ICES WGISDAA REPORT 2018

ICES CM 2018/ EOSG: 15

REF. ACOM, SCICOM

Report of the Working group on improving use of Survey Data for Assessment and Advice (WGISDAA)

3–5 July 2018

Copenhagen, Denmark

International Council for the Exploration of the Sea

Conseil International pour l'Exploration de la Mer

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

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

Recommended format for purposes of citation:

ICES. 2018. Report of the Working group on improving use of Survey Data for Assessment and Advice (WGISDAA). 3-5 July 2018. Copenhagen, Denmark. ICES CM 2018/EOSG: 15. 30 pp. https://doi.org/10.17895/ices.pub.8164

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.

The document is a report of an Expert Group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the views of the Council.

© 2018 International Council for the Exploration of the Sea

Contents

Executive summary **2**

1 Administrative details **3**

2 Terms of Reference **3**

3 Summary of Work plan **3**

4 List of Outcomes and Achievements of the WG in this delivery period **4**

5 Progress report on ToRs and workplan **4**

5.1 Investigating inconsistencies in the North Sea cod assessment 4

6 Divergent indices for whiting in Division 6a from the Q4 Scottish and Irish surveys **12**

6.1 The mackerel (and horse mackerel) egg surveys (MEGS) in the Northeast Atlantic and in the North Sea 18

6.2 The Mackerel and Horse Mackerel Survey in the Northeast Atlantic 19

6.3 The North Sea MEGS 24

6.4 Evaluation of the effect of potential design or sampling intensity changes in survey series 26

6.5 North Sea IBTS: Post-stratification options 28

7 Revisions to the work plan and justification **29**

8 Next meetings **29**

Annex 1: List of participants **30**

Executive summary

The Working Group on Improving use of Survey Data for Assessment and Advice (WGISDAA) met in Copenhagen, Denmark, 3–5 July 2018 for its first meeting in its three-year multi-annual cycle.

WGISDAA cooperated with assessment WGs on two ad-hoc questions on stock assessment issues related to survey data used in the cod (North Sea) and haddock (west of Scotland) assessments. The WG discussed causes and possible solutions to survey index divergence issues and made suggestions for further investigations for these specific cases.

WGISDAA cooperated with survey WGs on three survey design questions ranging for ways to evaluate methodologies for spatial interpolation and appropriateness of design for the mackerel-egg survey, appropriate post-stratification of the NS-IBTS survey for different objectives, and the effects of survey effort changes on the stock assessment uncertainty.

These topics are ongoing work and any conclusions will be reported to the cooperating WGs when they become available and will be fully reported at the end of the multi-annual cycle.

1 Administrative details

Working Group name
Working Group on Improving use of Survey Data for Assessment and Advice (WGISDAA)
Year of Appointment within current cycle
2018
Reporting year within current cycle (1, 2 or 3)
1
Chair(s)
Sven Kupschus, UK
Meeting dates
03–05 July 2018
Meeting venue
Copenhagen, Denmark

2 Terms of Reference

a) To work together with assessment working groups to provide resolution to assessment issues prioritized by the assessment working groups.

b) To work together with survey working groups to provide resolution to problems associated with index calculations, survey design changes (proposed or realized) to ensure efficient and effective use of survey resources.

c) Initiate with ACOM and secretariat a process to identify upcoming issues associated with the use of survey data in benchmarks. This should be initiated as soon as the benchmark process is started.

d) Review the output from the topic specific workshops initiated by WGISDAA and document more general principles learned from the specific cases dealt with in ToR a) and b).

3 Summary of Work plan

Year 1

Continue and update process eliciting advice requests from other elements of the ICES system; assessment, survey and benchmarking groups. Identify priorities within requests, and set up meeting and personnel accordingly. Prepare for topic specific workshops.

Year 2

Continue and update process eliciting advice requests from other elements of the ICES system; assessment, survey and benchmarking groups. Identify priorities within requests, and set up meeting and personnel accordingly. Review output from the topic specific workshops.

Year 3

As in year 2, plus appraisal of the success of the process, and make proposals for changes and any continuation

4 List of Outcomes and Achievements of the WG in this delivery period

First year of WG multi-annual period, currently no outcomes.

5 Progress report on ToRs and workplan

TOR a): WGNSSK had noticed some further issues with regards to the cod assessment with quite a strong revision of the cohort strength since last year and there seemed to be some conflict between the indices used in the assessment

5.1 Investigating inconsistencies in the North Sea cod assessment

The latest assessment of North Sea cod resulted in a downscaling of SSB in recent years. This is due to the addition of new data and thought to be primarily caused by lower than expected catch rates of older age classes in the IBTS Q3 in 2017 and the IBTS Q1 in 2018. This downward revision results in a reduction in the advised F due to SSB now falling below MSY $B_{trigger}$, consequently leading to a 47% reduction in the advised catch for 2019.

Exploratory work carried out by WGNSSK confirmed that revisions to the assessment were due to the addition of new data (catch 2017, IBTS Q3 2017 and IBTS Q1 2018) by showing minimal visual differences when rerunning the 2017 assessment with any revised time-series data truncated at 2017. Sequential removal of the new data showed the two survey indices to have the largest effect on the revisions to SSB and F. Work carried out by WGISDAA aimed to further investigate these revisions to the North Sea cod assessment.

Further data analyses

It was noted that the highest concentrations of older cod are found near the border of the assessment area towards the west of Scotland. It could therefore be hypothesized that migrations in and out of the assessment area are causing year effects in the survey indices. It was attempted to verify this hypothesis by combining NS-IBTS data with the SWC-IBTS survey data. However, the SWC-IBTS data from 2017 and 2018 (which was specific years of interest) were not available in the DATRAS database (neither under the "SWC-IBTS" nor the "SCOWCGFS" survey acronyms). Hence, the hypothesis could not be investigated at the time. Survey CPUE plots by year and age show the distribution of age 4 and 5 cod in the most recent survey period (Q3 2017–Q1 2018) to be similar to distributions in previous years (Figures 5.1.1–5.1.2), suggesting no significant changes to migrations in and out of the area. Furthermore, no clear outliers are apparent in these plots, with low catch rates widespread across the survey area, suggesting the results are not driven by a single observation or nation.

Figure 5.1.1. Distribution of cod ages 1–5 caught in the NS-IBTS Q1 survey 2014–2018.

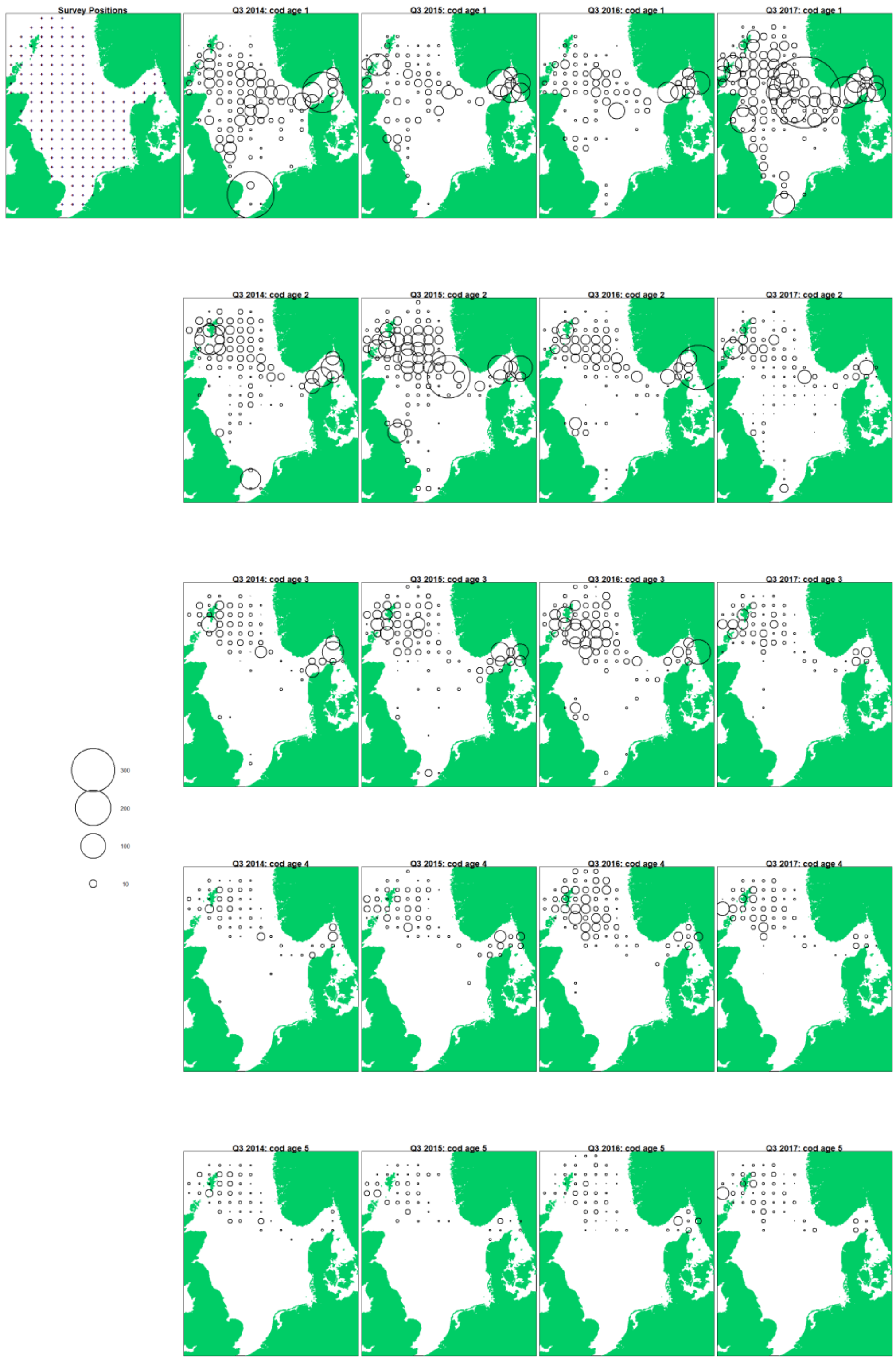


Figure 5.1.2. Distribution of cod ages 1–5 caught in the NS-IBTS Q3 survey 2014–2017. Note age 5 are not included in Q3 index.

