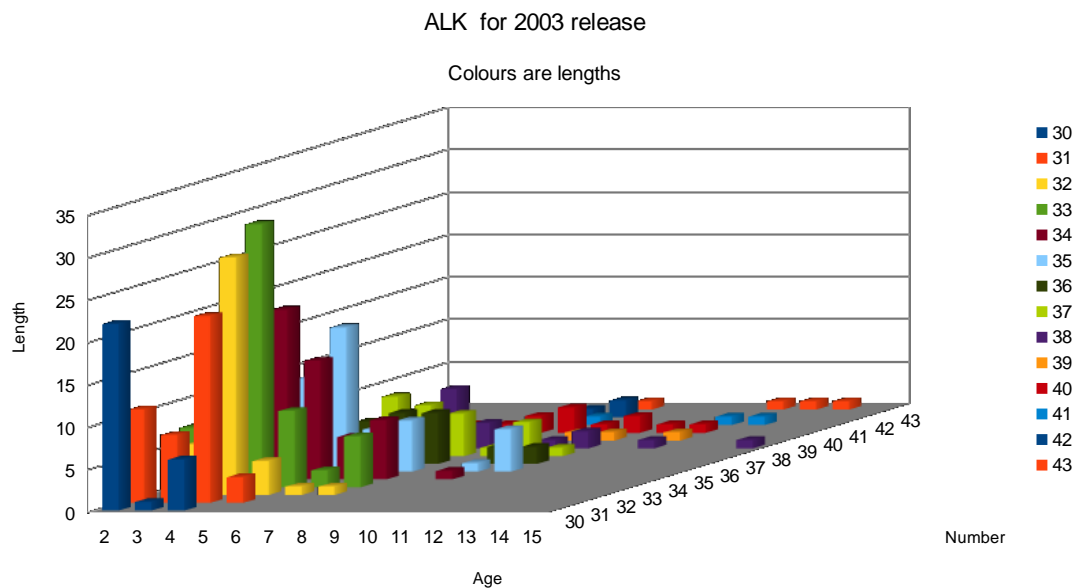


Figure 1. Example of an age-length key from a tag-release survey (2003).

2.1.4. Survival after release. Some of the newly tagged fish will die shortly after release. This may be due to damage or stress, and in some cases due to predation by birds. Mackerel is very vulnerable, so details in the handling may matter. The release mortality is difficult to measure, and experiments are few and incomplete.

In the assessment, the survival is a parameter. It just functions as a scaling, which may adapt to the fit to other data, right or wrong. A major draw-back is that the (inverse) recaptures become a relative measure of abundance, with the survival as an analogue to catchability. The assessment then will only regard trends, not absolute values.

The survival as estimated in the assessment appears to quite different before and after the time when release procedures were changed. Experiments comparing the recapture rate of tags that were released according to old and new procedures did not support a different survival with the old and new arrangements on deck. Rather than assuming that the survival has changed, one should suspect a problem with the assessment method or with other data that are included in the assessment.

2.2. Recapture:

2.2.1. Identity of recaptured tags. This is not a problem, each tag is uniquely identified.

2.2.2. Recapture efficiency. The efficiency in detecting a tag that goes through the detector is not completely free of problems, and in some cases factories have been excluded because of this. It is controlled by sending 'control fish' that have got tags inserted through the detection process. Some loss does occur, but only minor, and it can hardly explain a gross overestimate.

2.2.3. Volume screened. This is the recorded landing. In the past, catches may have been under-reported. Under-reporting used to be a problem both in international fisheries and probably in Norwegian, both as black landings and as slipping/high-grading at sea. Slipping/high-grading should not alter the amount actually screened, black landings might potentially have done so, if they have been screened but not reported correctly. Under-reported catches of course lead to underestimates of stock abundance in the assessment, which may be concealed as a higher estimated survival.

It also might have happened that some landings that were screened without finding any tags went unreported. This would also lead to too low values for Nscan. This problem has been thought of, but there is no solid evidence. With the RFID tags logistic, this is highly unlikely. However, it may be checked that all screened catches, and none else, are recorded.

2.2.4. Number screened. The number screened is derived from the catch by dividing by the mean weight. The mean weight is taken from samples, and is probably representative for the catch. In previous years, this may be more questionable, but it is hard to get firm enough evidence to suggest corrections.

In the last 3 years, about 20% of the catch has been screened, and in 2017 almost 700 tags were recaptured see Figure 2). The total number recaptured per year has increased gradually, presumably as more tags accumulate after more releases. A very rough calculation could be that about 30000 tags are released each year, which is also the yearly loss at equilibrium. Assuming $F = 0.3$ and $M=0.15$, one would expect that about 2/3 of the lost tags are lost in the fishery, which is 20000 tags. The equilibrium may not have been reached yet for the RFID tags, but after 6 years, about 90% of the fish at release time should be lost with these mortalities. When 20% of the catch is screened, one might expect something like 4000 tags recaptured each year. In practice, it will be less, because of a low survivor rate, but 700 is perhaps on the low side of what one should expect. On the other hand, the survivor rate of 10% estimated by the assessment seems very low, and if that is right the relative change by small alterations in survival may become a matter of concern.

Numbers screened and recaptures are shown in Figure 2 below, together with the total international catch at age from the assessment input (from WGWIDE report 2018)

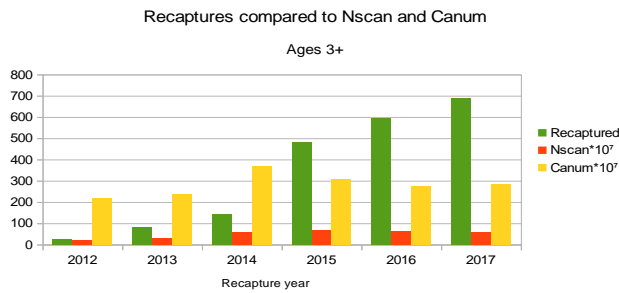


Figure 2. Numbers recaptured, screened and caught as summed up from the tag3.dat input file to the SAM assessment in 2018.

2.2.5. Age distribution of screened catch.

Figure 3 shows the age distribution of screened fish each year, compared with the age of the fish that was released in those years. In gross terms, both should reflect the age composition in the stock.

The age distribution of the recaptured tagged fish is rather similar to the released, as it should be. The age distribution in the screened fish is very different from the age distribution of releases, in particular in the more recent years. The younger year classes (2010 and younger) are much better represented in the screened material than in the tagged fish at release. In the releases before 2014, these year classes were hardly tagged, while they were well represented in the screened material later on. At best, this is a source of uncertainty, since the ratio between release and recapture has some very small release numbers.

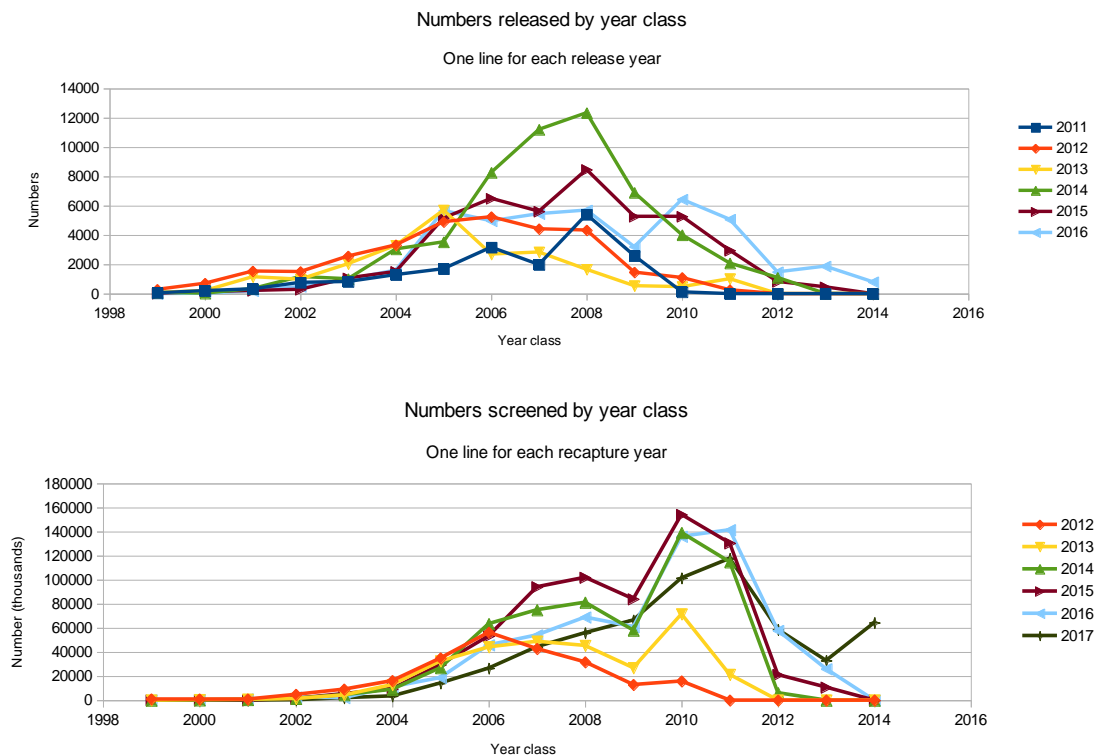


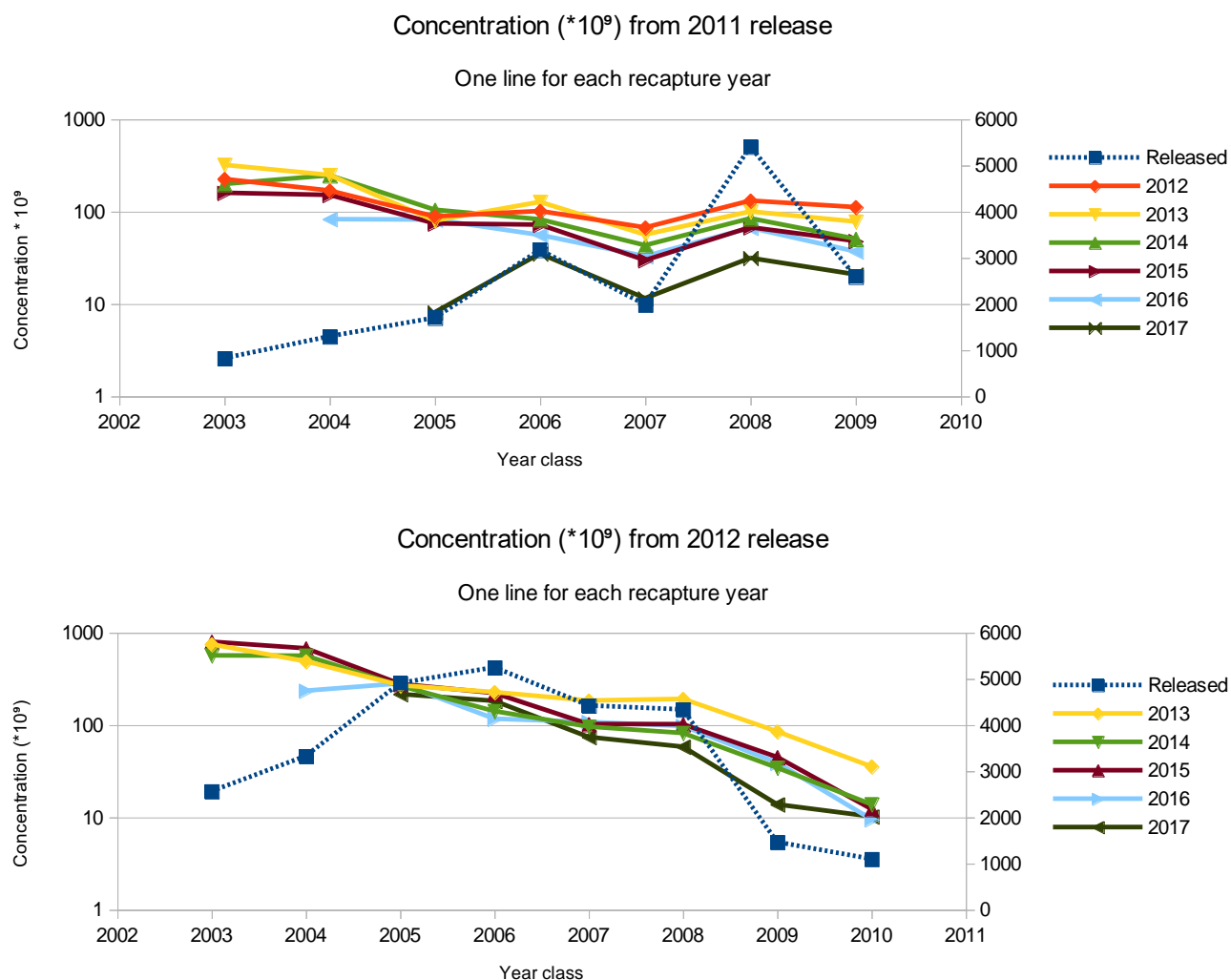
Figure 3. Year class distribution at release and in the screened catches.

2.3. Behavior in the sea between release and recapture - concentrations of tags.

A basic assumption is that the concentration of tags in a year class is the same in all recapture years, and as at release. That is equivalent to assuming that the mortality is the same for tagged and untagged fish.

The concentration can be measured at recapture, as the ratio between recaptured and screened tags within each year class. This was calculated for each release, for all subsequent recapture years. One should expect this to be constant, the number recaptured should be the same fraction of the number screened irrespective of how many fish were screened.

The collection of figures (Figure 4) show the concentrations by year class, one panel for each release year. The lines are the concentrations as seen in different recapture years. One would expect these lines to be on top of each other. That is only partially true, and the discrepancy in the figures is partly concealed by the log scale.



