DTU Library

Working Group on Improving Survey Data for Analysis and Advice (WGISDAA; outputs from 2020 meeting)

Berg, Casper; Beukhof, Esther; Börjesson, Patrik; Costas, Gersom; Cunningham, Curry; Jaworski, Andrzj; Kloppmann, Matthias; Kotwicki, Stan; Kupschus, Sven; Nash, Richard

Total number of authors:
16

Link to article, DOI:
10.17895/ices.pub.8250

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

WORKING GROUP ON IMPROVING SURVEY DATA FOR ANALYSIS AND ADVICE (WGISDAA; outputs from 2020 meeting)

VOLUME 3 | ISSUE 68

ICES SCIENTIFIC REPORTS

RAPPORTS SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA

CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

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

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

ISSN number: 2618-1371

ICES Scientific Reports

Volume 3 | Issue 68

WORKING GROUP ON IMPROVING SURVEY DATA FOR ANALYSIS AND ADVICE (WGISDAA; outputs from 2020 meeting)

Recommended format for purpose of citation:

ICES. 2021. Working group on improving survey data for analysis and advice (WGISDAA; outputs from 2020 meeting)
ICES Scientific Reports. 3:68. 57 pp. https://doi.org/10.17895/ices.pub.8250

Editors

Sven Kupschus

Authors

Casper Berg • Esther Beukhof • Patrik Börjesson • Gersom Costas • Curry Cunningham
Andrzj Jaworski • Matthias Kloppmann • Stan Kotwicki • Sven Kupschus • Richard Nash
Ismael Núñez-Riboni • Kotaro Ono • David Stokes • Ioannis Thasitis • Nicola Walker • Kai Wieland
.

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

Contents

i Executive summary

The Working Group on Improving use of Survey Data for Assessment and Advice (WGISDAA) provides a forum where survey practitioners, statisticians and stock assessment scientists can share expertise and knowledge to advance the science of fisheries surveys. The working group serves to bring together data collection and data users in a supportive and cooperative environment and to improve communication and coordination across survey groups in the ICES network.

The working group developed a strategy for making decisions on survey design and implementation that maximises the benefits of long-term survey data and minimises the impacts of enforced changes to survey design and implementation. WGISDAA concluded that the increasing use of model-based indices is beneficial in dealing with change in data collection but requires an approach that integrates across data collection and data use to control overall uncertainty in indices. This is especially important in cases where multiple objectives are addressed by a single survey but also across different surveys. Existing practices for decision making are now documented and gaps in the strategy identified. Further workshops are being planned to develop the tools that quantify the trade-offs between different objectives and establish useful estimates of uncertainty for different survey designs / data collections.

WGISDAA assisted The Working Group on Mackerel and Horse Mackerel Egg Surveys (WGMEGS), The International Bottom Trawl Survey Working Group (IBTSWG), The Working Group on Beam Trawl Surveys (WGBEAM) and The Working Group on Surveys on Ichthyoplankton in the North Sea and adjacent Seas (WGSINS) with survey design questions, most notable was the cooperation with WGMEGS on developing an improved spatial interpolation for the egg abundance data given the considerable changes in the in spawning distribution and the associated consequences to data consistency due to organisational difficulties in providing adequate survey coverage. A spatio-temporal annual generalized additive model (GAM) of egg distribution was developed which allowed for much more appropriate interpolation between samples and indicated that the total egg production and hence the standing stock biomass (SSB) was likely to be considerably smaller than previously thought. However, the temporal trend in production which matters most to the stock assessment was largely unaffected by the improved model except for the 2013 data point. Sensitivity analysis of the egg production calculation suggests that other factors in the index particularly temporal changes in fecundity or changes in catchability in the trawl survey are more likely causes of the current stock assessment difficulties.