Log-mean standardized plots of the survey indices indicate a year effect Q3 2017–Q1 2018, with most ages showing a decrease (Figure 5.1.3). It was highlighted that observed catch rates in the IBTS Q1 were low across all species in 2018 due to the severe weather conditions that year. The CPUE plots indicate that a year effect may have occurred during the previous season, Q3 2016–Q1 2017, as similar catch rates are observed when tracking cohorts from 2015 to 2016 in the IBTS Q3 (Figure 5.1.2) and from 2016 to 2017 in the IBTS Q1 (Figure 5.1.1), suggesting either very little mortality on cohorts or a year effect in the survey data.

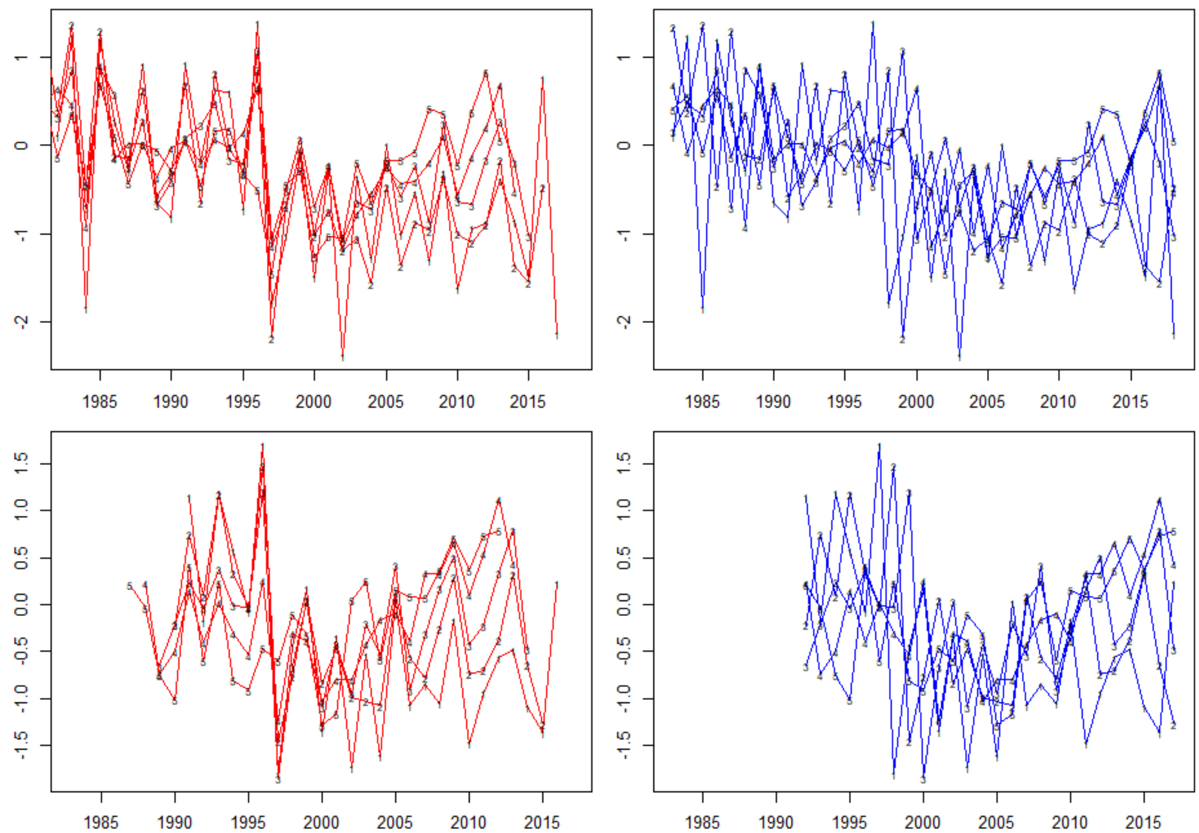


Figure 5.1.3. Log-mean standardized survey indices plotted by cohort (red) and year (blue) for the NS-IBTS Q1 (top) and NS-IBTS Q3 (bottom).

Runs of the SURBAR survey-based assessment model performed in 2017 and 2018 were analysed to further investigate possible year effects in the survey indices. Addition of the IBTS Q3 2017 and IBTS Q1 2018 data (i.e. going from the 2017 to the 2018 assessment) resulted in an upward revision of mortality and downward revision of SSB from 2015 (plots not shown). The residuals from the 2018 assessment suggest the IBTS Q1 2017 data are over-optimistic, with positive residuals for all age classes, and the IBTS Q3 2017 pessimistic, with negative residuals for all but age 1 (Figure 5.1.4). However, when examining the previous SURBAR assessment, performed in 2017 without the new IBTS data, the model fits the survey data Q3 2016–Q1 2017 reasonably well, with the residuals balanced around zero (Figure 5.1.4). These results suggest that year effects may be present for either or both periods Q3 2016–Q1 2017 and Q3 2017–Q1 2018 and that the SURBAR model is drastically changing its parameters to fit the survey residuals.

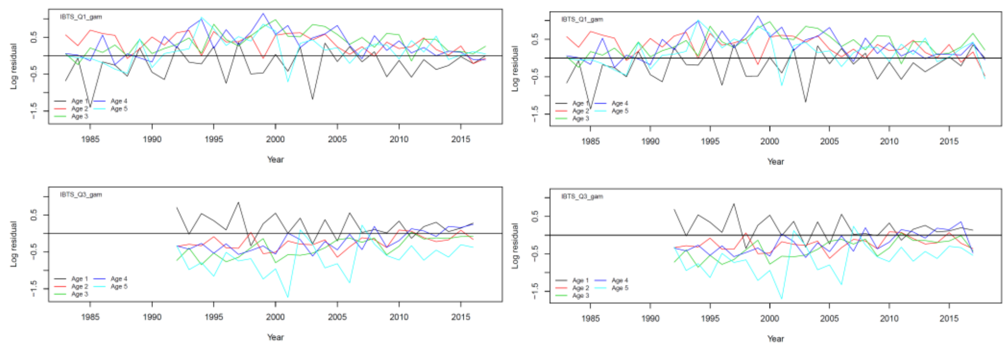


Figure 5.1.4. Residual plots from SURBAR runs performed in 2017 (left; without IBTS Q3 2017 and IBTS Q1 2018 data) and 2018 (right; with the new survey data).

The catch-at-age matrix, presented to WGNSSK as standardized proportions-at-age, shows no obvious differences in 2017 abundance compared to previous years catch. A separable VPA was performed as an additional check on the catch data (2006–2017; avoiding the prior period of uncertain catch levels) and presented no large step changes in fishing mortality.

Exploratory assessments

The accepted assessment of North Sea cod was rerun leaving out (1) the IBTS Q3 2016–Q1 2017 data (high catch rates) and (2) the IBTS Q3 2017–2018 data (low catch rates). Both options resulted in slight improvements to the residuals, with the first option producing a more pessimistic assessment and the second option a more optimistic assessment in line with that of 2017 (Figure 5.1.5).

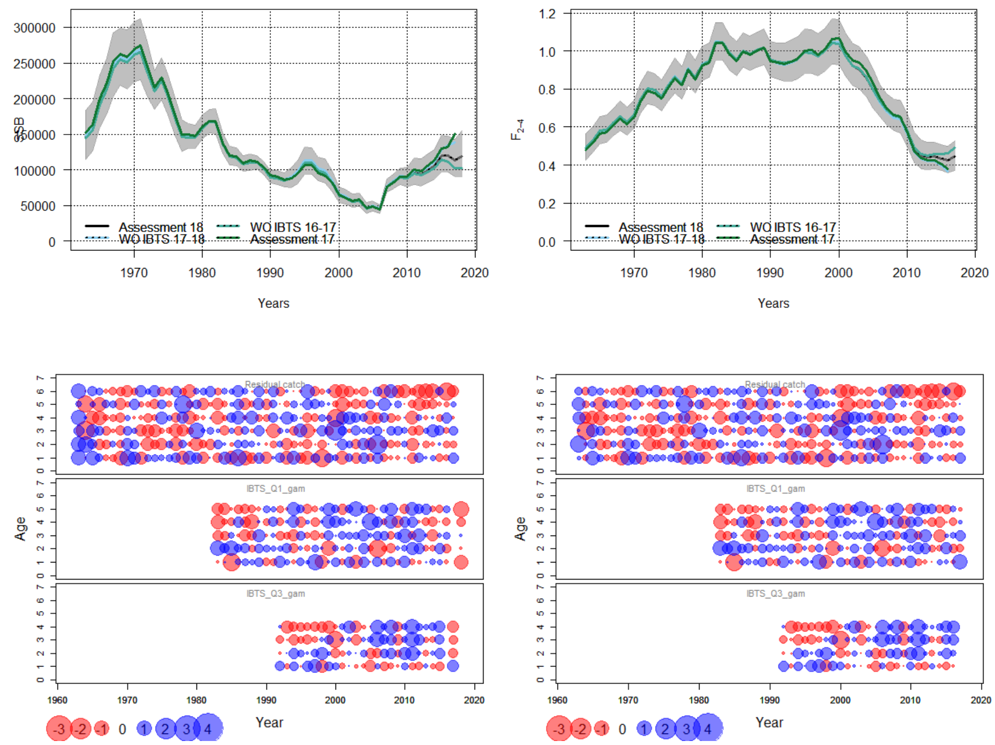


Figure 5.1.5: SSB, F (top) and observation residuals (bottom) from assessment runs without IBTS Q3 2016–Q1 2017 (left) or IBTS Q3 2017–Q1 2018 (right) survey data. SSBs and Fs from the accepted 2017 and 2018 assessments are plotted for comparison.