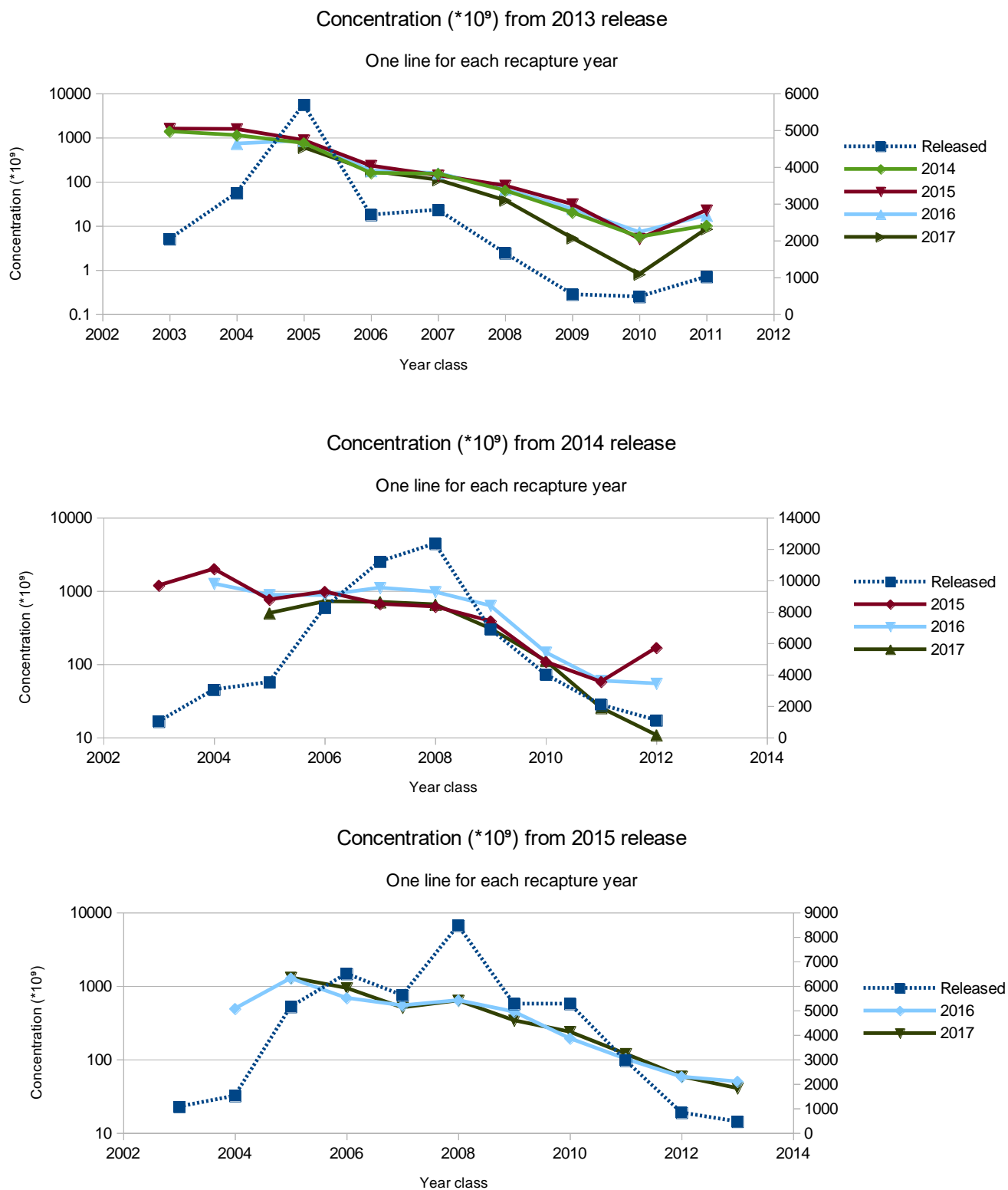


Figure 4. Observed concentrations of tags (number recaptured/ number screened) by year class. There is one panel for each release year. and one line for each recapture year. The number released in the release year is shown as a blue hatched line.

The deviations can be because the concentration in the sea is not constant, or that either the number of tags recaptured or the number of tags screened are wrong. One may suspect the latter in some of the cases here:

The oldest year classes get very high concentrations, and the concentration generally gets lower towards the younger year classes. One may suspect that this may be related to the age distribution of the screening, which is skewed towards young fish compared to the age distribution of the releases.

The constant concentration assumption itself is violated if:

- The tags are released in a sub-population, and that sub-population is represented to a variable extent in the screened catches. A sub-population here is not necessarily a sub-stock in a meta-population, but perhaps just a tendency for fish spawning in different places to get fished in different fisheries.
- Tagged fish has a different mortality also long after the tagging.
- Different tagging survival for different ages will violate the assumption of a common survival fraction in the assessment, but not the assumption of constant concentration by year class.

Different tagging survival for different ages (for example due to a rising selection with age in the fishery) will violate the assumption of a common survival fraction in the assessment, but not the assumption of constant concentration by year class.

We have only limited insight in these obstacles. In particular, we have very limited insight in the mortality of tagged vs. un-tagged fish, we only know that tags can be found after many years and the fish does not appear to be affected by them. That of course, does not exclude that some get affected, we probably will not see them.

The subpopulation structure has caught some attention recently (For example: Jansen T, Gislason H (2013). Population Structure of Atlantic Mackerel (*Scomber scombrus*). PLoS ONE 8(5): e64744. doi:10.1371/journal.pone.0064744), which may matter as migration routes are changing. It may lead to misleading results if the scanned catches are dominated by a different subpopulation than the one that is tagged. However, but the numbers scanned and the age distributions of tagged and scanned fish introduces an uncertainty that perhaps is more disturbing.

2.4. Recaptures in the release year.

The standard practice has been to ignore recaptures of tags released in the same year. The argument has been that these tags will not be properly mixed in the population. In addition, they will only be fished after the tagging in May, so the numbers screened to find them is lower than the number screen for the year. Figure 5 below shows the ratio between Number recaptured of the release one year before and Number recaptured in the release year. The ratios are presented by recapture year and year class.

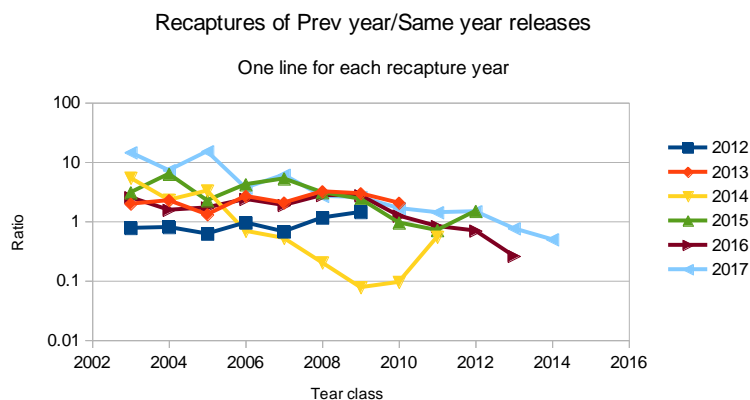


Figure 5. The ratio : Number recaptured of the release one year before / Number recaptured in the release year, presented by recapture year and year class.

Ideally, the ratio should be close to one. This was the case with recaptures in 2012, but not in later years, where the number recaptured in the release year was markedly lower, except in a few cases. The 2014 recaptures were exceptional, and in the 2015 - 2017 recaptures..there is a trend towards lower ratios (release year tags more in line with or even more frequent than previous years tags) in the youngest year classes. This may have some relevance to discussions o effects of migrations, and also be a warning that recaptures of fish released the same year may be misleading if used uncritically.

3. Measures derived from the tag-recapture data.

The measures that can be derived from this material relate to stock numbers and mortalities. Our practice has been to analyze the data as year-class dis-aggregated, to follow the developments over time. We can regard a year class as a closed cohort, in the sense that the only cause of change in abundance is mortality.

3.1. Stock numbers

The stock numbers in a year class at release time is estimated as

$$N = R/r*Nscan$$

In the assessment, the expected recaptures by release and recapture and year class is modeled as a function of modeled stock numbers:

$Expected_r = R*Nscan/Nassm$. These are compared with the observed values in the likelihood function, assuming a negative binomial distribution.

The numbers estimated derived from the tagging material are tabulated in Table 3 at the end of the document. A summary (estimated stock numbers in each release year according to the recaptures in each recapture year, summed over all year classes that were at least 3 years old at release time) are shown in Figure 6. These numbers should be the same for all recapture years, i.e. the lines should be horizontal. More often, they tend to rise. Many of the individual numbers are very large, several of them well over 10^{11} . This may be caused by small recapture numbers, large number screened or both.

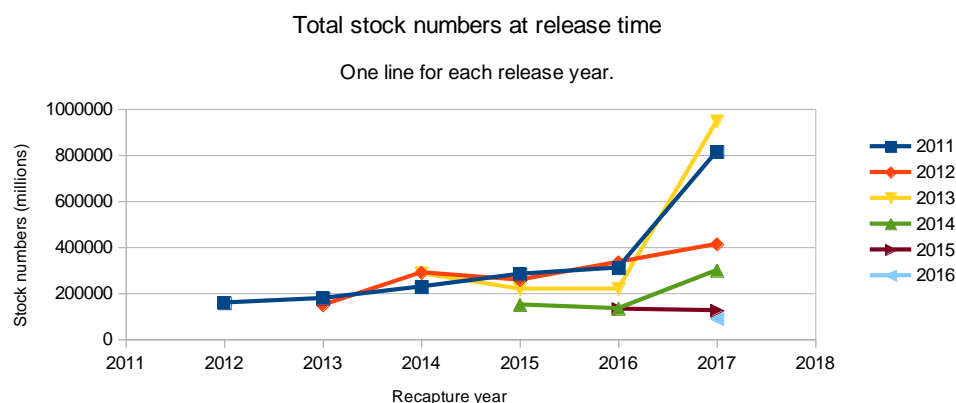


Figure 6. Summary of stock numbers at release time, as estimated at recapture time. The sum over all year classes corresponding to age 3+ at release time is shown.

Figure 7 below show that most of the extreme stock number estimates are based on one recaptured tag or less. Most large numbers are associated with small numbers recaptured, but some with quite large numbers screened. Therefore, the wide spread of ages of fish and to a lesser extent large numbers screened, seem to contribute to stock numbers that do not seem convincing.

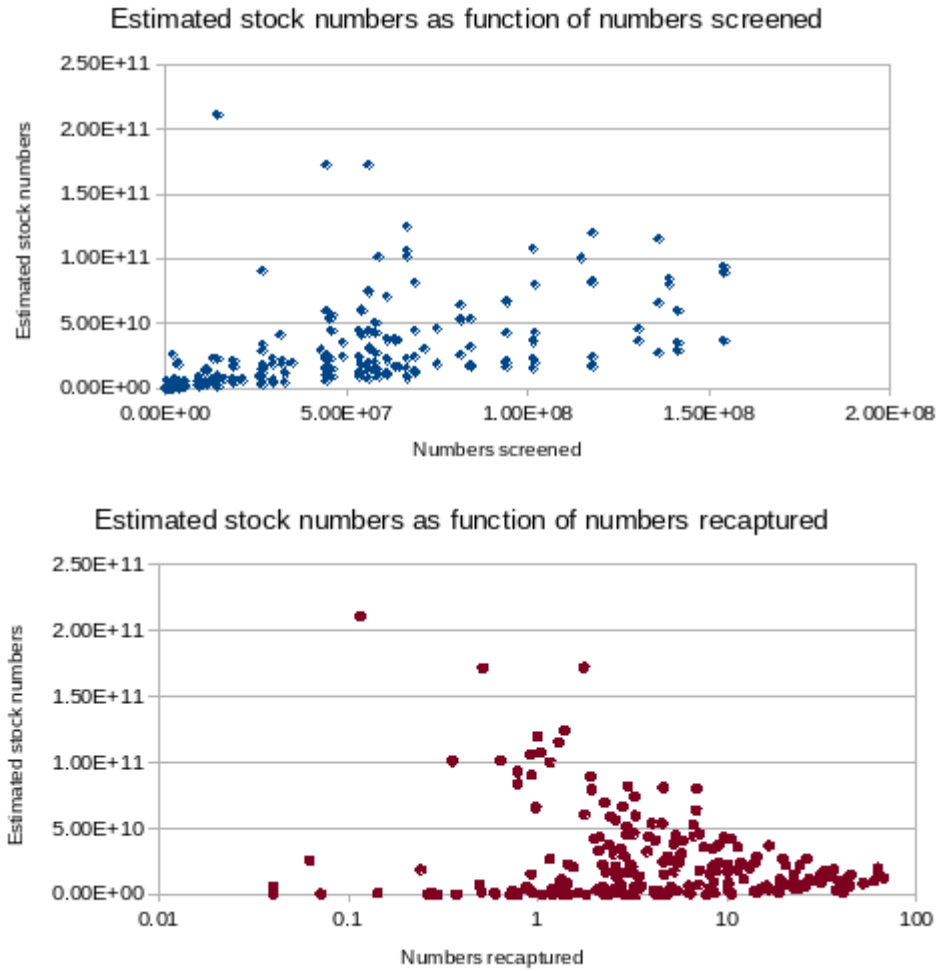


Figure 7. The relation between estimated stock numbers (ordinate axis) and numbers recaptured (upper panel) and numbers screened (lower panel.) Note the log scale on some of the axes. All data except one extreme outlier.

3.2. Z-estimates.

Two approaches have been examined.