The group worked with stock assessors for North Sea cod, West of Scotland whiting and Celtic Sea gadoids. As for the mackerel assessment, cooperation was essential for resolving disparities in the information provided by different data sources. The group helped develop more appropriate combined indices that took account of changes in spatial distribution of the stock and the survey effort. For North Sea cod however, this still proved to be a poor explanation of the underlying dynamics and further investigation suggested that there has been increased mixing with an adjacent population of cod particularly at the older ages. A more appropriate method to deal with this failure of the closed population assumption is being developed.

ii Expert group information

Expert group name	Working group on improving survey data for analysis and advice (WGISDAA)
Expert group cycle	Multiannual fixed term
Year cycle started	2018
Reporting year in cycle	3/3
Chair(s)	Sven Kupschus, UK
Meeting venue(s) and dates	3-5 July 2018, Copenhagen, Denmark (9)
	8-10 October 2019, Copenhagen, Denmark (8)
	6-8 October 2020, Virtual due to corona (16)

1 Background

The WG on improving survey data for analysis and advice focuses on evaluating scientific survey data collection and analysis methods for the advisory processes. The group provides answers, options and recommendations to other expert groups and the wider ICES networks. In this 3-year cycle the group has focused on 2 primary topics. First the evaluation of the implementation risks in the Mackerel Egg Survey (support for WGMEGS) and improving the interpolation methodology given decreased survey coverage and second the more general question of the tools and processes required to effectively manage survey implementation beyond strict consistency when technical, technological, ecological, political and financial factors force change in survey effort or methodology (support for WKUSER). In addition to these main topics the group has supported and assisted stock assessors in preparation for benchmarks with questions on North Sea cod, West of Scotland whiting, and Celtic Sea gadoids (support for WKESIG) and cooperated with the IBTSWG on prioritizing changes for the North Sea IBTS-survey (support for WKNSIMP).

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

Until recently, the Northeast-Atlantic mackerel stock was considered as consisting 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 Northeast 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 considered to be 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.

2 The Mackerel and Horse Mackerel Survey in the North-east Atlantic

2.1 Introduction

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, i.e. every other transect is sampled by each survey participant during a first pass of their assigned survey area, the remaining gaps being filled during the return pass. A transect has to be followed until 2 consecutive zero counts of freshly spawned mackerel eggs were encountered. A strategy referred to in the MEGS manual (ICES 2019) as adaptive survey design. TAEP is then calculated from stage 1 egg abundance data. With the fecundity values estimated during the same survey, the TAEP of mackerel is converted into a corresponding SSB value, which is used as 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 2019 (Figure 2.1). The considerable expansion of spawning area and increase in the duration of spawning period (and divergence of the timing of peak spawning for the two-target species) meant that the sampling resolution was substantially reduced despite recruitment of additional survey partners at the international level. In an effort to maintain full coverage of spawning area and period transect distance was increased by sampling only alternate latitudinal transects. The changes to sampling and indications of conflicting trends in some of the indices used in the assessment raised the question whether the indices provided by WGMEGS continued 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.