As data analyses indicate revisions to the assessment of North Sea cod are driven by year effects in the survey indices, an assessment that is more robust to year effects was explored. This was achieved by introducing correlated errors on the survey indices of older age classes (3+) and, although it made little improvement to the observation residuals, the effect was shown to be significant and resulted in a more parsimonious model in terms of AIC and negative log likelihood (Table 5.1.1). This exploratory assessment resulted in a further downward revision of SSB and upward revision of F (Figure 5.1.6).

Table 5.1.1: Model fitting diagnostics for an exploratory assessment run with correlated survey residuals and the accepted assessment for 2018.

Model	log(L)	No.par	AIC	Pval
Correlated survey residuals	-161.38	36	394.77	
Assessment 2018	-169.79	34	407.57	0.000225

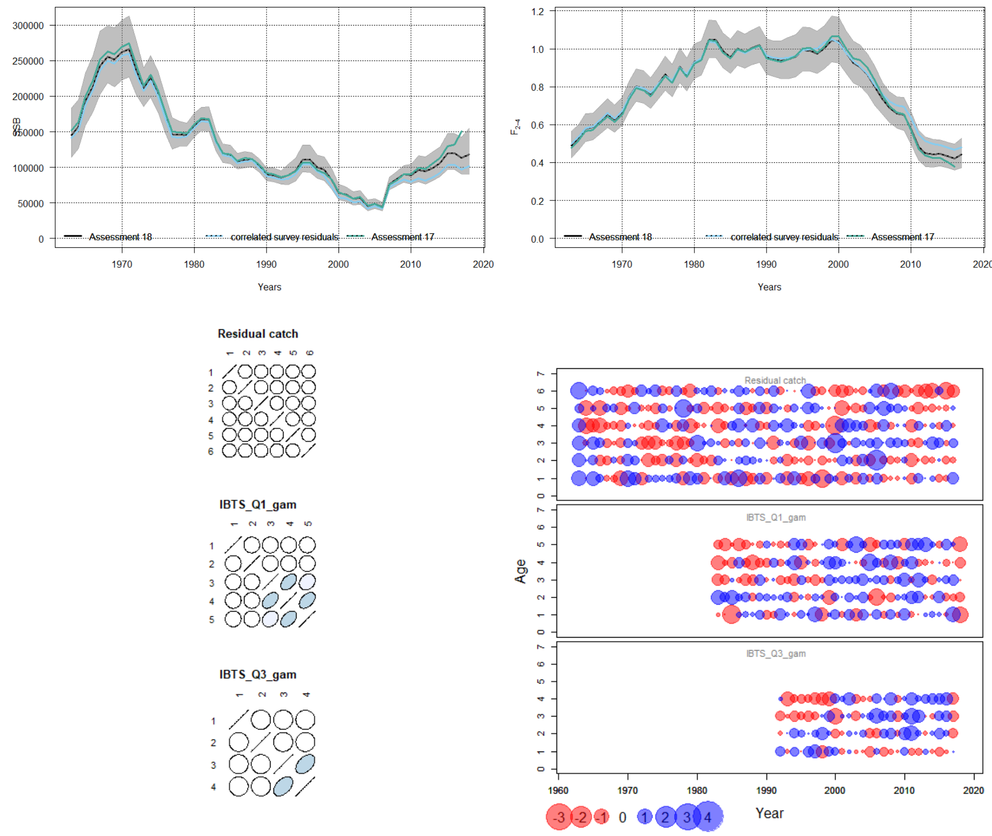


Figure 5.1.6: SSB, F (top), correlation matrices and observation residuals (bottom) for an assessment run with correlated errors on the survey indices. SSBs and Fs from the accepted 2017 and 2018 assessments are plotted for comparison.

A robust assessment was performed to check restrictiveness of smoothing in the accepted assessment. Mixture distributions consisting of 10% t_3-distribution and 90% normal distribution were assumed for log F, N and survey and catch observations, allowing for jumps in these processes and observations. There were no visual differences between the accepted and robust assessments, confirming that the Gaussian process and observation assumptions are not restrictive, and that no single observations appear to be driving the assessment.

Conclusions

Analyses conducted by WGISDAA find year effects in the recent survey indices. These year effects are likely caused by changes to the survey catchability, although the reasons behind these changes remain unresolved and could be due to several factors e.g. environmental conditions, changes to migrations or position in the water column potentially related to recent weather/sea conditions etc. This presents a problem for the assessment as the dome-shaped selection pattern leads the model to estimate a population of older fish that are not being seen in the survey. It is unclear whether this issue is due to sampling (i.e.

the fish are there but are not being caught in the survey) or modelling (i.e. the fish are not there and are not being removed from the model).

6 Divergent indices for whiting in Division 6a from the Q4 Scottish and Irish surveys

Introduction

The assessment of whiting in Division 6a is carried out yearly with catch and survey data (ICES, 2018). Currently, three surveys are in operation in the assessment area, two Scottish surveys (UK-SCOWCGFS-Q1 and UK-SCOWCGFS-Q4) and one Irish survey (IGFS-WIBTS-Q4). The data from these surveys are used to produce indices for the stock.

During the ICES Inter-Benchmark Protocol for West of Scotland Roundfish stocks (IBPWS-Round) in 2015, the option was considered of using one combined index for the two Q4 surveys (ICES, 2015). One rationale for combining the two indices was the fact that the Irish survey is mainly limited to the southern part of Division 6a (Figure 6.2.1, showing the most recent years, 2014–2017). It was hoped that a combined index could potentially provide a more robust and precise index. The comparative analysis conducted at that time for 2011–2014 (with the exclusion of 2013) suggested some differences in CPUE for whiting between the two surveys, the Irish survey tending to show higher catch rates. The common problem was high variability observed in the data which made the inference less plausible. Eventually, the IBPWS-Round concluded that the Q4 Scottish and Irish surveys should be retained as separate indices.

With more data collected in subsequent years and being available for analysis, the concept of using a combined index has been revived recently. This issue became important in view of the next benchmark for the stock being planned within the next 1–2 years.

Data and analysis

For the GAM analysis, haul data collected in the period 2011–2017 (excluding 2013) were used. Hauls were selected that were taken in the southern part of Division 6a. In that area, roughly, both surveys were operating and it is referred to here as the "common area". In 2013, the Scottish survey was not fully implemented and it covered only the northern half of Division 6a. This resulted in 391 valid hauls that were subsequently subject to modelling (Figure 6.2.2). The haul data included, among others, year, vessel (two vessels were operating in the area at that time), geographical position (longitude and latitude), depth (m), day (ordinal day of the year), time (Coordinated Universal Time, UTC), haul number and tow duration (in minutes). The variable 'depth' was log-transformed to achieve a more even spread of data along the depth gradient (Wood, 2006).

A statistical model (a negative binomial GAM for counts with a log link function) was used to estimate catch. The dispersion parameter, k, for the negative binomial was found during the model optimization within a range of likely values for this parameter (0.1–1.3). This was run separately for age groups (0, …, 6 and 7+), as well as for an additional aggregate age group (1+). The age-0 group were omitted in this aggregation as they usually show distinct characteristics before the settling transition.