3.2.1 Using the loss of tags between two consecutive releases, as seen in subsequent recaptures.

$$Zest(y1, y2) = \log(\Sigma_{yrec} r(y2, yrec, ycl) / (\Sigma_{yrec} r(y1, yrec, ycl) * R(y1, ycl)) / R(y2, ycl))$$

This is shown for $y2=y1+1$ in Table 1 and Figure 8,

Under the assumption that the concentration of tags is the same in all years after a release, the ratio between the number of tags surviving release 1 and the number released in release 2 should be independent of the recapture year. Therefore, the recapture tags are summed over all years after $y2+1$. The first year after the last year is excluded, to make the mixing of tags more complete. This measure was used with the steel tags for many years in the past, as a confirmation of the mortality estimate in the analytic assessment.

Table 1. Mortality estimates by comparing recaptures form subsequent releases.

$$Zest(y1, y2) = \log(\Sigma_{yrec} r(y2, yrec, ycl) / (\Sigma_{yrec} r(y1, yrec, ycl) * R(y1, ycl)) / R(y2, ycl))$$

First release year	2011	2012	2013	2014	2015	
Second release year	2012	2013	2014	2015	2016	
2005 ycl	2.31452	1.23715	-0.56521	0.96316	0.88699	
2006 ycl	1.31701	-0.54263	2.61306	-0.28865	0.14656	
2007 ycl	1.95024	-0.06955	3.13463	-1.25087	0.62192	
2008 ycl	0.14992	-1.25435	4.40695	-0.63931	-0.16801	
2009 ycl	-0.68101	-1.48380	5.57481	-0.43442	-0.36087	
2010 ycl	*	-1.68958	5.38945	0.75968	0.82005	
2011 ycl	*	*	1.80640	1.26400	1.43748	
2012 ycl	*	*	*	0.30264	1.66265	
2013 ycl	*	*	*	*	3.80825	

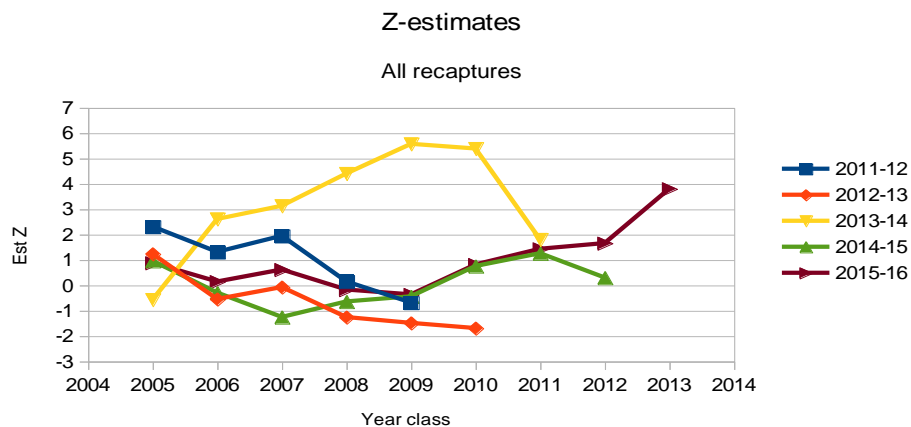


Figure 8. Estimates of Z as tabulated in Table 1.

There are some irregularities and inconsistencies that look quite alarming. In particular, the contrast between the 2013 and 2014 releases is very strange, but there are strange signals in the other mortalities as well and apparently, the noise is quite prominent. This measure is independent of numbers screened, but is quite sensitive to noise in the recapture numbers at age, which we have noted above, and the problem is amplified by relying on the ratio between small, noisy numbers.

3.2.2. Disappearance of tags between release and recapture.

Hoping to get a more robust measure of mortality, the gradual disappearance of tags between release and recapture, seen from the recapture year, was computed. The logic is that as years pass after a release, tags disappear because the fish dies. That reduces the numbers recaptured. They will of course also depend on the number released, and on the number screened. We can compare recaptures of tags from different releases, normalized by dividing with the number released, and we can do that in one recapture year, where the number screened will be the same all the way. The log of the rate of decline is the mortality between the points, and the slope of the line for a year class is the mean mortality in the period.

Hence, for each **recapture** year, we consider the ratio between recapture and release by year class:
 $\text{recapture}(\text{yrelease}, \text{yrecapture}, \text{ycl}) / \text{release}(\text{yrelease}, \text{ycl})$.

Basically, this is the same measure as above (fraction of the release that is recovered), but rather than presenting the ratios between two year classes directly, we just present the trend in the fractions as estimate of the average mortality.

Below is one example of the results (ratios for releases recaptured in 2015)

Table 2. Log (number recaptured in 2015/number released in years indicated)

Recaptures in	2015			
Ycl Release	2011	2012	2013	2014
2005	-6.649291	-6.388470	-5.398102	-5.053588
2006	-6.713177	-6.096016	-5.393785	-5.066084
2007	-6.565759	-6.130076	-5.381049	-5.190177
2008	-6.665117	-6.032453	-5.297377	-5.290934
2009	-6.463820	-5.961967	-5.332004	-5.356804
2010	*	-6.362585	-6.405748	-5.486320
2011	*	*	-5.851884	-5.623909
2012	*	*	*	-5.740505
2013	*	*	*	*

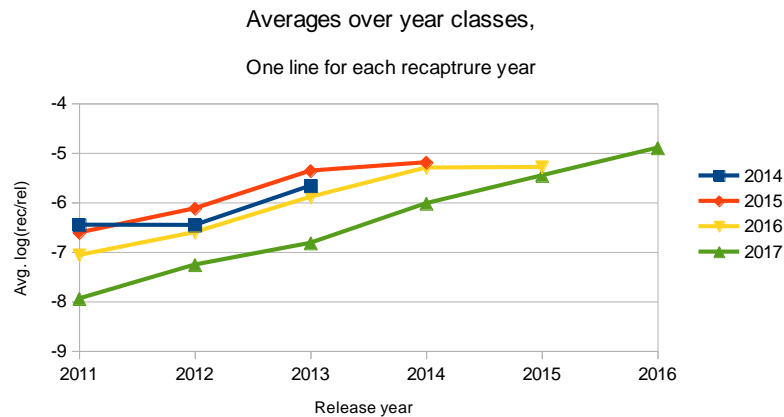


Figure 9. Disappearance of tags. Log (number recaptured in the year indicated/number released in years on the abscissa). Average over year classes. The slope of the curves is an estimate of the total mortality.

These results may be further smoothed by merging or taking the average of several year classes. This is shown in Figure 9. Here, averages are taken over year classes that are aged 5-10 at recapture, and at least 3 years old at release.

The text table below shows the slopes of these curves. The second (gray) column shows the slopes when points have been removed that make the curve flatter (2015 release in 2016 recapture, 2014 release in 2015 recapture and 2011 release in 2014 recapture). There is no good reason to regard these as outliers, but excluding them gives an indication of the sensitivity of even such aggregated results to single data points.

	Mean Z	Excl 'flat'
2014	0.39	0.8
2015	0.51	0.64
2016	0.49	0.6
2017	0.64	0.64

4. Merging year classes

In this study, we explore briefly 'cohorts', that we call year class bins, that consist of several year classes. Thwe purpose is to avoid some of the noise caused by inaccurate age attached to each tag. A similar approach using length bins has been discussed, but time did not permit exploring that.

The year class bins are closed cohorts, but they are not disjoint as year classes are. Such year class bins will only be valid in years where all its year classes are represented in the material. Here, we have used ages 3-12. One can derive stock numbers by year class at release time, and mortality between releases or between release and recapture.

4.1. Stock numbers.

Stock numbers were estimated as described above (Section 3.1), but for a single year class bin instead of estimating each year class separately as in Table 3. One example is shown in Figure 10 (upper panels), which is the bin with 2005-2009 year classes at various releases. These numbers should ideally be similar for all recaptures, i.e. the lines should be straight and horizontal, but at different levels, which does not seem to be the case. The averages over recapture years make a fairly smooth exponential decline, however, with a slope corresponding to $Z = 0.22$.

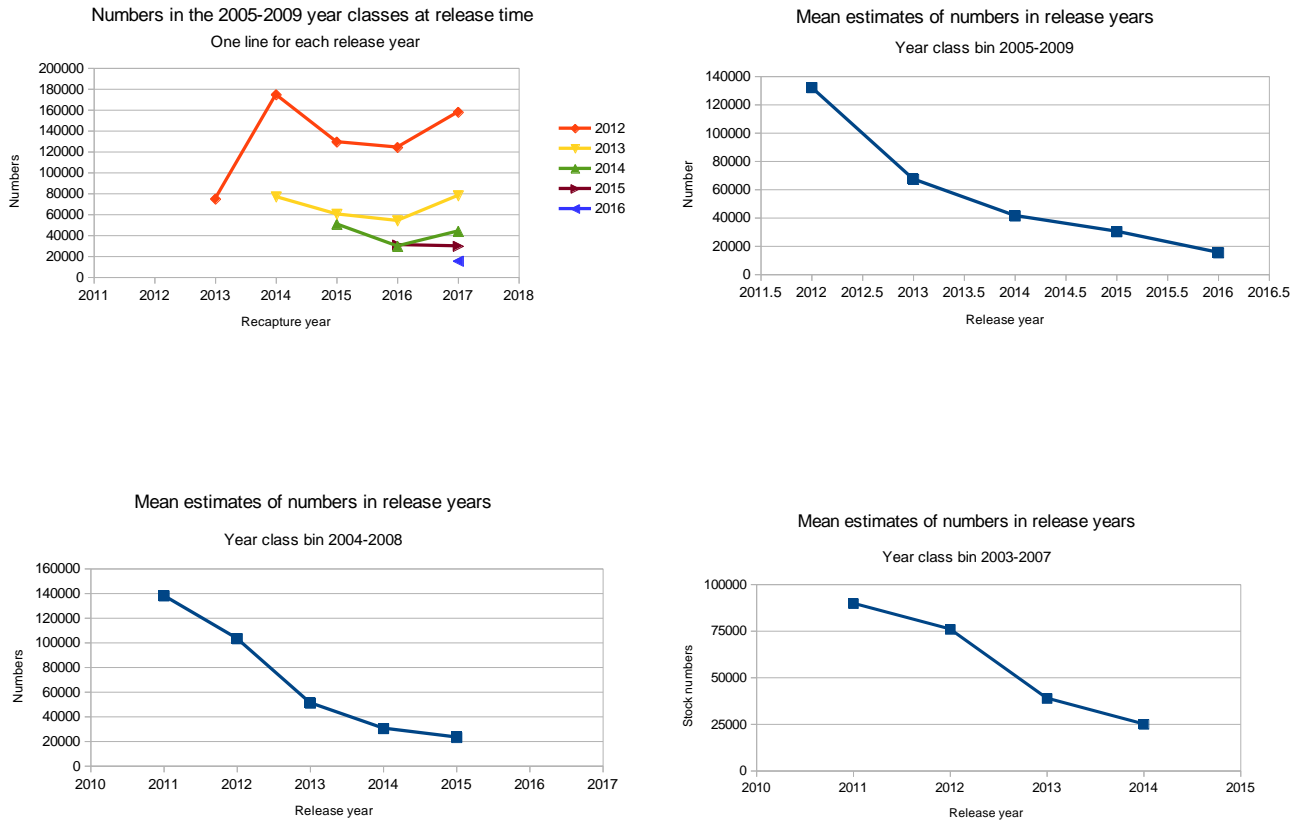


Figure 10. Upper left: Estimated stock numbers in the year classes 2005-2009 at release time as indicated. Other panels; Means over all recapture years, for the year class bins indicated.

Other year class bins are less regular. Both the 2004-2008 and the 2003 - 2007 year classes have a sigmoid pattern, both giving a Z -estimate of 0.20. Altogether, merging year classes is helpful, although there still is noise in single year estimates. In all these calculations, recaptures in the release year were not included.

4.2. Mortality estimates.

When estimating mortality as the *ratio of recaptures in two consecutive releases*, merging year classes may remove some noise, but the results are not satisfactory. The figure 11 below can be compared with Figure 7. The 2013-14 releases still give a strange pattern, and the others vary between 2 and -1, which is not very informative.

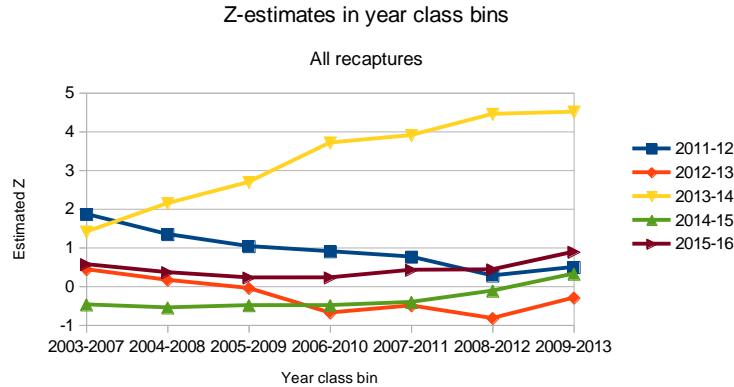


Figure 11. Mortality estimates by comparing recaptures from subsequent releases, by year class bins.. One line for each pair of releases.

When estimating mortality by taking the linear regression of **log recapture-release ratios from several release years in the same recapture year**, the results with year class bins look more consistent than with single year classes. As an example, the figures xy below compare the two year class types in the 2016 recapture. With single year classes, some releases stand out, which are those where the year class is still young. In both, the youngest (below 3 years at release) are not included.