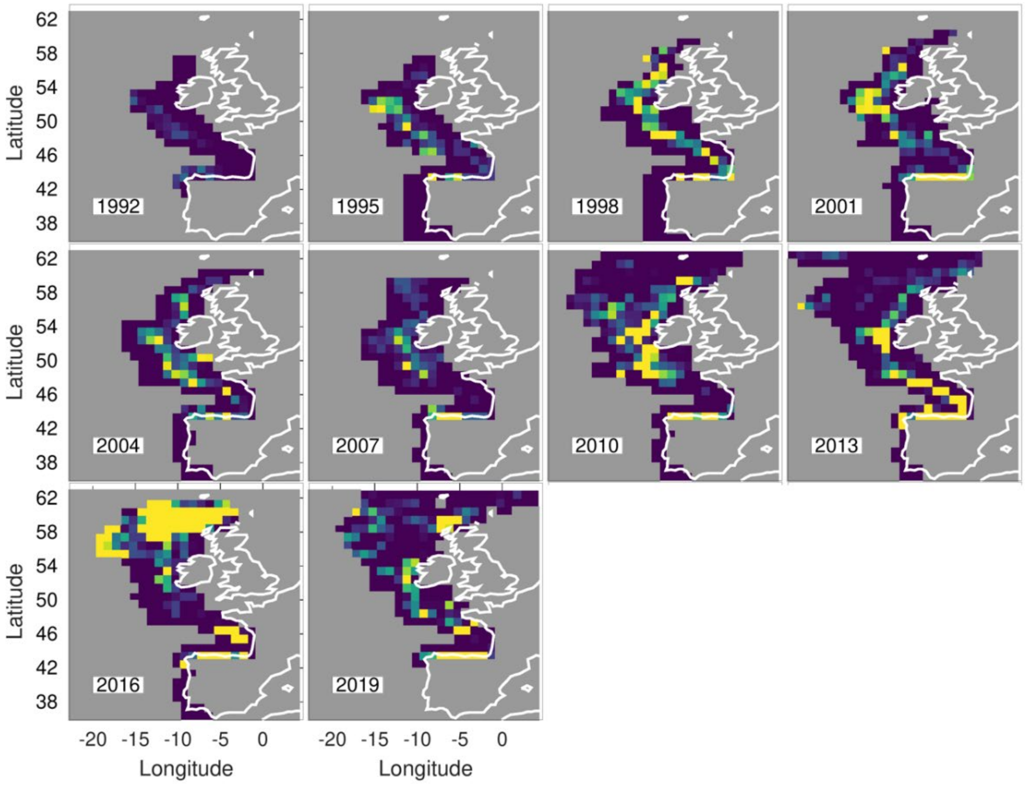
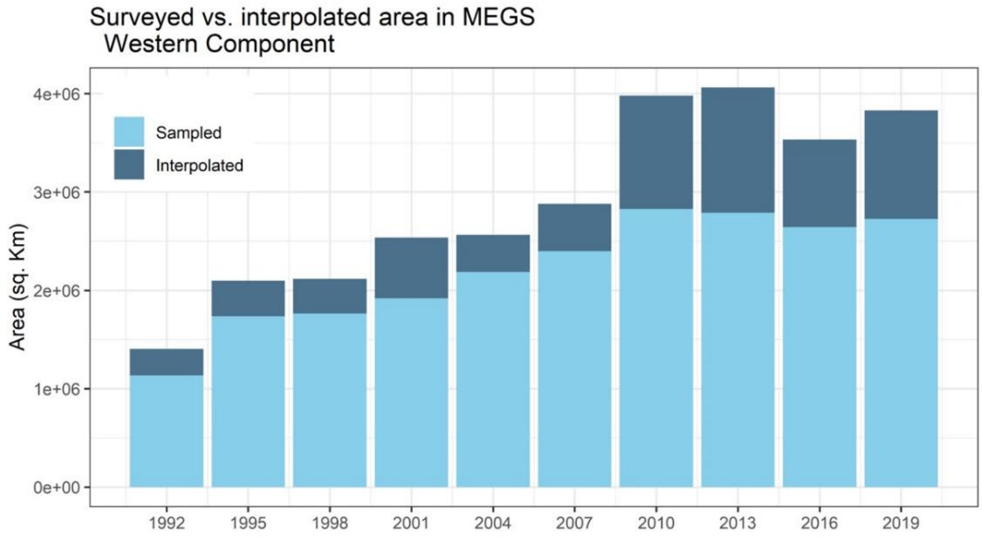


Figure 2.1: Map showing the evolution of mackerel spawning area from the 1992 to 2019 MEGS. Color coding from blue to yellow. The lighter the color, the higher the egg production

The latter recommendation gained recognition the MEGS manual, which had currently been revised (ICES 2019). 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 and 2019 surveys it appeared that spawning maximum had returned to the traditionally later period in April/May and happened just in the northern (2019) and northwesterly (2016) expanded area. Again, estimation of TAEP relied heavily on between transect interpolation (ICES 2019). An inspection of the MEGS time-series showed that the area, where interpolation had to be applied, did indeed increase for particularly the 5 recent surveys (Figure 2.2) 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 2.3). In the southern stock component, the contribution of interpolation was not as imminent.