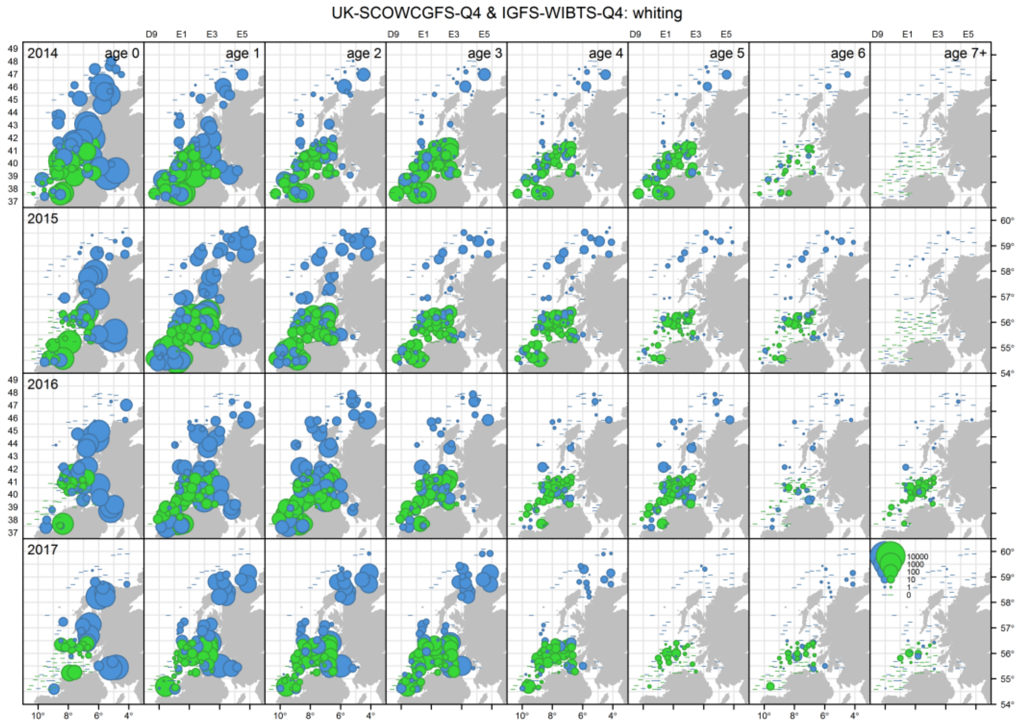


Figure 6.2.1. CPUE from the two Q4 groundfish surveys on the West Coast, Scottish (UK-SCOWCGFS-Q4, in blue) and Irish (IGFS-WIBTS-Q4, in green), in 2014–2017. Numbers are standardized to 30 minutes towing.

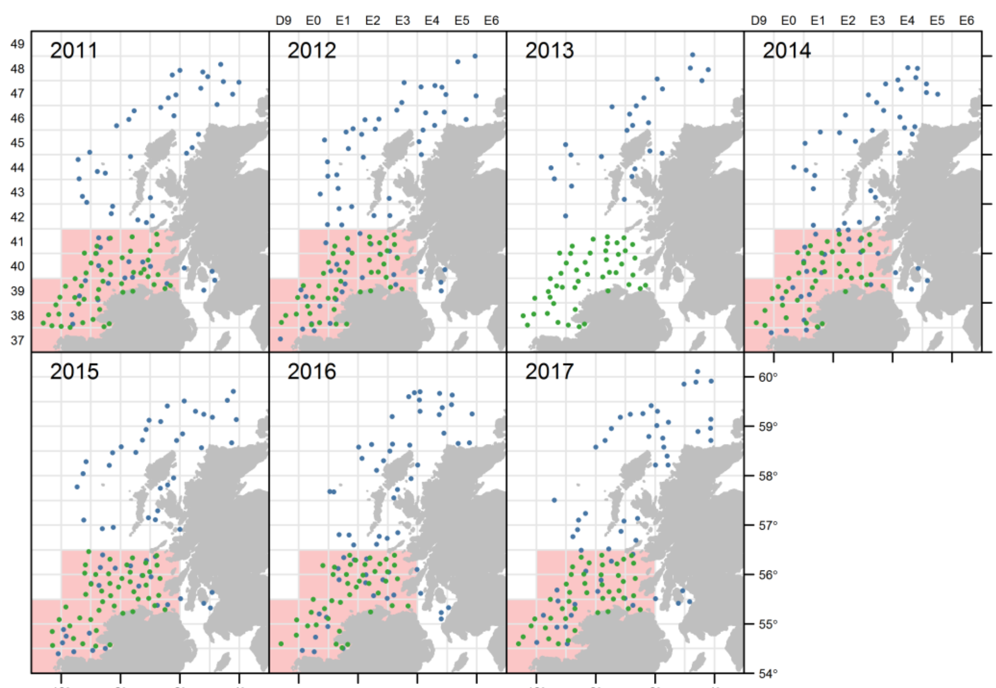


Figure 6.2.2. The location of hauls in the two Q4 surveys: UK-SCOWCGFS-Q4 (in blue) and IGFS-WIBTS-Q4 (in green), in 2011–2017. The shaded red area marks the statistical rectangles where both surveys were operating ("common area").

The raised numbers (N_i) were the response variable in the model. To account for the different tow duration, it was given as a proportion of the standard tow duration of 30 minutes (thus, the catch rate was expressed as the number of fish per standard tow rather than as the number of fish per minute). The tow duration term was log-transformed and added as an offset to the model (Zuur *et al.*, 2009).

In the first approach, a full model was considered that included vessel, year, day, time, depth, and longitude and latitude:

$$N_i \sim NB(\mu_i, k)$$

$$E(N_i) = \mu_i = e^{g(x_i)} \qquad \text{and} \qquad \mathrm{var}(N_i) = \mu_i + \frac{\mu_i^2}{k}$$

where

$$g(x_i) = \log(Td_i / 30) + \alpha + \beta_1 \times Vessel_i + \beta_2 \times Year_i + f_1(Day_i) + f_2(Time_i) + f_3(\log(Depth_i)) + f_4(Lon_i, Lat_i)$$

μ_i is the mean and k is the dispersion parameter in the negative binomial distribution; $g(x_i)$ is the link function; Td_i is the tow duration; α is the intercept; β_1 and β_2 are parameters to be estimated; f_1 is the smoothing function of Day_i; f_2 is the smoothing function of $Time_i$; f_3 is the smoothing function of the log-transformed depth, $Depth_i$; f_4 is the smoothing function of the interaction of longitude (Lon_i) and latitude (Lat_i) in the ith sampling location.

Smoothing parameter estimation for the model maximized likelihood (delivering maximum likelihood, ML; Wood, 2006).

In the next step, a Chi-square test was conducted with the above model and the reduced model (without the 'vessel' variable) to establish how significant the difference between the two surveys was. In addition, confidence intervals for the estimated difference were constructed by applying the bootstrap.

Results

Some variability in the whiting CPUE, with the observed data, was observed between the two surveys in the "common area". In some years and for some age groups (especially for fish at age 3–5, and most consistently for age 4), the CPUE tended to be higher in the Irish survey (Figure 6.2.3). Apart from this bias, the overall trends were rather similar in the two surveys for all age groups. Also, no considerable difference was observed for the aggregate 1+ age group.

An examination of the observed length distributions showed a shift between the Irish and Scottish surveys (Figure 6.2.4). It was most pronounced for the smallest fish. For the bigger fish (over 20 cm in length), it was notable only in some years (in 2011 and 2012). In general, the length distributions were rather similar in shape. In 2014, the difference in catch rate between the two surveys was considerably bigger for most sizes.

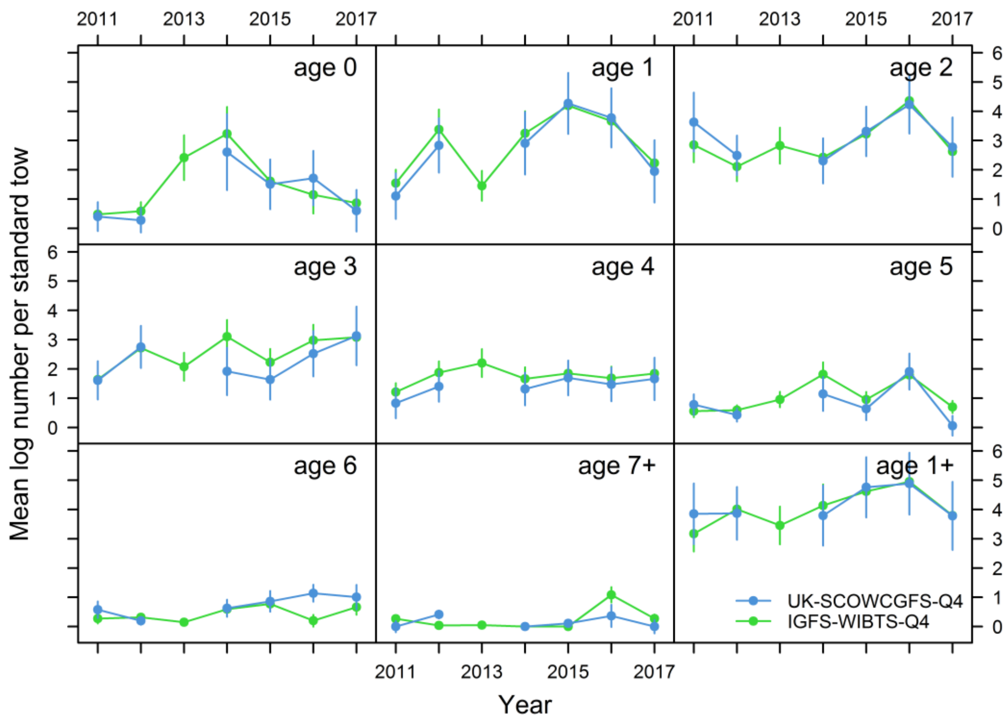


Figure 6.2.3. The observed CPUE of whiting (log values with 95 % confidence limits) by age in the two surveys in the "common area".