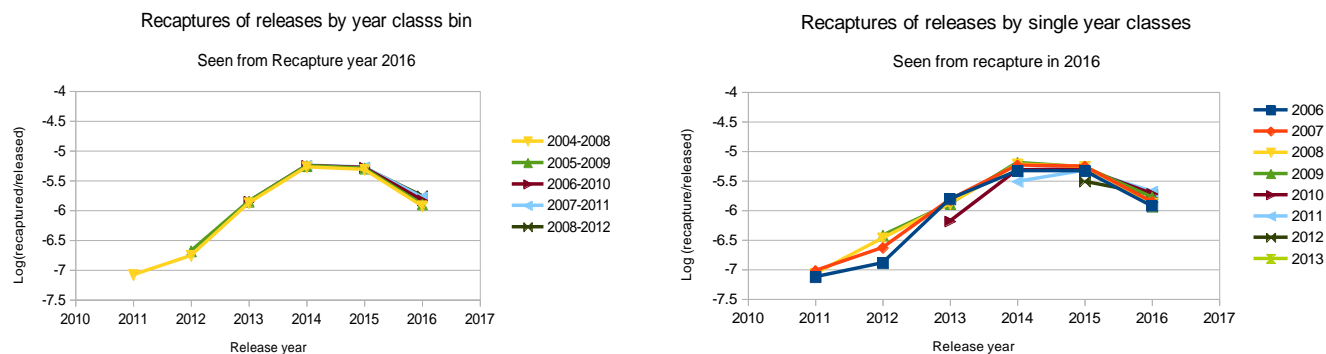


Figure 12 Disappearance of tags. Log (number recaptured in 2016/number released in years on the abscissa). Calculated for the year class bins indicated (left) and single year classes (right). The slope of the curves is an estimate of the total mortality.

In both, the curves bow down towards the most recent releases. This is even more prominent in the 2017 recaptures, shown in Figure 13. The most likely reason for this is less recaptures of fish tagged in the recapture year, and to some extent in the year before, as discussed previously.

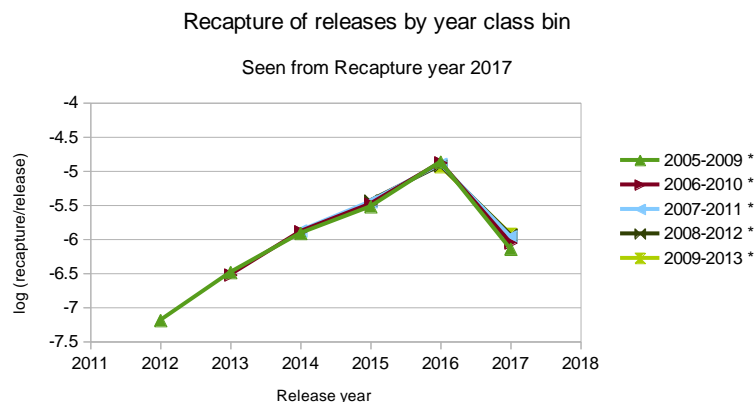


Figure 13 As for Figure 12, but for recaptures in 2017.

The slope of the curves should indicate the total mortality. If the most recent releases are excluded, most regression slopes are in the order of 0.5 - 0.6. This is more than indicated in the N-estimates. The reason for that has not been clarified, and might warrant a more in depth examination. One should at least keep in mind that the estimates here is of the apparent mortality, which includes both fish dying and fish becoming less accessible to recapture. Including the most recent releases reduces the slope, of course, but then the curves are not straight any more.

It is also worth noticing that the curves are very much on top of each other. This may indicate a rather flat selection at age, which is different from what is estimated in the assessment.

5. Suggestions for improvements.

Two major problem areas have emerged from this study, the ageing of released fish and the amounts and age distribution of the screened fish. Two further problems are the mortality at release, which we have to assume is similar from year to year, and the incomplete mixing between sub-stocks, by which the screened part of the population may not be representative of the part that was tagged.

To reduce the impact of these problems, we suggest that:

1. One should try to avoid relying on the amount scanned. That can be done by using the tag returns as a measure of mortality rather than of abundance.
2. One should avoid relying more than necessary on ratios between numbers that may sometimes be small with a huge *relative* error. The second variant of the Z-estimate outlined above may be one possible way, in particular combined with merging of year classes.
3. One may also consider ways to avoid applying a wide range of ages to each fish, referring it to several year classes with only a small fraction to each. We probably want to stick to cohorts that we can regard as closed, like year classes, where disappearance is equivalent to mortality.
 - One alternative could be to merge ages at release, and operate with multiple-age cohorts. Some examples showing the effect of doing so are described in Section 4.
 - Another suggestion might be to consider length cohorts. Admittedly, they are not perfectly closed - a fish may go from one length cohort to another. Nevertheless, regarding the central part of the length distribution as a cohort, and increment it according to mean growth each year would perhaps cause less problems than a broad age distribution.

Table 3 For each year class: Numbers released and estimated stock numbers (millions) by recapture year

R at release and Nscan R/r*Nscan

Estimated N (millions) at Release in 2011

Ycl Nrelease Recap. year

- -	2012	2013	2014	2015	2016	2017
1999	51	650.	-1	-1	-1	-1
2000	202	469.	340.	-1	308.	276.
2001	326	353.	375.	1171.	146.	523.
2002	768	1479.	1101.	2160.	1735.	2661.
2003	825	3665.	2579.	4159.	5172.	2264.
2004	1297	7697.	5221.	5257.	8577.	15690.
2005	1701	19197.	21357.	16269.	22897.	20788.
2006	3169	31287.	24943.	38459.	43940.	56904.
2007	1985	29557.	35297.	46116.	66833.	60549.
2008	5407	41179.	54028.	64065.	80029.	81179.
2009	2588	23204.	33263.	51139.	53931.	70241.

Estimated N (millions) at Release in 2012

Ycl Nrelease Recap. year

- -	2013	2014	2015	2016	2017
2000	707	436.	399.	41.	6337.
2001	1538	469.	541.	77.	1892.
2002	1503	882.	1552.	831.	7695.
2003	2565	3434.	4532.	3221.	4409.
2004	3327	6811.	5927.	4927.	14172.
2005	4907	18226.	18510.	17640.	17436.
2006	5248	23161.	37208.	23704.	44963.
2007	4422	24049.	46124.	43229.	41055.
2008	4336	22642.	53045.	42509.	44534.
2009	1465	17232.	42621.	32650.	37606.
2010	1098	30949.	79751.	89399.	115438.

Estimated N (millions) at Release in 2013

Ycl Nrelease Recap. year

- -	2014	2015	2016	2017
2001	1151	184.	30.	466.
2002	984	457.	294.	1129.
2003	2041	1484.	1290.	865.
2004	3291	2918.	2122.	4526.
2005	5696	7613.	6552.	6649.
2006	2698	17121.	11745.	15253.
2007	2836	18777.	20440.	18188.
2008	1658	26254.	20382.	24869.
2009	537	26929.	17390.	22071.
2010	474	84141.	93342.	65964.
2011	1017	100518.	45420.	59262.

Estimated N (millions) at Release in 2014

Ycl Nrelease Recap. year

-	-	2015	2016	2017
2002	1166	201.	704.	204.
2003	1014	855.	629.	1216.
2004	3049	1527.	2435.	2202.
2005	3530	4643.	4005.	7112.
2006	8271	8463.	9471.	11522.
2007	11206	16888.	10168.	15852.
2008	12343	20251.	12747.	18895.
2009	6902	17826.	10980.	22579.
2010	4013	37220.	27720.	35976.
2011	2082	36161.	35044.	82099.
2012	1102	6593.	20245.	101939.

Estimated N (millions) at Release in 2015

Ycl Nrelease Recap. year

-	-	2016	2017
2003	1072	519.	749.
2004	1513	3088.	1147.
2005	5148	4053.	3952.
2006	6500	9477.	6935.
2007	5620	10354.	11054.
2008	8470	13263.	13382.
2009	5272	11838.	15517.
2010	5274	27277.	22186.
2011	2955	29035.	24834.
2012	829	14291.	14131.
2013	464	9277.	11441.

Estimated N (millions) at Release in 2016

Ycl Nrelease Recap. year

-	-	2017
2004	1602	532.
2005	5667	1972.
2006	4965	3496.
2007	5454	5590.
2008	5699	7168.
2009	3201	8210.
2010	6436	14547.
2011	5070	17356.
2012	1484	8580.
2013	1874	4138.
2014	804	17118.

Spatio-temporal distribution of RFID tag recaptures

Konstantina Dimitrakopoulou and Aril Slotte

Contact: Email: Konstantina.Dimitrakopoulou@hi.no and Aril@hi.no, Institute of Marine Research, Bergen, Norway

Background

During ICES WGWIDE 2018, and the inter-mediate process after this, it has been discussed whether the underlying assumptions of mixing in the RFID tag data is true; that is that the tagged fish within a year class is behaving in the same way as the rest of the fish within this year class, with same mortalities after initial tagging mortality. It has been demonstrated a tendency that tagged fish being out for several years are being recaptured at a lower rate than expected, suggesting a long term extra mortality relative to non-tagged fish. However, it has also been suggested that imperfect mixing may cause similar tendencies.

It seems that an overview of distribution of recaptures from the various release years over all recapture years may be relevant to look more at with regard to the discussion about mixing. Moreover, given that fish at different sizes have different migrations, it would also be of interest to look more into spatio-temporal distribution between different sizes of tagged fish. Another issue is if fish tagged in different areas may have similar distribution of recaptures and recapture rates; here the potential of components with different migration patterns that may occur in a situation we have been in in recent years with expanding distribution and a tendency of more north and westward shift in distribution

Objectives

The objective of the present WD is simply to give an overview of recaptures from all different release years, over all different recapture years, from different size groups, and from different release areas. To see if there are potential trends that may interfere with the way the data are used in the SAM model.

Methods

Data used for the analysis are the ones going into the assessment, coming releases of Ireland and the British Isles and recaptures from landings at factories with RFID systems in Norway, Iceland, Faroes and Scotland (Figure 1). A lot of maps has been produced showing the exact coordinates for release data (from GPS), whereas for recaptures the ICES rectangles center coordinates are used. Release data are shown in green, whereas recaptures are shown in light to dark red colours, where more recaptures within same ICES rectangle is shown by darker red. Note that the recapture data of 2018 cover the period until February 2018 as data base was not updated at point of analyses. PS! Note that maps are very small, but they are made to be zoomed up for further interpretation, they are grouped as many maps on one page to give an overview. Note that we have each of these maps on files that may be retrieved from the authors at request.

Results

At first, we analyzed the Haversine distance distribution (distance was calculated for every pair of release and recapture coordinates), to see if the median distance migrated changed between release years and years out (Figure 3). There seems to be a tendency with highest distance between tagging position and recapture in release years 2012-2013, being somewhat lower afterwards. Also a tendency of reduced

distance between release and recapture position with number of years between release and recapture. One might expect the opposite if the tagged fish is growing, and migrating longer as it grows, but this may also indicate a tendency that a larger proportion of the scanned fish is now closer to tagging area.

When looking at the distribution of recaptures from all release years, and recaptures years (years out) (Figure 3), it is clear that a change with more recaptures in Iceland and Central Norwegian Sea is developing over time, but this is also mainly a result of the new factories coming in Iceland and Faroes, as well as a change in the Norwegian fishery going from North Sea to central Norwegian Sea. However, when one are moving closer to the latest tagging years, and also looking at recaptures several years after release, there is also a tendency that the recaptures are clumping around Scotland, which again may be linked to the change in scanning with Scottish factories joining in from 2014 onwards. Yet, again there also seems to be a change which is more related to a movement of the Norwegian and Scottish fishery in autumn and winter being closer to Scotland, a change that may have affected recapture rates, if they are higher closer to the tagging area.

The overall recapture years distribution of recaptures related to size at release, i.e. the recaptures all years out from splitted data into length bins at release (Figure 4), show a tendency that the smallest and largest fish does not migrate as far north and west, such as to Iceland, as the medium sized fish. When you for example look at Figure 4, length bins 3 and 4, in 2013 and 2014, they have rather different the distributions, and seems to be now recaptures of the largest bin 4 fish around Iceland in 2013, but a lot of bin 3. Some of this may be due to differences in numbers recapture in total between the two size groups, but one would have expected more recaptures of the biggest fish around Iceland following the idea of length dependent migration. Hence, this is worth looking more into, if these are signs of components with different migration routes etc. In Figure 5a-d, size differences in recapture distribution are also shown for each length bins for each recapture years instead of merged years. An interesting trend here that the smallest fish distribute closer to tagging area in first year after tagging than larger fish, indicating that the youngest tagged fish does not fully distribute in the area the rest of their conspecifics are distributed in terms of fishery as well as the IESSNS survey.

When comparing recaptures from three different release areas from south Ireland to west of Scotland (Figures 6a-c), there seems to be some effects of where the fish is released, even though distances are not as large. As an example below is maps picked from Figures 6a and 6c, showing distribution of recaptures from Area1 and Area 3 for in year (Year 0) and first year after release (Y1). It seems that mackerel tagged at the south tip of Ireland is more recaptured of Iceland than those tagged of north Ireland.

As a follow up on this and example from 2014, where we compared recaptures from two experiments, Experiment 29 at south tip of Ireland, and Experiment 27 at north Ireland. 20,000 mackerel were released in both experiments, but recaptures from experiment 27 in north Ireland was 30% lower than experiment 29, with a tendency to have less tags coming back in Icelandic waters. We have few examples like this where large numbers tagged in different areas, but they suggest that we may have to do with some components, with different migrations, which may interfere with the stock assessment, and is worth following up in the

future. The tag data certainly may help understanding migrations of this stock in addition to the use in assessment of the stock.

In 2011 we purse seined young mackerel off the Norwegian coast in the North Sea, and tagged totally 31253 fish, of which 27635 fish were estimated to be 2010 year class (1 year olds). The maps of recaptures (Figure 8) indicate how this year class developed it's migrations from age 2 to age 8. It seems that it did not really distribute southwards along British Isles in large proportions before ages 5-6.