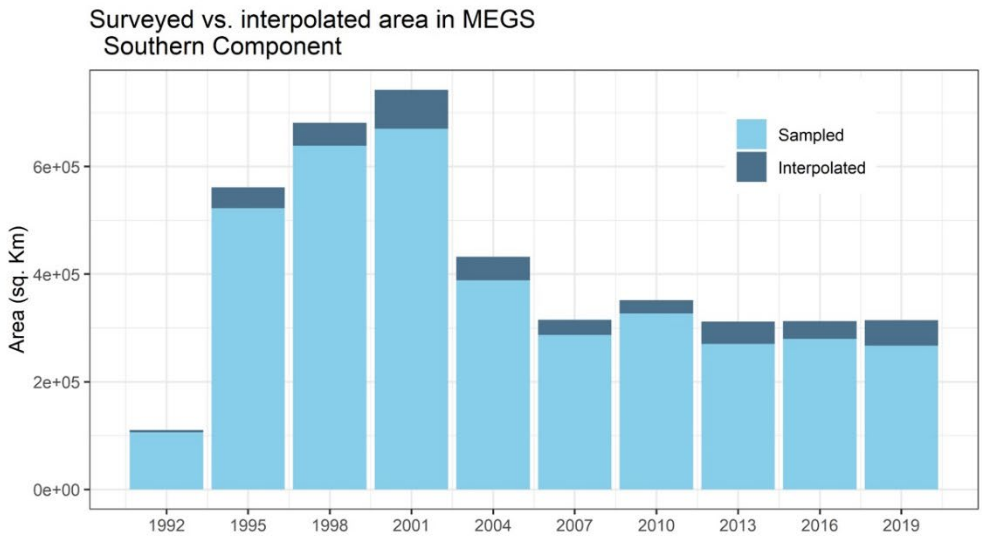
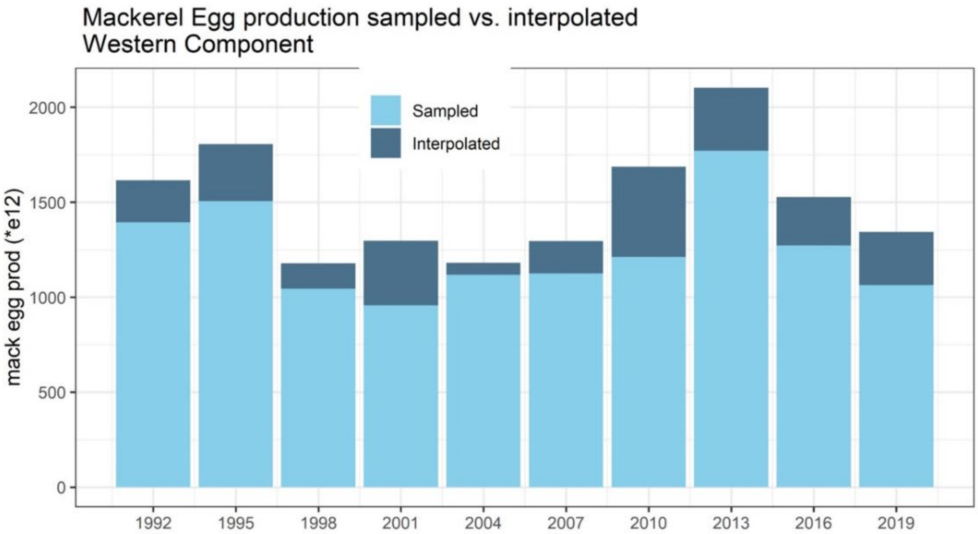


Figure 2.2 Mackerel. Prospected and interpolated area ratio within the area where mackerel eggs were found for the Western (top) and Southern (bottom) mackerel spawning stock components for the MEGS time-series. Dark blue: actual sampled area, light blue: interpolated area during the surveys.