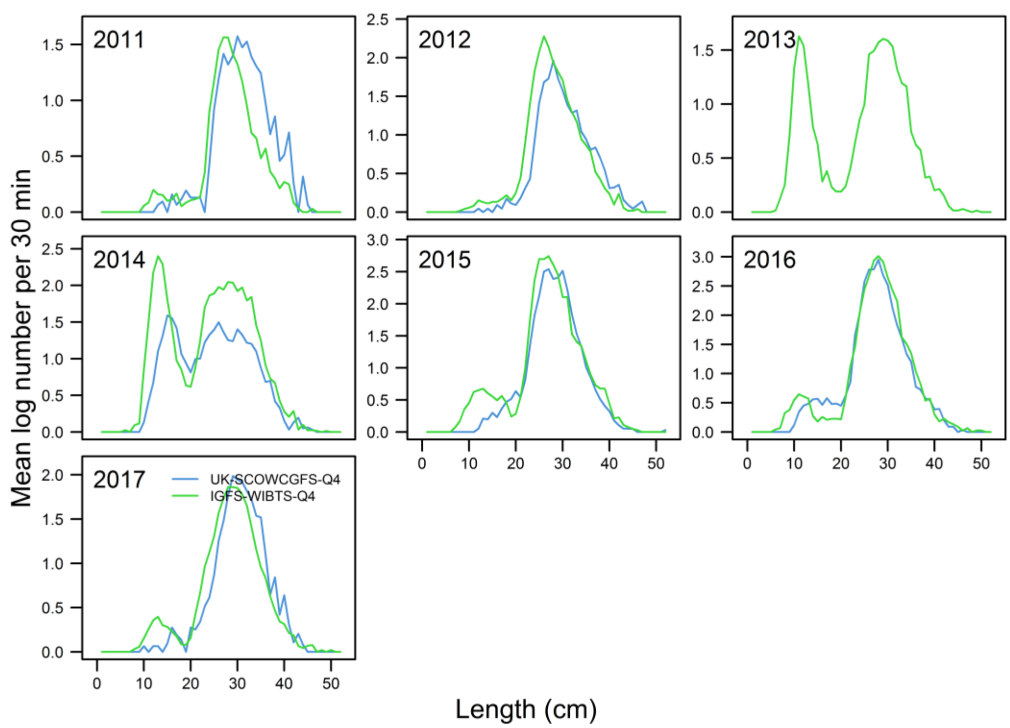


Figure 6.2.4. The length frequency in the two surveys in the "common area".

The GAM analysis provided more insight into the difference between the two surveys. The dispersion parameter in the negative binomial distribution varied within a wide range from 0.27 to 0.80 (Table 6.2.1). Overall, the 0-group tended to form aggregations which is demonstrated by a low value of the dispersion parameter. This tendency weakened with age. Depth and location (within the "common area") were important (and highly significant) explanatory variables in the model, explaining a large amount of the variation observed in the data. The spatial distribution varied among the different age groups, but the common feature was that whiting densities gradually decreased along the western edge of the study area (not shown). The effect of the two other explanatory variables, day of the year and time of the day, was less pronounced. Despite the different timing of the two surveys (the Scottish survey usually followed the Irish survey), it appeared to be of less importance. The catch rate slightly changed during the day with a minimum in the afternoon/evening hours, but this tendency was not observed across all ages, not in the least for the 0-group.

The effect of vessel was quite variable across the different age groups. While the catch rates tended to be higher in the Irish survey for most age groups, for fish at age 6, they were significantly higher in the Scottish survey (see the vessel effect in Table 1, Figure 6.2.5). In relative terms, the ratio of catch rates in the Scottish survey to those in the Irish survey was 0.6–1.8, depending on the age group. For the aggregate age group 1+, it was 0.8. The model explained 40–75 % of the deviance, with decreasing (with fish age) ability of the model to explain the variation in the data.

Table 6.2.1. Summary of the GAM analysis

Age group	Dispersion parameter	*Day* significance	*Time* significance	log(*Depth*) significance	*Lon* × *Lat* significance	*Vessel* effect	*Vessel* effect significance	exp(*Vessel* effect)	Deviance explained (%)
0	0.265	0.227	1.000	<0.001	<0.001	-0.102	1.000	0.90	75.1
1	0.526	0.111	0.128	<0.001	<0.001	-0.360	0.762	0.70	66.9
2	0.620	0.076	0.071	<0.001	<0.001	0.157	0.997	1.17	62.3
3	0.655	0.006	0.037	<0.001	<0.001	-0.552	0.105	0.58	58.4
4	0.799	0.117	0.261	<0.001	<0.001	-0.551	0.014	0.58	51.9
5	0.760	0.107	0.008	<0.001	<0.001	-0.530	0.312	0.59	56.3
6	0.798	0.055	0.375	<0.001	<0.001	0.580	0.115	1.79	39.7
7+	0.730	0.205	0.024	<0.001	<0.001	-0.575	0.424	0.56	56.8
1+	0.617	0.055	0.116	<0.001	<0.001	-0.202	0.995	0.82	61.4

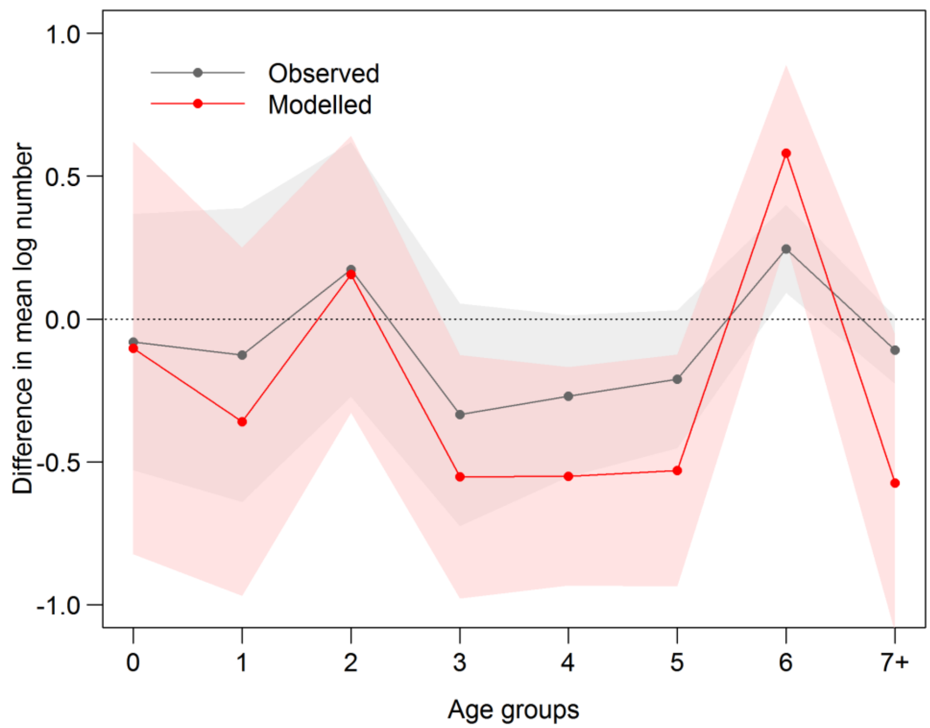


Figure 6.2.5. The observed and modelled difference in CPUE of whiting (with 95 % confidence limits) by age between the two surveys in the "common area".

Discussion

This preliminary analysis demonstrates that there were some differences between the two Q4 surveys in terms of catchability. While the observed differences were previously considered to result from the different survey timing (IBPWSRound; ICES, 2015), the GAM analysis showed that this factor was less important or it was not consistent over years.

It is unclear why the vessel effect was variable among the different age groups. Given that all the important explanatory variables were included in the model (among them, the year effect), it was expected that the modelled differences among age groups would form a more regular pattern. It is likely that more data would be needed to ascertain such a long term pattern.

The IBPWSRound (ICES, 2015) considered the possibility of estimating the relative catchability in the two surveys in the common area and then adjust the CPUE in one survey. One of the outcomes of the present analysis was an attempt to quantify the as for the different age groups. Provided that such a long-term bias is established and quantified, it could subsequently be used to standardize the two surveys.

Further exploration is needed to establish the origin of the differences between the two surveys. A re-formulation of the model may provide more insight into the data structure and the real differences (for example, by including more interactions). More data being available within the next 1–2 years will provide more ground to support the inference of the difference.

As the final conclusion, despite some differences observed for whiting in the two Q4 surveys, using a combined index offers a plausible alternative to using two separate indices, especially if a more in-depth exploration, possibly with more data, follows the present analysis.

References

ICES. 2015. Report of the Inter-Benchmark Protocol of West of Scotland Roundfish (IBPWSRound), February–April 2015, By correspondence. ICES CM 2015/ACOM:37. 72 pp.

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

Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC, Boca Raton, Florida, 382 pp.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., and Smith, G. M. 2009. Mixed Effects Models and Extensions in Ecology with R. Springer, New York, 574 pp.

TOR b): A number of participants brought questions on survey index calculation and survey evaluation questions that the group started to address. Most of the work done at the WG focused on developing analyses that could be applied to the available data to address the question. Such work will be implemented intersessionally and reported back to WGISDAA next year.

6.1 The mackerel (and horse mackerel) egg surveys (MEGS) in the Northeast Atlantic and in the North Sea

The Northeast-Atlantic mackerel stock consists of 3 spawning stock components: the western, the southern and the North Sea component. There are currently 2 surveys in place that deliver independent SSB abundance indices:

The Mackerel and Horse Mackerel Egg Survey in the North East Atlantic and the Mackerel Egg Survey in the North Sea. Both surveys aim at the annual egg production of mackerel, and in the case of the Northeast Atlantic Survey also of horse mackerel, to produce a relative index for SSB estimation in the target species. Results of the Northeast Atlantic survey are used in the assessment for both species, mackerel and horse mackerel. Results of the North Sea survey, however, are not used in the assessment because the contribution of that component is currently only around 4 % of the total SSB (ICES 2017a). Nevertheless, the

North Sea mackerel egg survey is still considered useful by the corresponding assessment working group WGWIDE in order to monitor the status of the formerly much more important North Sea component (see e.g. Jansen 2014).