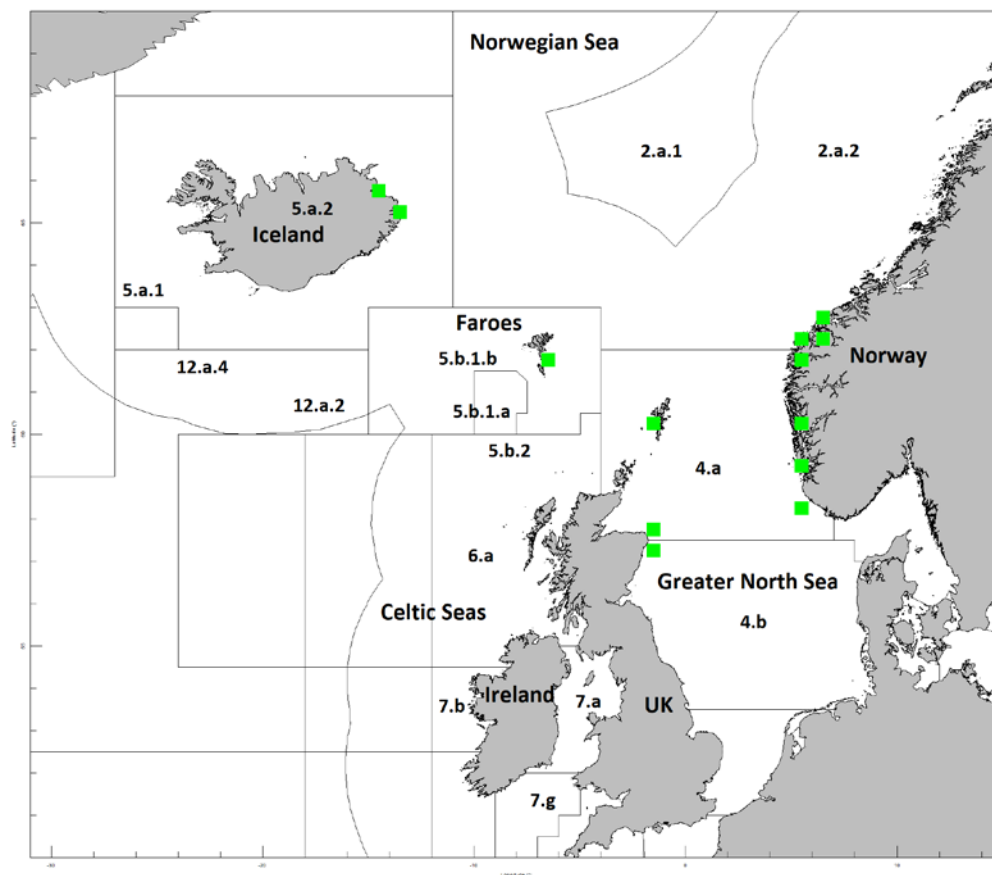


Figure 1: Factories: map showing the locations where the processing of catches is done.

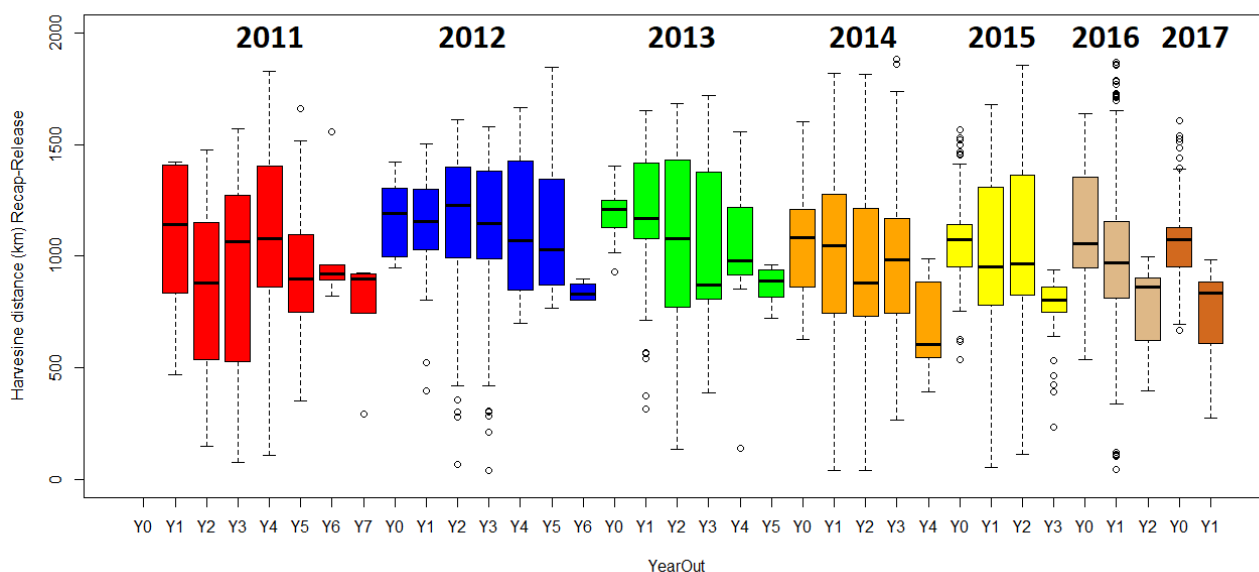


Figure 2. Distance_boxplot: each color refers to one release year and Y0 corresponds to the recaptures within the same year, Y1 to the recaptures one year later etc. The y-axis shows the Haversine distance distribution (distance was calculated for every pair of release and recapture coordinates).

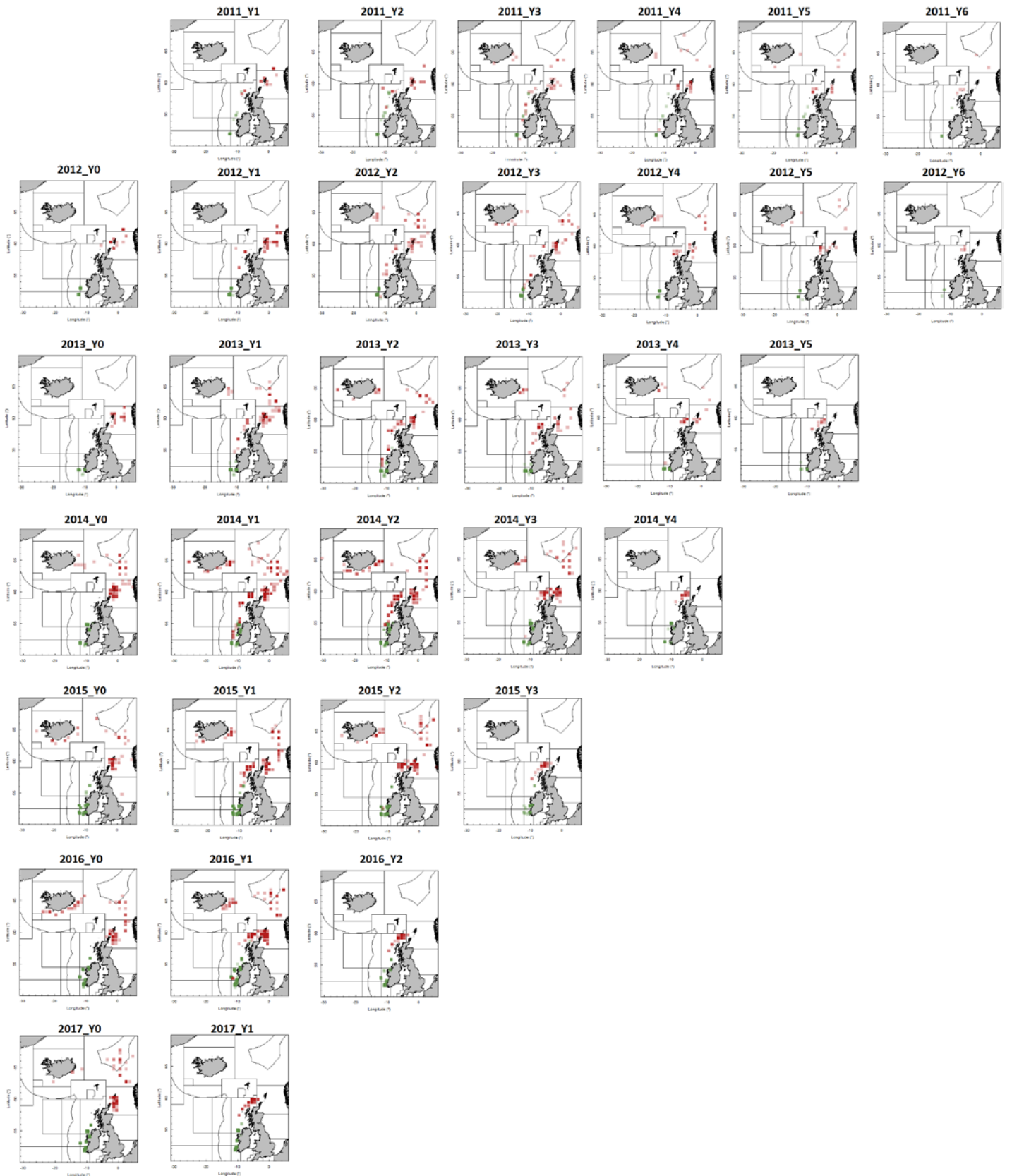


Figure 3. Recapture maps per Year Out : each map shows per release year and per recapture Year Out the locations of releases and the corresponding recaptures (green- releases, red-recaptures, density of color increases as points overlap). E.g. the per_2012_Y0 map shows the subset of fish that was released in 2012 and recaptured within the same year. The per_2012_Y1 shows the subset of fish that was released in 2012 and recaptured one year later etc.

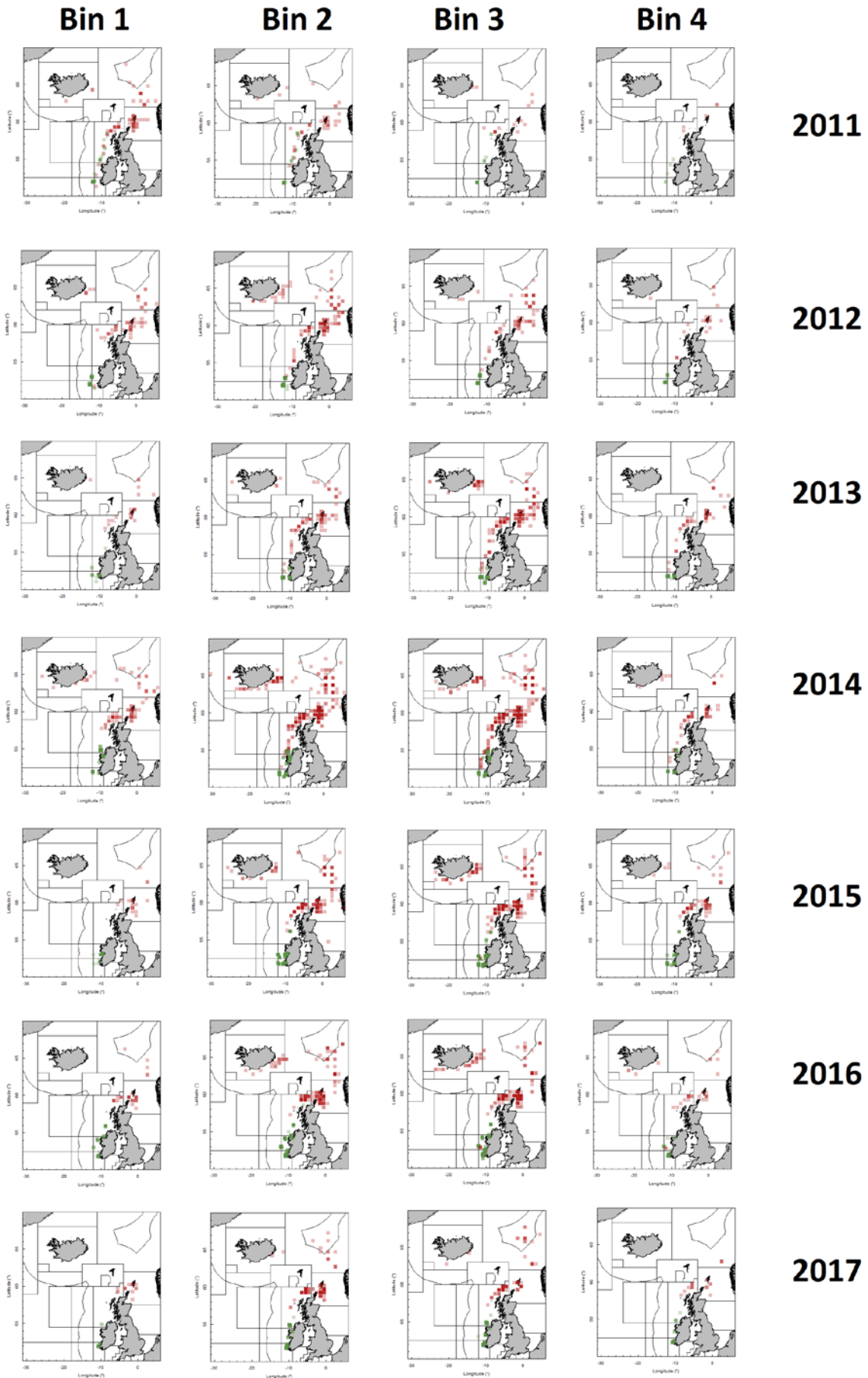


Figure 4. Recapture maps per length bin: each map shows per release year and per Length bin (as defined at the release), the locations of releases and the corresponding recaptures (green- releases, red-recaptures, density of color increases when points overlap). E.g. the merged_bin1_2012 map shows where the Bin 1 fish released in 2012 were recaptured at any time point after release. The merged_bin2_2012 shows where the Bin 2 fish released in 2012 were recaptured at any time point after release. **Length bins (cm).** Bin 1: <32 cm, Bin 2: 33,35 cm, Bin 3: 36-38 cm, Bin 4: > 38 cm.