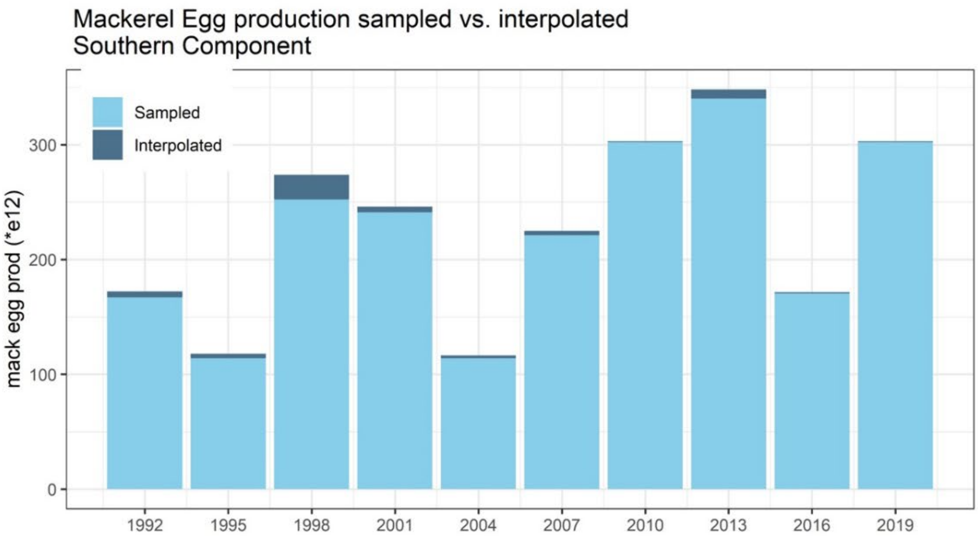


Figure 2.3 Effect of spatial interpolation on mackerel daily egg production estimates for Western (top) and Southern (bottom) mackerel spawning stock components for the MEGS time-series. Dark blue: actually, sampled egg production, light blue: interpolated egg production during the surveys.

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 2.4). Temporal interpolation was particularly relevant in the surveys for the southern spawning component.

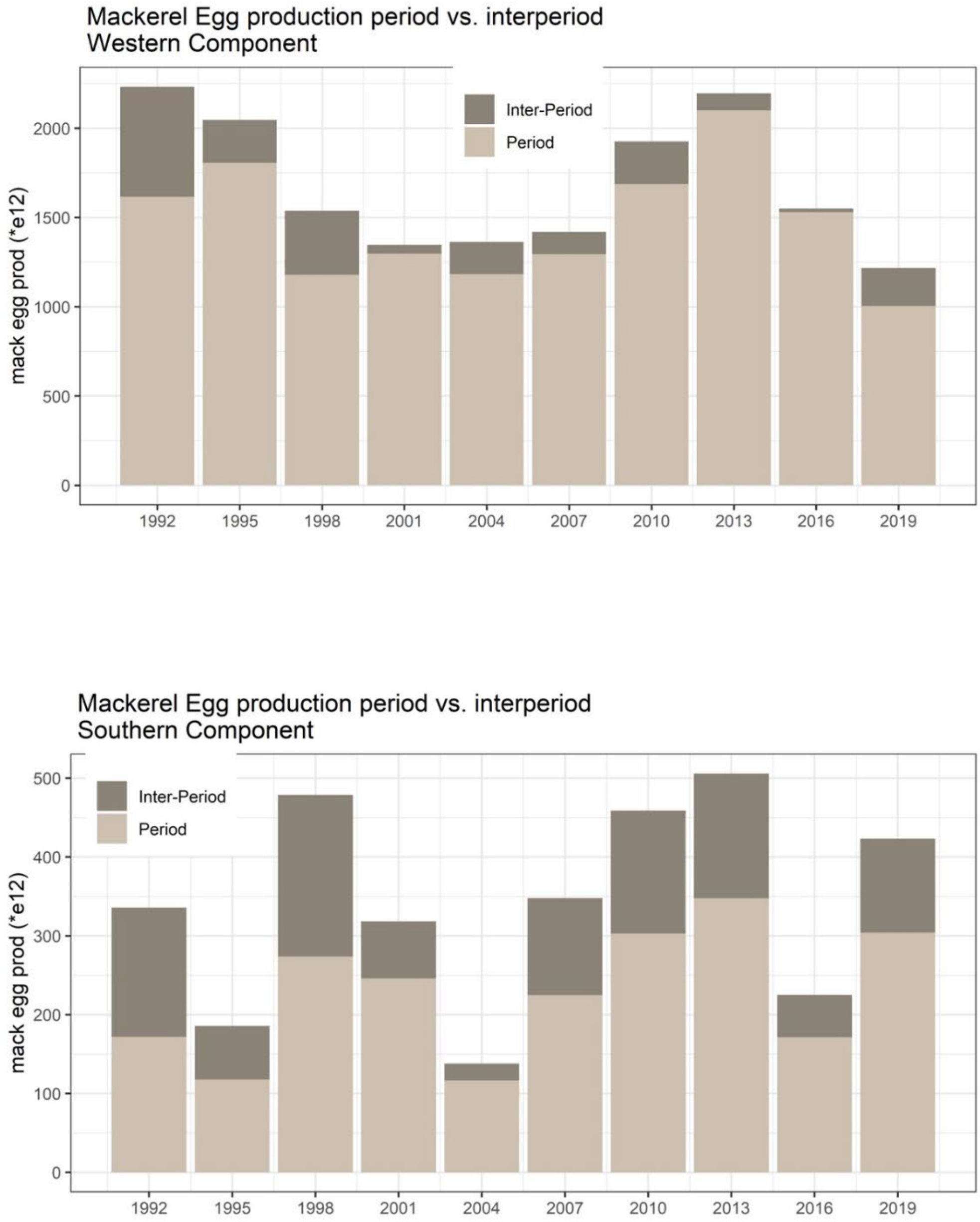


Figure 2.4 Effect of temporal interpolation on annual egg production of mackerel for Western (top) and Southern (bottom) mackerel spawning stock components for the MEGS time-series. Dark purple: means egg production in sampling period, light purple: egg production in interpolated period (inter periods).