Both surveys face recurring problems with survey execution, which have the potential to considerably increase the uncertainty in the abundance index calculation for all target species.

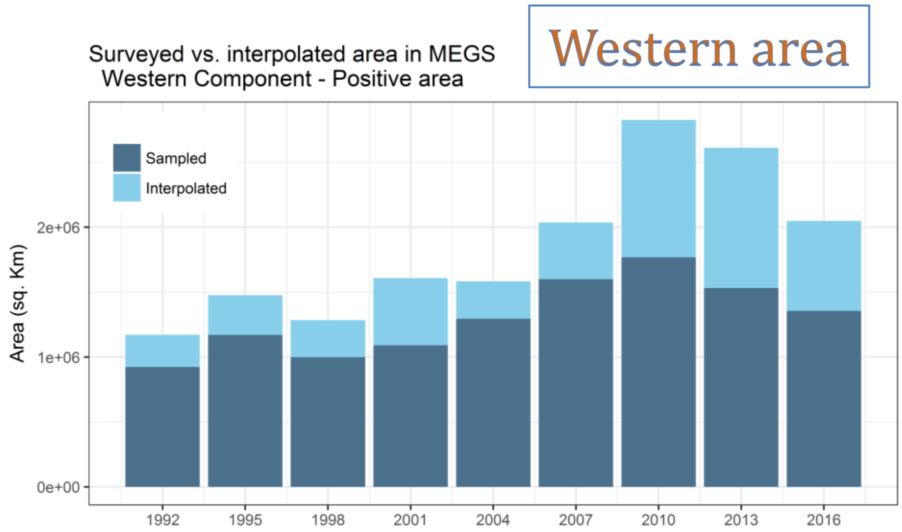
6.2 The Mackerel and Horse Mackerel Survey in the Northeast Atlantic

The MEGS is carried out triennially since 1977 and delivers the only fishery-independent data for the assessment of Northeast Atlantic mackerel and horse mackerel. Total annual egg production (TAEP) is calculated from counts of freshly spawned eggs taken from tows with Gulf VII type samplers. Plankton samples are taken on stations on predefined zonal transects every full half degree latitude using the alternate transect strategy. In addition, survey participants are requested to follow an adaptive strategy while following the transects, i.e. each transect should only be finished after encountering 2 consecutive zero counts of freshly spawned mackerel eggs. TAEP is then calculated from stage 1 egg abundance data. With the fecundity values estimated during the same survey, the TAEP of mackerel is then converted into an SSB value for mackerel, which is used a relative index in the assessment. For horse mackerel, the TAEP is used directly as a relative index for SSB in the assessment (for more details see first interim report of the previous WGISDAA term, ICES 2015).

Starting in 2007, an extension of the spawning area together with a forward shifting of peak spawning had been observed until 2013. Though new survey participants could be recruited, the vast expansion of both spawning time and area necessitated leaving out of every other transects in some of the survey periods in order to achieve a full coverage of the survey area. Additionally, it also became increasingly difficult to represent the annual egg production for both target species of the survey, mackerel and horse mackerel, as their time of peak spawning appears to drift further apart. This raised concern that the traditional survey design will not be able to provide reliable and defendable estimates of TAEP for mackerel and horse mackerel. These problems were thoroughly discussed during the recent term of WGISDAA (2015 – 2017) and recommendations of the working group had already been implemented by WGMEGS:

- To estimate the contribution of northwestern spawning extension to TAEP, which would indicate areas where effort savings could be made that have minimal impact on the precision and accuracy of the index.
- And to replace the double zero rule for transect termination by a more meaningful one.

The latter recommendation gained recognition the MEGS manual, which is currently under revision. For the 2007 to 2013 MEGS, it was estimated that the contribution of the northwestern expansion of the spawning area to the total annual egg production was low and stable (ICES 2015), while the core area further south appeared to be more temporally variable and larger in absolute terms. This suggested that thinning out sampling in the expansion area for the benefit of concentrating sampling in the traditional mackerel spawning core area would not greatly bias the survey results. However, in the 2016 survey it appeared that spawning peak had returned to the traditionally later period in May and happened just in the northwesterly expanded area. Again, estimation of TAEP relied heavily on between transect interpolation (ICES 2017b). An inspection of the MEGS time-series showed that the area, where interpolation had to be applied, did indeed increase for particularly the 3 recent surveys (Figure 6.3.1) to up to more than 30 % of the survey area. However, it also showed that in the western component interpolated egg abundances always contributed more than 10 % to the TAEP estimation (Figure 6.3.2). In the southern stock component, the contribution of interpolation was not as imminent.



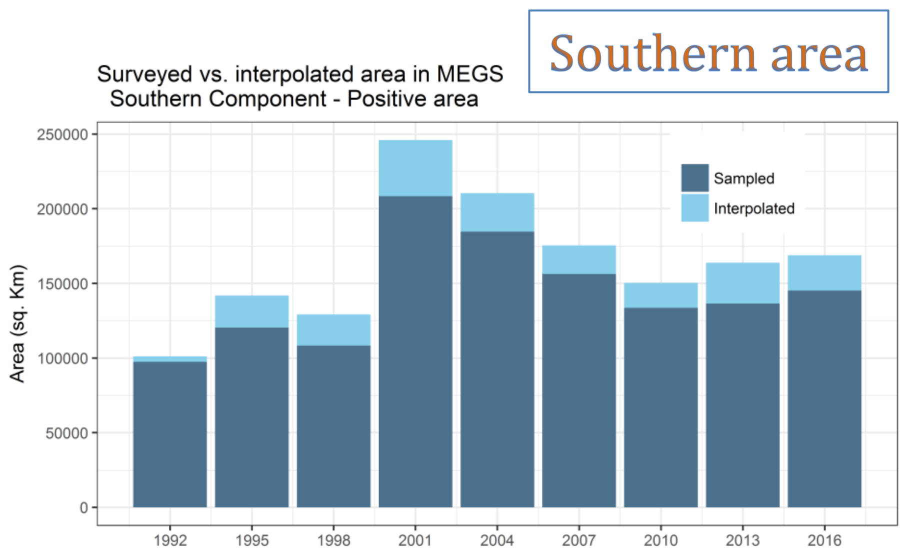


Figure 6.3.1 Mackerel. Prospected and interpolated area ratio within the area where mackerel eggs were found for the Western (top) and Southern (bottom) mackerel spawning stock components for the MEGS time-series. Dark blue: actually sampled area, light blue: interpolated area during the surveys.

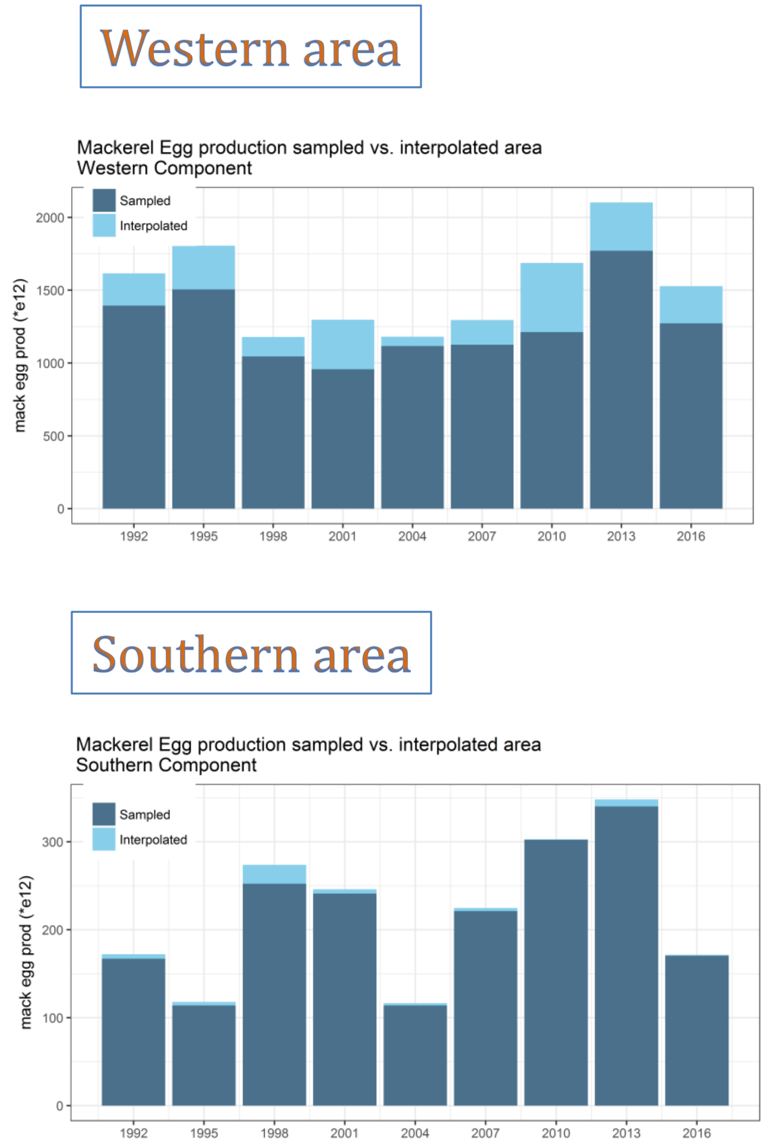
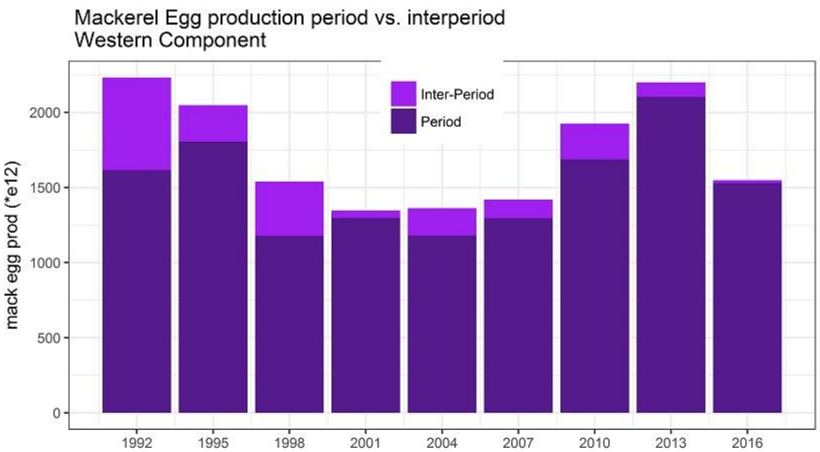