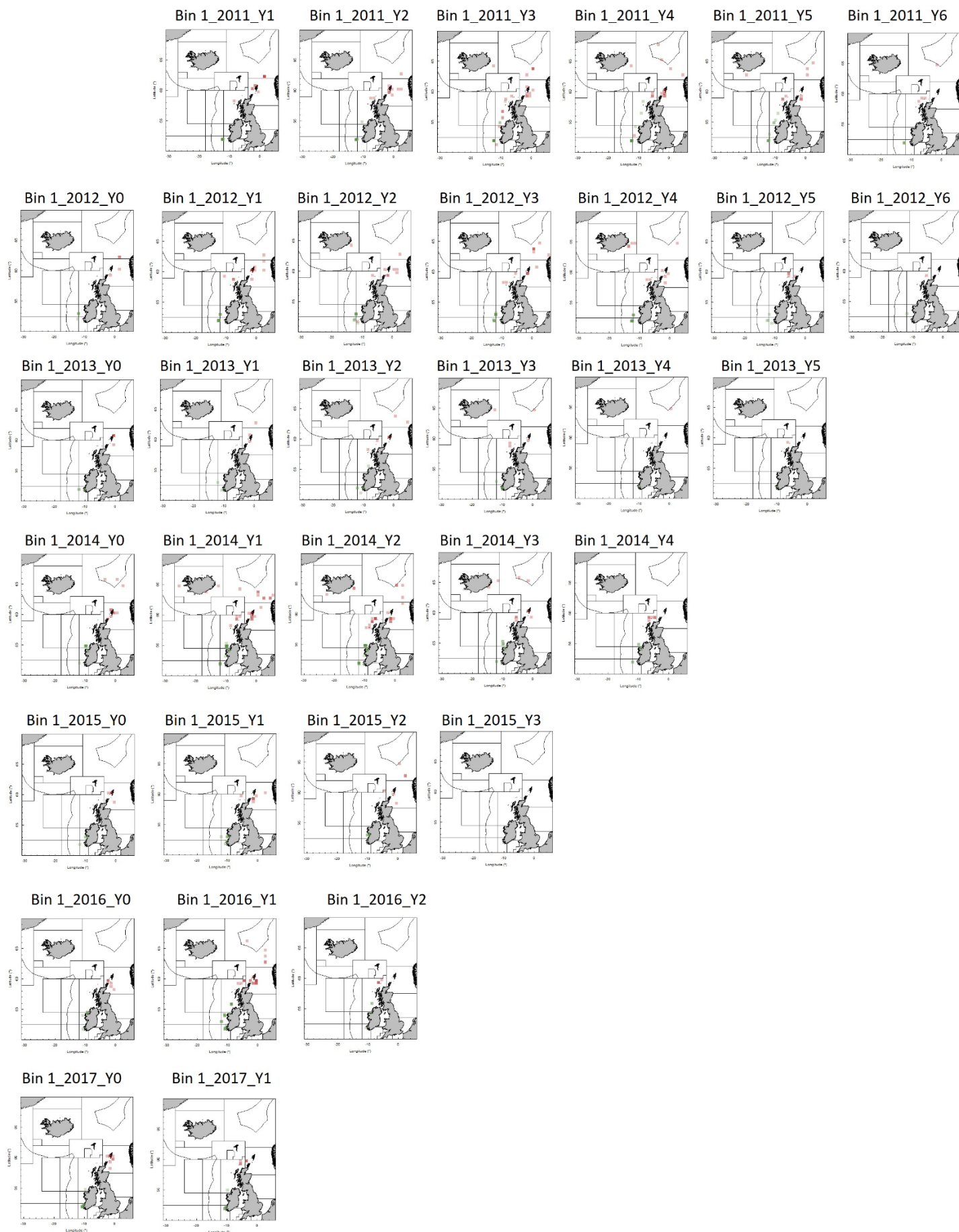


Figure 5a. Recapture maps per Year Out and Length bin. Here Bin 1: <33 cm: each map shows per release year, per recapture Year Out and per Length bin (as defined at the release), the locations of releases and the corresponding recaptures (green-releases, red-recaptures, density of color increases when points overlap). E.g. the bin1_2012_Y0 map shows where the Bin 1 fish released in 2012 were recaptured within the same year. The bin1_2012_Y2 shows where the Bin 1 fish released in 2012 were recaptured two years later.



Figure 5b. Recapture maps per Year Out and Length bin. Here Bin 2: 33-35 cm: each map shows per release year, per recapture Year Out and per Length bin (as defined at the release), the locations of releases and the corresponding recaptures (green-releases, red-recaptures, density of color increases when points overlap). E.g. the bin1_2012_Y0 map shows where the Bin 1 fish released in 2012 were recaptured within the same year. The bin1_2012_Y2 shows where the Bin 1 fish released in 2012 were recaptured two years later.



Figure 5c. Recapture maps per Year Out and Length bin. Here Bin 3: 36-38 cm: each map shows per release year, per recapture Year Out and per Length bin (as defined at the release), the locations of releases and the corresponding recaptures (green-releases, red-recaptures, density of color increases when points overlap). E.g. the bin1_2012_Y0 map shows where the Bin 1 fish released in 2012 were recaptured within the same year. The bin1_2012_Y2 shows where the Bin 1 fish released in 2012 were recaptured two years later.



Figure 5d. Recapture maps per Year Out and Length bin. Here Bin 4: >38 cm: each map shows per release year, per recapture Year Out and per Length bin (as defined at the release), the locations of releases and the corresponding recaptures (green-releases, red-recaptures, density of color increases when points overlap). E.g. the bin1_2012_Y0 map shows where the Bin 1 fish released in 2012 were recaptured within the same year. The bin1_2012_Y2 shows where the Bin 1 fish released in 2012 were recaptured two years later.



Figure 6a. Recapture maps per release area – Here Area 1: Latitude <52: each map shows per release year, per recapture YearOut and per area, the locations of releases and the corresponding recaptures (green- releases, red-recaptures, density of color increases when points overlap). E.g. the area1_2017_Y1 map shows where the fish from area 1 released in 2017 were recaptured one year later.



Figure 6b. Recapture maps per release area – Here Area2: Latitude [52-54]: each map shows per release year, per recapture YearOut and per area, the locations of releases and the corresponding recaptures (green- releases, red-recaptures, density of color increases when points overlap). E.g. the area1_2017_Y1 map shows where the fish from area 1 released in 2017 were recaptured one year later.

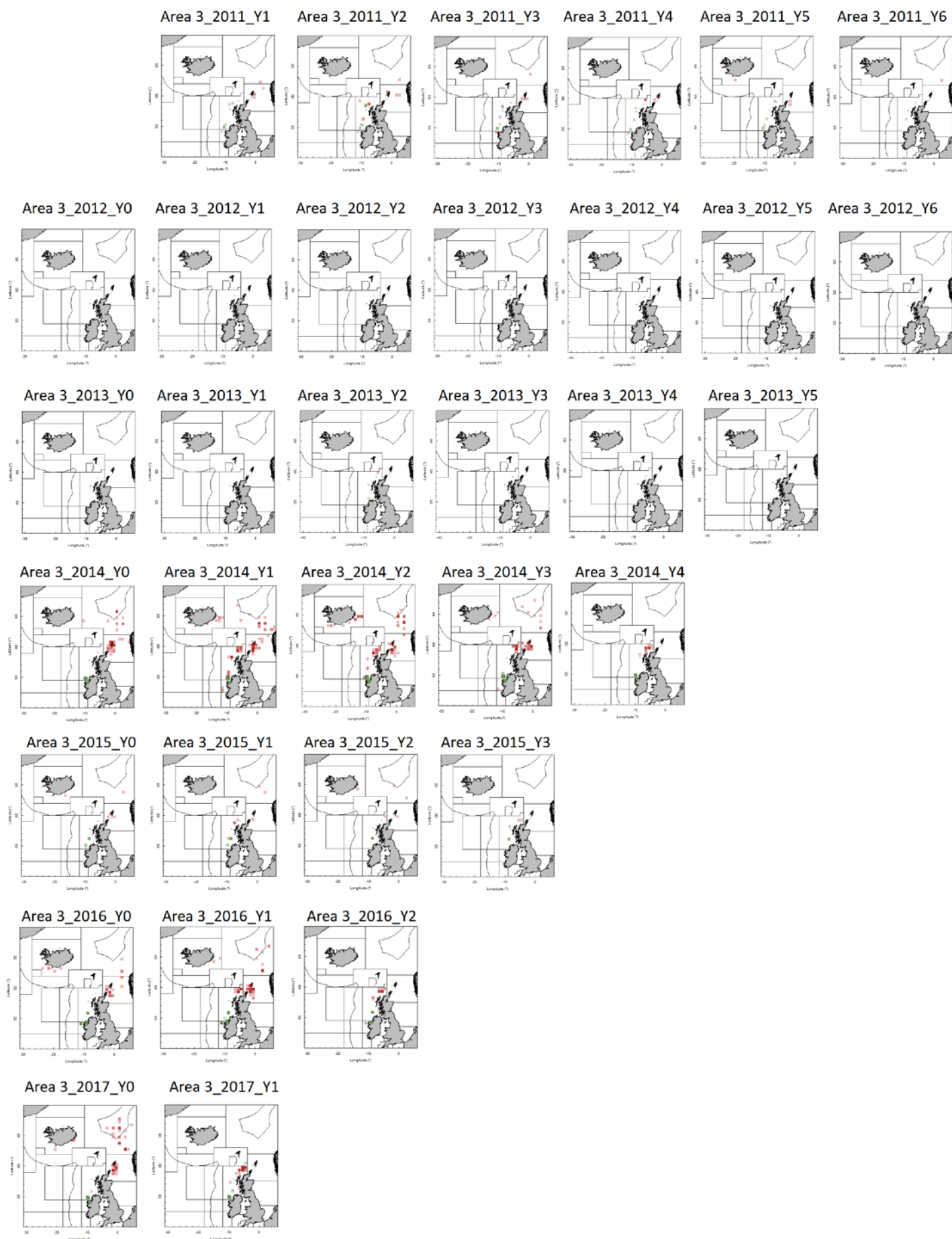


Figure 6c. Recapture maps per release area – Here area 1: Area3: Latitude ≥ 54 : each map shows per release year, per recapture YearOut and per area, the locations of releases and the corresponding recaptures (green- releases, red-recaptures, density of color increases when points overlap). E.g. the area1_2017_Y1 map shows where the fish from area 1 released in 2017 were recaptured one year later.

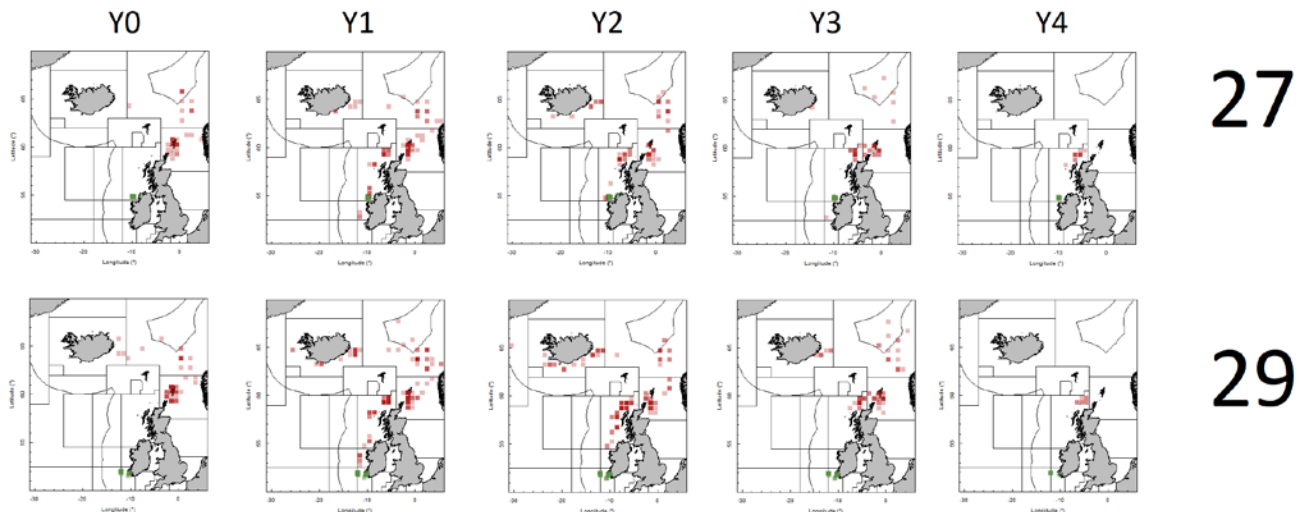


Figure 7. Recapture maps from 2 experiments in 2014. Experiment 29 at south tip of Ireland, and Experiment 27 at north Ireland. 20.000 mackerel were released in both experiments, but recaptures from experiment 27 was 30% lower than experiment 29, with a tendency to have less tags coming back in Icelandic waters.