It is evident, that interpolation is widely used in calculating TAEP in both mackerel and horse mackerel (see also ICES 2019). 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 (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 and ended in 2020, WGMEGS pursued this recommendation and has asked WGISDAA for support. Results of the exercise, which was at first only carried out for the western component, are presented below.

2.2 Estimation of total annual egg production using GAMs

The annual egg production curve for the western spawning mackerel component was calculated trying to follow the traditional method (Section 10.2 of ICES 2019) as much as possible. Only the spatial interpolation has been replaced with estimates from a spatial generalized additive model (GAM; Hastie and Tibshirani 1986) as described in detail in what follows.

The estimated daily egg production per square-meter data (EP) per station from the triennial egg surveys were used as input data for the egg production model. In deviation from the usual practice of using variable sampling periods, which arise from variable ship's availability dates among survey years, and in order to avoid inter-period interpolation, the model uses six fixed sampling periods of roughly 30 day long, defined with the Julian day as shown in Table 2.1. EP was mapped with the following GAM with Tweedie distribution (Shono 2008; Augustin et al 2013; Peel et al 2013) and (canonical) logarithmic link, for each sampling period:

$$log(\widehat{EP}) = s(lon, lat), \qquad (1)$$

where s is a penalized thin plate spline smoother (Wood 2017) and (lon, lat) represents the geographical position of the egg samples. The basis dimension k has been chosen as $k = 0.6\times N$, where N is the number of observations in each sampling period. Defining k as a large fraction of N warrants that the model optimally explains the observed variance but keeping the computational time short.

The GAM was fitted using the package MGCV (Mixed GAM Computation Vehicle; Wood, 2017) from R, version 3.6.3. To map EP with the model, a regular grid with the same average resolution of the ICES survey (0.5°×0.5°) was constructed matching the great majority of the survey's sampling positions. EP was predicted with the model only on grid points lying inside triangles from a Delaunay triangulation (Swan and Sandilands, 1995) with all 3 sides smaller than 2 geographical degrees (Figure 2.5). Such convex mapping region warrants that the model estimates arise only from observations on the edges of the sampling region or from nearby surrounding observations. Therefore, extrapolated estimates and, with it, "boundary effects" of the smoother spline are avoided (Figure 2.6).

Table 2.1 Definition of 6 sampling periods along the year in terms of Julian Day (JD). The periods are non-overlapping, i.e. excluding the 1st period, sampling periods do not include their starting day.

Period	Start (JD) >	End (JD) ≤	Length	Month (approx)
1	35 (4 Feb)	65 (6 Mar)	30	February
2	65 (6 Mar)	94 (4 Apr)	29	March
3	94 (4 Apr)	124 (4 May)	30	April
4	124 (4 May)	153 (2 Jun)	29	May
5	153 (2 Jun)	183 (2 Jul)	30	June
6	183 (2 Jul)	212 (31 Jul)	29	July

Once EP was mapped on a regular grid, total annual EP was integrated using the classical approach as described in what follows (see also Section 10.2 of ICES 2019). First, the area A_i of the i^{th} model grid was calculated with:

$$dA_i = dX_i \times dY_i = (cos(lat_i \times \pi\ /\ 180) \times sin(S \times \pi\ /\ 180) \times R) \times (sin(S \times \pi\ /\ 180) \times R)$$

where R is the Earth's radius, i.e. $R = 6{,}371{,}000$ m and the cell size $S = 0.5°$, as mentioned above.