Figure 6.3.2 Effect of spatial interpolation on mackerel daily egg production estimates for Western (top) and Southern (bottom) mackerel spawning stock components for the MEGS time-series. Dark blue: actually sampled egg production, light blue: interpolated egg production during the surveys.

Besides spatial interpolation, TAEP estimates also have to rely on temporal interpolation in cases when survey times do not entirely cover the a-priory defined spawning periods (Figure 6.3.3). Temporal interpolation was particularly imminent in the surveys for the southern spawning component.

Western area



Southern area

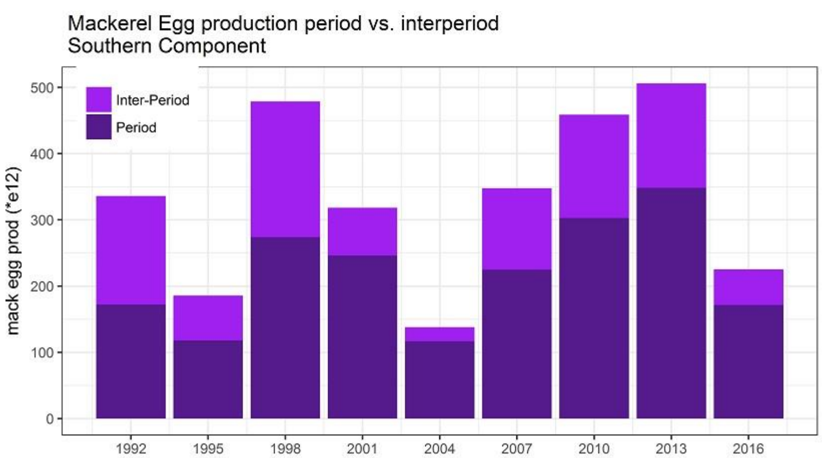


Figure 6.3.3 Effect of temporal interpolation on annual egg production of mackerel for Western (top) and Southern (bottom) mackerel spawning stock components for the MEGS time-series. Dark purple: means egg production in sampling period, light purple: egg production in interpolated period (interperiods).

It is evident, that interpolation is widely used in calculating TAEP in both mackerel and horse mackerel (see also ICES 2018). The manual for the MEGS only stipulates linear interpolation methods for both, spatial and temporal interpolation, disregarding the high spatial and temporal variability that may occur in egg abundance (see e.g. Kloppmann et al. 2012). Already during its 2015 meeting, WGISDAA recommended that WGMEGS should consider investigating temporal and spatial variability of mackerel and horse mackerel egg production in order to design an unbiased estimation TAEP in the target species. For their current term, which started in 2018, WGMEGS will pursue this recommendation and has asked WGISDAA for support. Results of the exercise will be presented and discussed during the following WGISDAA meetings.

6.3 The North Sea MEGS

The North Sea mackerel stock component is currently very small, which is why the results from the egg survey are not used for the assessment. However, the North Sea MEGS is still considered useful by WGWIDE because it is the only survey, which delivers information on the status of the stock component. Like the Northeast Atlantic MEGS, the North Sea MEGS is carried out triennially and is normally carried out 1 year after the Northeast Atlantic survey. In the past, it was conducted by Norway and The Netherlands, but after the 2011 North Sea MEGS, Norway withdrew from the survey leaving The Netherlands as the only responsible participant.

The survey suffered substantially from the withdrawal. Not only was there a reduction in the coverage of the mackerel spawning season in the North Sea, but also there occurred inconsistencies in the calculation of the TAEP rendering the recent 2 estimates unreliable for the North Sea MEGS. This becomes particularly apparent for the 2015 North Sea MEGS (Figure 6.3.4 and table 6.3.1). The egg production curve is almost over its entire course below the curves of the 2 previous surveys but still delivers a higher TAEP estimate. Also, the area under the 2017 curve does not appear to be particularly higher than for the 2005 one, but the 2017 TAEP is considerably increased compared to the 2005 TAEP.

The problems of that survey were discussed at the WGISDAA meeting and the group concluded that before all necessary analyses and adjustments for a continuation of the survey are carried out, the respective working groups that either plan or utilize the egg surveys (WGMEGS and WGWIDE) should decide on the priorities of the egg surveys, in particular on the usefulness of the North Sea mackerel abundance index. If the North Sea MEGS is still considered useful and necessary for the assessment of the Northeast Atlantic mackerel stock, the analysis of the time-series should be carried out along the lines WGISDAA also recommended for the Northeast Atlantic survey.

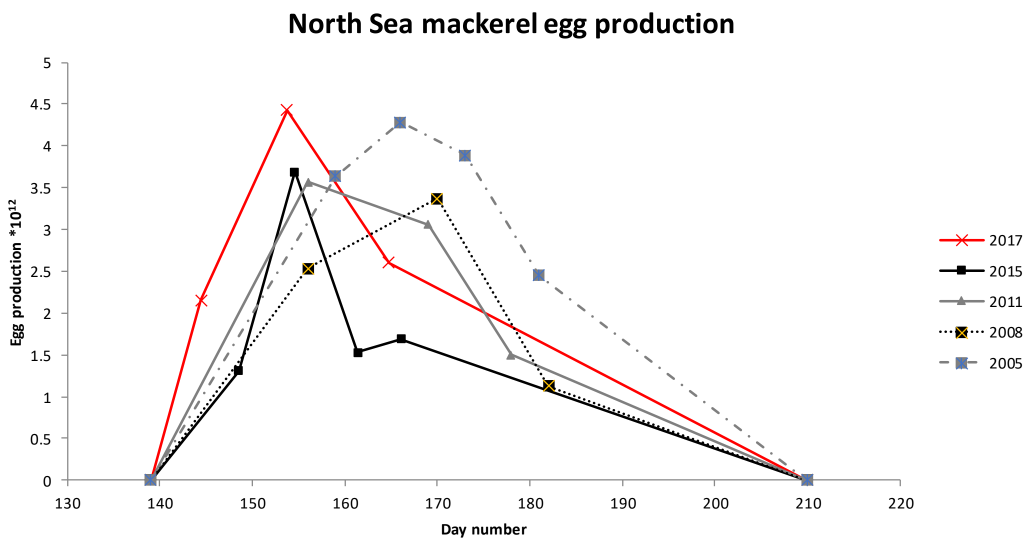


Figure 6.3.4 Annual egg production curves for North Sea mackerel (prior to 2015 the Lockwood egg development equation was used, since 2015 the Mendiola equation was used).

Table 6.3.1 Egg production estimates from egg surveys 2005 – 2017 in the North Sea and corresponding SSB based on a standard fecundity of 1401 eggs/g/female.

Year	Egg prod $*10^{12}$	SSB $*10^3$ tons
2005	155	223
2008	108	154
2011	116	165
2015	119	170
2017	201	287

During its recent meeting WGMEGS decided to utilize the daily egg production method (DEPM) for the coming surveys, because this would only require one full coverage of the spawning area over a shorter time period (ICES 2018). The remaining ship time plus a possible Danish participation would potentially also enable to collect the required large number of adult samples for fecundity analysis. However, because the inconsistencies in the 2 recent surveys need to be eliminated and in order to get the AEPM time-series in line with the newly created DEPM time-series, the existing data need to be thoroughly analysed. As for the Northeast Atlantic survey, WGMEGS asked for support in this process through WGISDAA.

The problems of that survey were discussed at the WGISDAA meeting and the group concluded that before all necessary analyses and adjustments for a continuation of the survey are carried out, the respective working groups that either plan or utilize the egg surveys (WGMEGS and WGWIDE) should decide on the priorities of the egg surveys, in particular on the usefulness of the North Sea mackerel abundance index. If the North Sea MEGS is still considered useful and necessary for the assessment of the Northeast Atlantic mackerel stock, the analysis of the time-series should be carried out along the lines WGISDAA also recommended for the Northeast Atlantic survey.