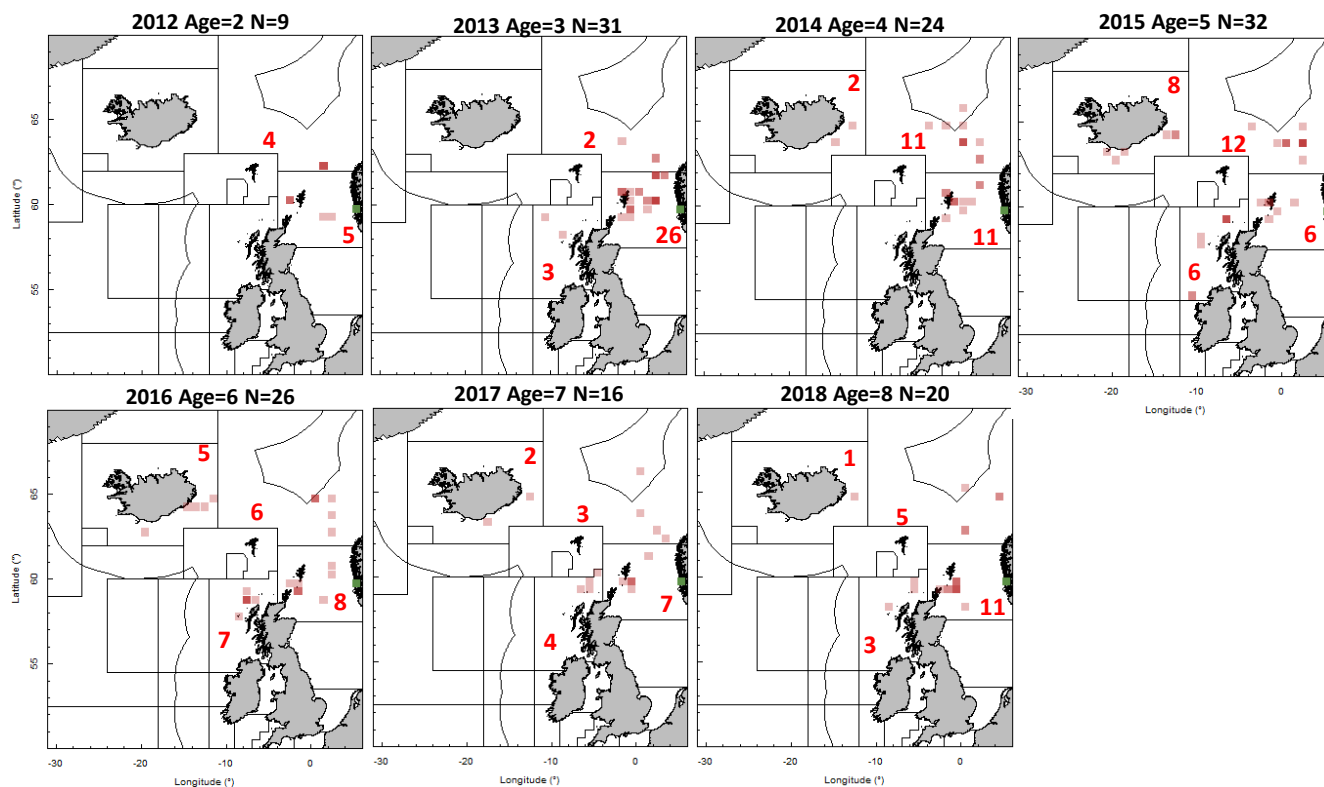


Figure 8. Recapture maps from and a tagging experiment mainly on 1 year olds of the 2010 year class. In 2011 we purse seined young mackerel off the Norwegian coast in the North Sea, and tagged totally 31253 fish, of which 27635 fish were estimated to be 2010 year class. The maps of recaptures therefore indicate how this year class developed it's migrations from age 2 to age 8. Total numbers recaptured are shown on top of each map, and number recaptured by ICES area are shown in red, number by ICES rectangle is shown by darker color when more recaptures.

Deriving Auxiliary Stock Information from the mackerel RFID Tags

A Working Document Submitted to the Inter-Benchmark Workshop on the Assessment of NE Atlantic Mackerel,

IBPNEAMac 2019

S. Subbey, B. Sandercock, K. Dimitrakopoulou, N. Jourdain, D. Skagen, A. Slotte, S. Vatnehol

February 4, 2019

Abstract

This document presents a model-based approach for obtaining survival rates from NE Atlantic mackerel tag data. Estimates of natural mortality rates and profiles are obtainable from the survival rates, when fishing mortality rates can be estimated as exogenous parameters.

Results obtained by applying the Brownie tag-recovery model to the RFID tags data show that the total (sum of natural and fishing) mortality rate is age-dependent, and convex (bathtub-shaped) in profile. The accuracy and validity of results is conditioned on a number of fundamental assumptions (e.g., that there is a good mixture of the tagged and untagged sub-populations). Because these assumptions have not been thoroughly investigated, the results and conclusions reported in this WD must be considered as preliminary.

1 Background

One source of data used in stock assessment is from two tagging time series. The first series, from steel tags, has data from 1980 to 2006, while the RFID tag data covers a period from 2011 to present. A brief description of the data sets follows.

1.1 Steel tags

Tagging involved the use of a hired purse seine boat where the mackerel was caught by manual jigging using four to five wheels alongside a boat (Antsalo, 2006). A typical steel tag is small, rounded at the ends, 20 mm long, 4 79 mm wide and 1 mm thick with a unique serial number. Once the mackerel were brought on board the vessel, the general health of each mackerel was evaluated. Those evaluated to be at good health were tagged, while the others were aged and then used to develop an age-length key. Length measurements were also taken. Fish were measured at 1 cm length group.

At the factories, the fish are scanned using metal detectors. When detected, tagged fish are taken out of the production line, where age and length readings are taken, as well as the tag ID-number. The ID-numbers are used to link the recaptured tag to its release time and location.

1.2 The RFID tags

This class of tags are small, and with a radio transmitter within a glass vial. During the tagging procedure, the length of each tagged fish is measured. The age-length distribution are constructed using untagged fish caught in the same location and time as the tagged.

At the factories, the fish are scanned using radio detectors that capture unique codes transmitted by the tags. In contrast to steel tags, the RFID-tagged fish is never removed from production. Its data is automatically uploaded to a database at the Institute of Marine Research (IMR, in Bergen) and linked to information at release (Tenningen et al., 2011).

1.3 Age Determination – Age-length Keys (ALK)

ALKs for fish tagged with steel tags are derived from the tagged fish. That is, when a fish is recovered, the length is measured and otoliths are removed for age determination. For RFID tagged fish, age is linked to the age length key (ALK) at time of release, see Tables 1 and 2. Missing ages are imputed using the closest neighbour length groups.

Length (cm)	Number at age a						Total
	1	2	3	4	5	6	
36	30	45	8	36	3	2	124
37	3		2	5	9	7	31
38	6	4	3	6	1	4	24

Table 1 Number at length and age

Length (cm)	Probability of being at age a						Total
	1	2	3	4	5	6	
36	0.24	0.36	0.06	0.29	0.02	0.02	1
37	0.10		0.06	0.16	0.29	0.23	1
38	0.25	0.17	0.13	0.25	0.04	0.17	1

Table 2 Probability of being at age a , given length

1.4 Assumptions for the use of tag data

1. Stable tagging mortality (mortality due to handling and tagging)/tag loss
2. Stable scanning efficiency (% tags detected when passing antenna)
3. Correct biomass reported landed at the factories (correct numbers scanned)
4. Representative mixing (same behavior as untagged fish)
5. Same mortality as untagged fish after initial tagging mortality (from year 1 after tagging)

2 Deriving Population Indices—The Brownie tag-recovery model

The Brownie tag-recovery model is a special case of the Cormack Jolley Seber (CJS) models (see e.g. Pledger et al., 2003), which is useful for estimating harvest mortality and survival. Like all CJS models, the Brownie model is conditional on the animals being previously marked and released. In this tag-recovery model, encounters happen in a single way: dead recovery, either killed and reported by hunter, or found dead and reported by the general public. The main assumption of this model is that animals survive to the first season. If mortality occurs between the time of tagging and the first hunting (harvest) season, the Brownie estimator would be negatively biased (Buderman et al., 2014).

The Brownie tag-recovery model is a parameterization of the tag-recovery probability when recovery is a result of hunter harvest (Buderman et al., 2014). In this model, *two* types of parameters can be estimated: **survival rates** (denoted by S) and recovery or "reporting" rates (denoted f or r depending on the specific model parameterization). If the parameterization is in terms of S and f , the model is the band or ring recovery model of Brownie et al. (1985). Here, S is "pure" survival as opposed to "apparent" survival. This is so because in recovery studies there is no notion of permanent emigration from the study area, and thus animals are always subject to encounter.

Let S denote the probability that a tagged and released fish survives the first year. The probability that the fish dies (either of natural causes or due to harvesting) is $1 - S$. If the probability that it is harvested is K , then the natural mortality probability is given by $1 - S - K$. If we c and r represent respectively, the probability of retrieving a killed fish, and reporting a retrieved individual, then without further information, it is impossible to separate K , c and r — only their product (Kcr) can be estimated.

The probability product, $K \cdot c \cdot r$ (being killed, recovered, and reported), is termed the recovery rate (f), and the probability of being killed and recovered ($K \cdot c$) is termed the harvest rate (H). It is trivial to derive that $f = Hr$, so that when reporting of harvest is certain, that is $r = 1$, then the recovery rate equals the harvest rate, (Buderman et al., 2014; Cooch and White, 2001). Combining S and f , one can determine the expected numbers of bands recovered in a particular time period.

Suppose N_1 fish were marked and released, the expected number of fish that will be harvested, retrieved, and reported (i.e., recovered) at the end of the first interval after marking is $N_1 f$. Figure 1 is an example fate diagram for the expected number of recoveries, assuming that both survival and recovery probabilities are time-specific. Note S_i is the probability of surviving from time $(i - 1)$ to time (i) , whereas f (recovery rate) is the probability of not surviving, retrieved and reported. For example, for individuals marked in year 1, the number of expected dead recoveries in the second interval after marking is given as $N_1 S_1 f_2$. This is so because for the fish to be a dead recovery in the second interval, it has to survive the first interval (with probability S_1), and then be harvested, retrieved and reported (with probability f_2) (see Cooch and White, 2001).

year marked	number marked	year recovered			
		1	2	3	l = 4
1	N_1	$N_1 f_1$	$N_1 S_1 f_2$	$N_1 S_1 S_2 f_3$	$N_1 S_1 S_2 S_3 f_4$
2	N_2		$N_2 f_2$	$N_2 S_2 f_3$	$N_2 S_2 S_3 f_4$
3	N_3			$N_3 f_3$	$N_3 S_3 f_4$
k = 4	N_4				$N_4 f_4$

Figure 1 Fate diagram for expected number of recoveries, given N_1 has been marked and released (Cooch and White, 2001)

The Computational Model and Software

Given the parameters as defined above, and the data, a multinomial model can be constructed to estimate survival S_i and recovery f_i rates using maximum likelihood methods. These estimates are determined under the assumption that no mortality occurs between tagging and the first harvesting season.

In our analysis, we have used the Brownie model as implemented in the Program Mark: (<http://www.phidot.org/software/mark/docs/index.html>).

3 Main Results

Figure 2 is a summary of the main results¹, which show that

- Recovery rates are extremely low (under 0.5%) for all years
- Survival rates (and therefore total mortality rates) are age-dependent and dome-shaped (survival) or bathtub-shaped (total mortality)
- The stock can only be sustained by high compensatory fecundity (from ages 5-10 fish), given a fishing mortality of between 20-30%.