References

ICES 2015: First Interim Report of the Working Group on Improving use of Survey Data for Assessment and Advice (WGISDAA). ICES CM 2015/SSGIEOM:28

ICES 2017a: Mackerel (Scomber scombrus) in subareas 1–8 and 14, and in Division 9.a (the Northeast Atlantic and adjacent waters). ICES Advice on fishing opportunities, catch, and effort. Ecoregions in the Northeast Atlantic and Arctic Ocean, mac.27.nea. Published 29 September 2017 DOI: 0.17895/ices.pub.3023

ICES 2017b: Final Report of the Working Group on Mackerel and Horse Mackerel Egg Surveys. ICESCM 2017/SSGIEOM:18

ICES 2018: Report of the Working Group on Mackerel and Horse Mackerel Egg Surveys (WGMEGS). ICESCM 2018/EOSG: 17

Jansen, T. 2014: Pseudocollapse and rebuilding of North Sea mackerel (*Scomber scombrus).* ICES Journal of Marine Science, 71(2), 299–307. doi:10.1093/icesjms/fst148

Kloppmann M. et al. 2012: Enhancing survey coverage with no net increase in survey effort. ICESCM 2012/F:11

6.4 Evaluation of the effect of potential design or sampling intensity changes in survey series

Cefas is developing a framework by which the effect of potential changes to sampling and survey design can be evaluated based on existing knowledge. This should be developed into an open form structure, but it seemed sensible to start developing this using some specific examples as proof of concept. IBTSWG has for some time now considered that a reduction in survey effort is required to ensure full coverage of the NS ecosystem as well as building some contingency into the monitoring system. Opinions differ as to whether shortening the 30-minute tow duration to 15 minutes adds significant uncertainty to the derived indicators used in single species assessments and MSFD indicators has varied widely. Up to now the outputs have been compared only at the level of the developed survey index, but for stock assessments using IBTS data at least it is possible to evaluate the impact at the level of the stock assessment metrics used in management.

Over the coming years WGISDAA aims to develop a comparison of the effect of two methods (100% of stations at 50% tow duration; 50% of stations 100% of tow duration of reducing the effective sampling effort in both NS-IBTS surveys (Q1 and Q3) on three gadoid assessments (Cod, Haddock and Whiting) using two index calculation methodologies (model- and design-based).

This year's work focused on discussing appropriate methods for subsampling the nested sampling designs (survey stations and biological samples given catches at a station) to implement as far as possible a realistic simulation methodology for all three stocks and the two types of indices.

50% of stations 100% of tow duration:

Through the use of "rectangle" strata, the IBTS survey ensures relatively even coverage over the extent of the NS. Such a systematic design has benefits in reducing bias when species distributions change, but with usually only 2 samples within a stratum in a year, this makes subsampling according to the strata difficult. Subsampling stations randomly within a year belies our understanding of the persistent patterns of spatial distribution of the three species encountered. An independent means of subsampling should be used that reflects patterns of ecological communities throughout the NS.

There was some debate then if the strata chosen for the post stratification should also then be used to calculate the design-based indices (ICES-indices), or whether for consistency of the variance structure in indices with existing index calculations the historic "roundfish areas" mainly used in the production of ALKs should be maintained.

Current otolith subsampling is conducted at the haul and length level for the species considered, but historically otolith targets have been at the roundfish area level so that all otoliths could come from a single rectangle so that historic data may underestimate the otolith that would have been collected if todays subsampling protocols were followed. Some form of supplementation may be necessary for some species to get reasonable estimates of variance out of the analysis. However, it is not clear how to do this appropriately and it is not thought to be a major issue for the species examined here.

100% of stations at 50% tow duration

All stations are used in all simulations so that all procedures for index calculations are implemented as currently implemented for both the design- and model-based indices. However, it is necessary to subsample the haul information at the length level to simulate a reduction in sampling effort. Simply halving the size of catches is unlikely to reflect the true error structure of a 15-minute tow. For this to be done more realistically it is necessary to calculate a binomial estimate of catches at each length (equivalent to selecting each individual caught with a mean probability of 50%) and summing the selected individuals by length over the sample.

An independent application of the same methodology cannot be applied to biological sample information as this would lead to a mismatch of lengths and age and one might have sampled different individuals for ages if those that were sampled had not been caught. To ensure consistency between the sampling approaches then we determined the number at lengths caught in an individual station in the simulated sample and retained all otoliths from that station where the number of lengths was greater or equal to the number of otoliths. Where the number of otoliths at length was greater than the number of lengths a random subsample was taken equal to the number of lengths observed in the simulation.

Theoretically at least the linear interpolation of the probabilities as modelled here through the 50% binomial is unlikely. There are edge effects such as an offset for the "starting catch" (fish caught before the start of the formal tow) and fish exhaustion (large fish tire more slowly so could be more frequently caught towards the end of the two than at the start). These are biases that are currently not quantifiable and therefore cannot be included in these simulations. Therefore, the proposed simulations represent a best-case scenario of the effects rather than a precise prediction.

The proposed simulations sample individual fish as if they were randomly distributed over the haul. However, it is likely that there is autocorrelation within and across length groups. Such contagion is unlikely to influence the central tendencies of the length distributions but could affect the availability of biological samples so that the impact on the index is not easily described. At the extreme condition, either all the fish or none of the fish are caught. It is simple to see that at that stage the data would roughly subsume to the station simulations with a halving of the "presence absence model" in the Delta-Gam model. The overall uncertainty would increase relative to the proportion of the presence absence model, but the central tendency would be unaffected. For the design-based indices variance is also likely to increase substantially compared to the "random" choice and the error distribution is likely to be more overdispersed. At present no information exists as to the magnitude of the contagion making appropriate modelling difficult. The scenario tested is a best-case scenario.

Conclusion:

The following two scenarios will be implemented in the code to evaluate the effect on stock assessment derived management metrics of changes to the survey design.

Scenario1: Samples are randomly selected stratified by ecological strata, and only the corresponding age data are used in the ALK calculations. Design based index calculations are carried out using the ICES protocols for ALK area substitutions and filling.

Scenario2: All hauls are used in the calculations consistent with the IBTS sampling design. Each haul is subsampled at the length level using a binomial distribution for each length and only otoliths to up to the number of lengths subsampled are retained in the analysis. Design based index calculations are carried out using the ICES protocols for ALK area substitutions and filling.

6.5 North Sea IBTS: Post-stratification options

Post-stratification can be applied for different purposes, one of them being the use of existing knowledge for the development of an improved survey design. Particularly in cases where the possibilities of a random stratified sampling design are to be explored, post-stratification is an informative approach.

The group discussed various options to investigate the effects of a post-stratification for the North Sea IBTS. Different methodological aspects were highlighted. In this, statistical analyses were presented, which looked at the effect of non-normality in the distribution of species-specific CPUE data, using bootstrapping and definition of uncertainty through quantiles, instead of estimates of uncertainty based on confidence intervals calculated from standard errors. While this took care of the error distribution in the raw data, it did not suffice to take the annual variation in the tested 10-yr dataset into account. Therefore, the group recommended that other methods should be tried for comparison. Specifically, modelling should be able to separate effects, which are due to year effects on the one hand, and due to the choice of strata on the other. Generalized linear models appear to be appropriate to serve this purpose, allowing for an analysis of variance, which can be used to compare alternative models.

Results from an analysis of the effects of sample size in IBTS surveys were also presented to the group, demonstrating how a set of statistical tools can quantify the effect of a reduction in survey effort on the variance in CPUE estimates for individual species. This method was applied to a stratification, which has previously been developed for the ecosystem model 'Atlantis' (Hufnagl et al. 2014), and has been applied in the EU projects JMP[1], and in ICES WKPIMP (Workshop to plan an integrated monitoring Program in the North Sea in Q3; ICES 2016).

References

Hufnagl et al. (2014) http://www.marine-vectors.eu/Deliverables_en in D 4.2.2 - Section 5

ICES. 2016. Report of the Workshop to plan an integrated monitoring Program in the North Sea in Q3 (WKPIMP), 22-26 February 2016, ICES HQ, Copenhagen, Den-mark. ICES CM 2016/SSGIEOM:11. 48

- *Cooperation with Advisory structures*

  WGISDAA has cooperated with WGCSE and WGNSSK on examining survey information for two stock assessments that are used to provide ACOM advice.

7 Revisions to the work plan and justification

No changes made to work plan.

8 Next meetings

The next meeting is scheduled for October 2019, Copenhagen.

[1] The EU-funded project Towards a Joint Monitoring Program for the North Sea and Celtic Sea (JMP NS/CS; 10/2013-06/2015) focused on investigating benefits and challenges of joint monitoring.

Annex 1: List of participants

Name	Address	Phone/Fax	Email
Sven Kupschus	Cefas	UK	sven.kupschus@cefas.co.uk
Matthias Kloppmann	Thuenen-Institute of Sea Fisheries	D	matthias.kloppmann@thuenen.de
Casper Berg	DTU Aqua	DK	cbe@aqua.dtu.dk
Stan Kotwicki (by correspondence)	NOAA	US	stan.kotwicki@noaa.gov
David Stokes	Marine Insitute	IR	david.stokes@marine.ie
Anne Sell	Thuenen-Institute of Sea Fisheries	D	anne.sell@thuenen.de
Nicola Walker	Cefas		Nicola.walker@cefas.co.uk
Ruben Verkempynck	Wageningen Research	NL	ruben.verkempynck@wur.nl
Andrzej Jaworski	Marlab	UK	A.Jaworski@MARLAB.AC.UK
Paul Fernandes (by correspondence)	Aberdeen University	UK	fernandespg@abdn.ac.uk