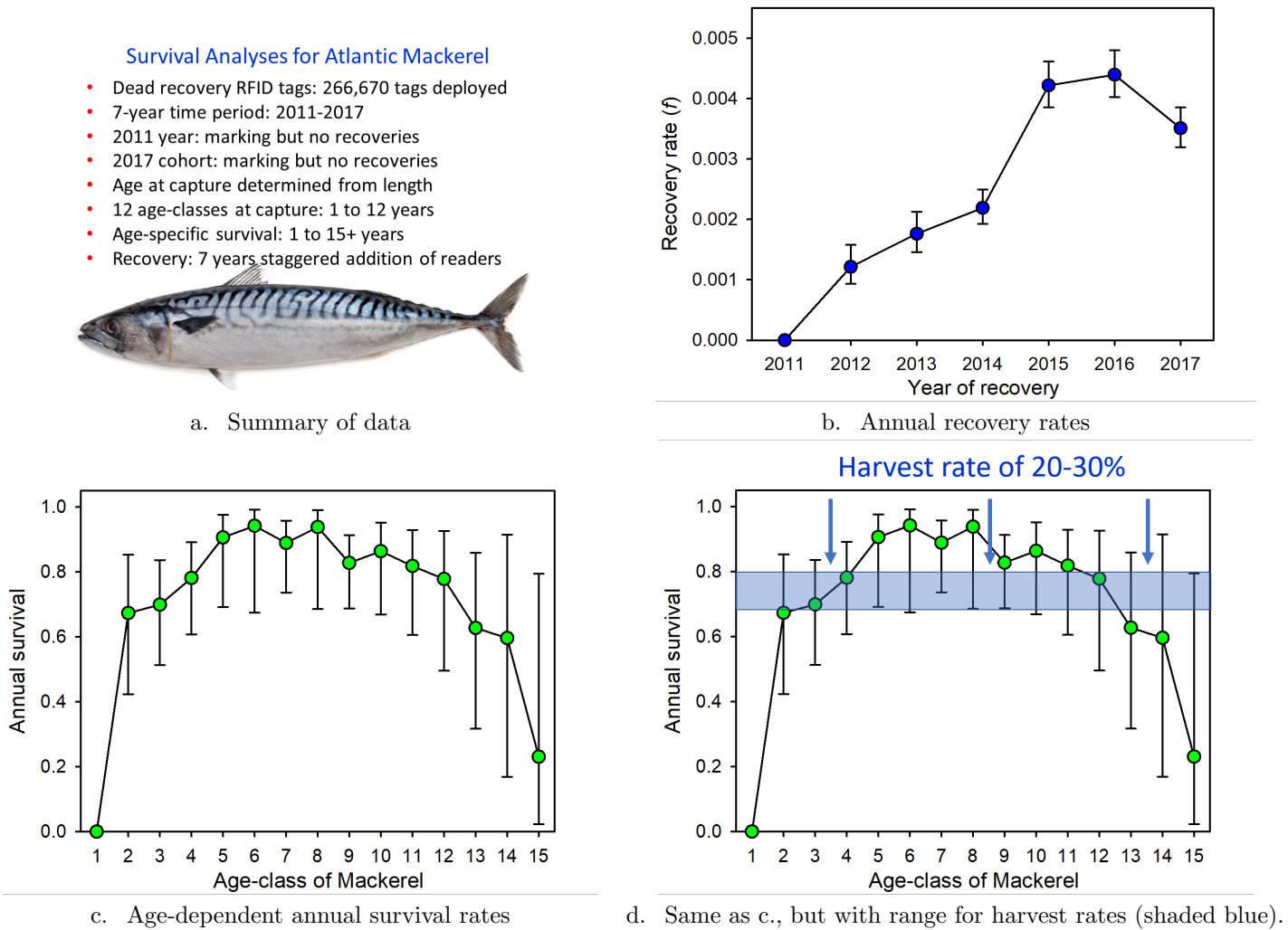


Figure 2 Main modeling results using only the RFID tags

¹See caveats in abstract

References

- Antsalo, M. (2006). *Abundance estimation of the Northeast Atlantic mackerel (*Scomber scombrus*) with use of Norwegian tag data*. University of Bergen, Department of Biology Bergen Norway. PhD thesis, Dissertation 64 pp.
- Brownie, C., Anderson, D. R., Burnham, K. P., and Robson, D. S. (1985). Statistical inference from band recovery data. *US Fish and Wildlife Service Resource Publication*, 156.
- Buderman, F. E., Diefenbach, D. R., Casalena, M. J., Rosenberry, C. S., and Wallingford, B. D. (2014). Accounting for tagging-to-harvest mortality in a brownie tag-recovery model by incorporating radio-telemetry data. *Ecology and evolution*, 4(8):1439–1450.
- Cooch, E. and White, G. (2001). Using mark-recapture introduction. *Cornell University, Ithaca*.
- Pledger, S., Pollock, K. H., and Norris, J. L. (2003). Open capture-recapture models with heterogeneity: I. cormack-jolly-seber model. *Biometrics*, 59(4):786–794.
- Tenningen, M., Slotte, A., and Skagen, D. (2011). Abundance estimation of northeast atlantic mackerel based on tag recapture data: a useful tool for stock assessment? *Fisheries research*, 107(1-3):68–74.

Some results for various way to handling tag and catch observations

Magne Aldrin, Norsk Regnesentral

February 28, 2019

Contents

1 Introduction	2
2 Analyses	3
2.1 Old tag data as single observations	3
2.2 Updated tag data as single observations	3
2.3 Updated tag data as aggregated observations	4
2.4 Updated tag data as indices.....	4
2.5 Comparing the “best” models	5
2.6 Including correlations in catch data.....	7

1 Introduction

The motivation for the work presented here can be found in my Working Document from December last year: “When several and quite different data sources are included for estimating a model, the importance of each data source depends on its contribution to the total likelihood, which again depend on the actual formulation of the model. Therefore, the exact model formulation may be more important when several data sources are included than if only one data source is used.”

In this note, I use the tag data in three different ways; as i) single observations, ii) as observations aggregated over recapture years, and iii) as a set of indices proportional to the stock size, similar to a survey index. Furthermore, I vary how the correlations in the tag data are taken into account. The fit of comparable models are evaluated by comparing AIC values, and the SSB estimates of each model are shown by plots. These analyses are not meant to be a complete validation of each model, but more to illustrate the consequences of some overall model choices.

In addition, I illustrate what happens if we allow for correlations between ages in the catch data.

2 Analyses

2.1 Old tag data as single observations

First, I have used the “old” RFID data (before the updated data downloaded to the share point the 12th of February). I have fitted the same models as Anders in his Working Document downloaded to the share point the 12th of February. Table 1 show the log likelihood and AIC values and I got exactly the same results as Anders. According to the AIC values, the “Correlated tag-recapture” and “Correlated release x recapture” models give the best fit.

The effects on the estimated spawning stock biomass (SSB) are shown in the upper panel Figure 1. The differences between the models are quite large also between the two models with best AIC values (blue and green curves).

Table 1: Models with tag data as single observations, old RFID data.

Model name used in plots	Model description	log(L)	No. par.	AIC
spallyOldData	Spaly	-2785.8	23	5617.6
RelYCorrOldData	Correlated tag-release	-2770.8	23	5587.7
RecapYCorrOldData	Correlated tag-recapture	-2766.7	23	5579.4
RelYRecapYCorrOldData	Correlated release x recapture	-2766.9	23	5579.8

2.2 Updated tag data as single observations

Next, I have fitted similar models to the new, updated RFID data (Table 2 and second panel of Figure 1).

I have also fitted a model which accounts for both correlations between observations within the same release year *and* between observations within the same recapture year, i.e. cases b) and c) mentioned in my WD from December 2018 (“... it is reasonable to expect that they (the tag data) are correlated in several ways, both a) between the various recapture years of the same age group tagged the same release year and b) between age groups tagged the same year, and c) perhaps also between observations within the same recapture year.”)

This model gives clearly the best fit out of the candidates in this set of models. To fit this model, I had to modify the SAM code slightly, since this involves two separate groupings of random effects. Note that this is not the same as grouping together observations within the same release year *and* recapture years, and then assuming independence between groups, as in the release x recapture model. The SSB estimates for this model is given by the cyan curve in Figure 1.

Table 2: Models with tag data as single observation-as, new RFID data.

Model name used in plots	Model description	log(L)	No. par.	AIC
spally	spaly	-2947.6	23	5941.2
RelYCorr	Correlated tag-release,	-2933.3	23	5912.6
RecapYCorr	Correlated tag-recapture,	-2911.4	23	5868.7
RelYRecapYCorr	Correlated release x recapture,	-2911.5	23	5868.9
RelYANDRecapYCorr	Corr. release and corr. recapture	-2894.9	24	5837.9

2.3 Updated tag data as aggregated observations

Then, I have aggregated the tag data over release years, as suggested by Thomas in an email in January, but I don't think this has been uploaded to the share point as a Working Document. Anyway, I think it is a very good idea to aggregate over recapture years. It makes the data and the model simpler and more transparent. I have fitted two versions of these data, one using all data, and another ignoring ages less than 5. The summary results are given in Table 3, but note that the log likelihoods and the AIC cannot be compared to each other, or to Table 2, since they are based on different data. I will come back to their SSB estimates later.

Table 3: Models with tag data aggregated over recapture years, new RFID data. The AIC values cannot be compared across models due to different data.

Model name used in plots	Model description	log(L)	No. par.	AIC
AggTagRelYCorr	Aggregated, corr. release,	-1498.4	23	3042.8
AggTag5YRelYCorr	Aggregated, corr. release, $a \geq 5$	-1119.4	23	2284.7

2.4 Updated tag data as indices

The tag data can be transformed to new indices by variants of the LincolnPeterson estimator. I have used indices computed by 1) the sum formula presented in a Working Document by Sindre Vatnehol, 2) the sum Bayes formula presented by Sindre Vatnehol and Aril Slotte in another Working Document and 3) a so called sum Chapman formula, that probably are better suited for small samples than the previous ones. Using the same notation as Aril and Sindre, the sum variant of the Chapman Lincoln-Peterson estimator is defined by

$$K_{a,y} = \left[(R_{a,y} + 1) \left(\sum_n k_{a,n} + 1 \right) / \left(\sum_n r_{a,n} + 1 \right) \right] - 1 \quad (1)$$

Most of these models are fitted with a common catchability parameter and a common observation variance for all ages, but separate for steel and RFID tags. Their estimates of SSB are almost equal in recent years (black, red and blue curves in the third panel in Figure 1). However, one model, with tag indices based on the sumChapman Lincoln-Peterson, divides the age parameters into ages 2-4, 5-9 and 10-11. This gives the best AIC value and is thus the preferred model among these. This model gives the lowest estimates of SSB in the recent years (green curve in the third panel in Figure 1).

Table 4: Models with tag data used as indices, new RFID data. The AIC values can only be compared for the last two models.

Model name used in plots	Model description	log(L)	No. par.	AIC
TagAsIndSum	Index by sum formula	-909.6	25	1869.3
TagAsIndBayes	Index by sumBayes	-850.7	25	1751.5
TagAsIndChap	Index by sumChapman	-970.0	25	1990.0
TagAsIndChapSepAges	Index by sumChapman, separate age-par	-944.6	33	1955.3

2.5 Comparing the “best” models

The lower panel of Figure 1 shows the SSB estimates for the best (according to AIC) models where tag data are treated either as single observations or as indices, and the two models based on tag data aggregated over recapture years.

As long as all data are used, there is little difference between the SSB estimates for the models with i) tags used as single observations (black curve), ii) aggregated tag data (red curve) or iii) tags used as indices (green curve). The SSB estimates for the aggregated tags data without ages less than 5 years are a little bit higher than for the other models.

Common for all models are that the highest SSB estimates are in 2014-2015, as opposed to in 2011 in the “spaly” model used on “old” tag data in the 2018 benchmark (black curve, upper panel).

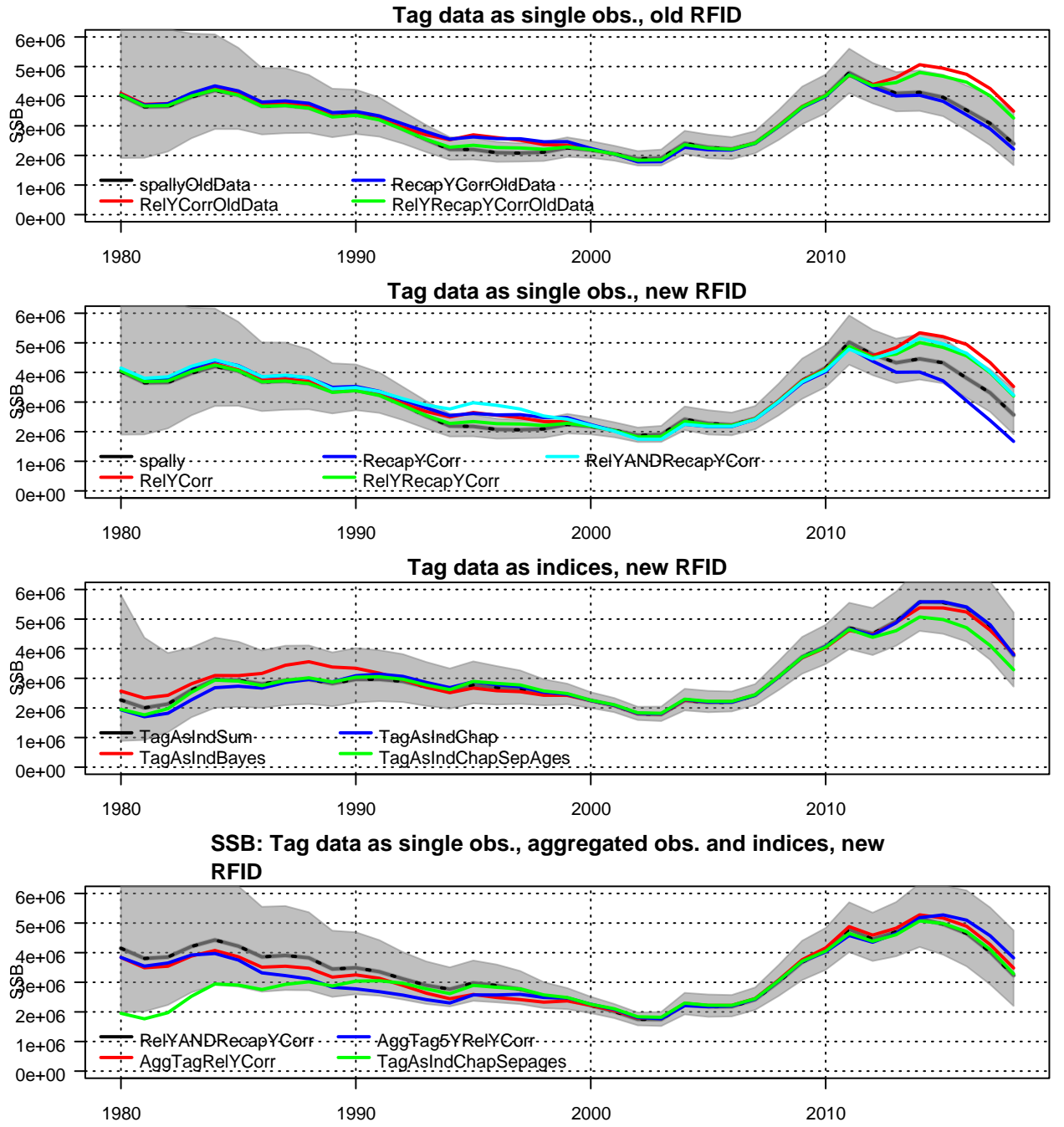


Figure 1: SSB estimates for various models.

2.6 Including correlations in catch data

As I stated in the Introduction, the exact model formulation is important when several data sources are included, and the relative weight of each data source depend on it. Therefore, one can argue that all data sources should be treated in an equally detailed level to give the correct weight to each data source.

In the models treated so far, correlations have been included in the survey index and in the tag data, but not in the catch data. To treat catch data in the same way as the other data sources, I have allowed for correlations between ages within years in the catch data, by using the irregular AR model implemented in SAM.

This gave an extreme improvement in the AIC values of nearly 800 in the models with the tag data used as single or aggregated observations. The corresponding correlations between ages became extremely high, in the magnitude 0.999. As a consequence, the catch data were down-weighted a lot compared to the other data sources, and the changes in the SSB estimates were large.

For the models where the tag data were used as indices, the AIC values were improved by nearly 200 for at least one model. The changes in the SSB estimates were less than for the models with raw tag data, but still large.

I find it worrying that treating the catch data in the same way as the other data sources has so dramatic consequences. Is there an explanation? Can we just ignore this?