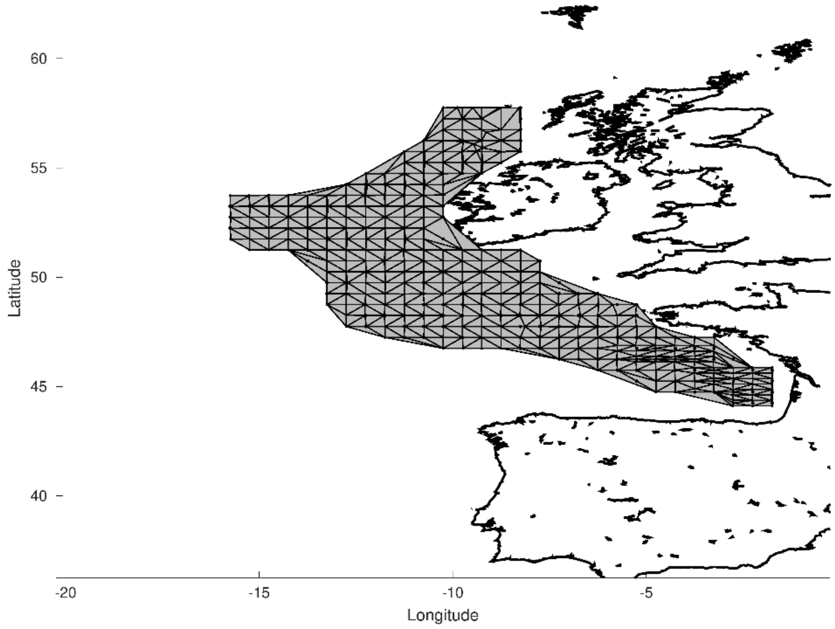


Figure 2.5 Delaunay triangulation of sampling positions of the ICES survey of the mackerel western spawning component for the 5th sampling period (Table 2.1) of the arbitrarily chosen year 1995. Lines joining the positions of the egg hauls are edges of Delaunay triangles.

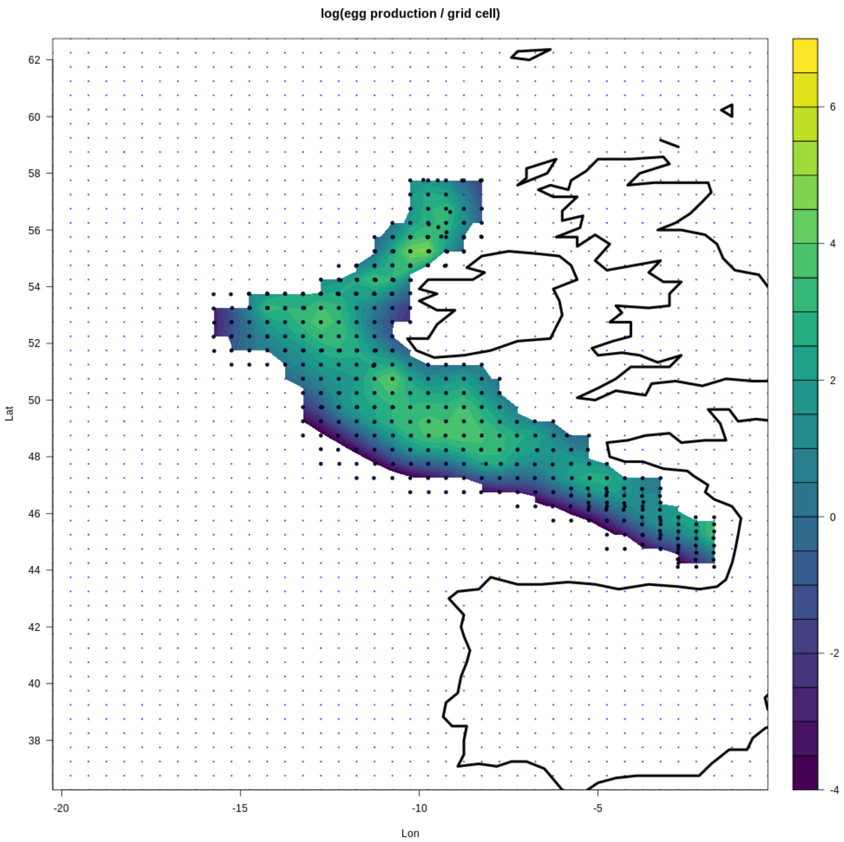


Figure 2.6 Logarithmic mackerel daily EP per grid point for the western spawning component for the 5th sampling period (Table 2.1) of the arbitrarily chosen year 1995. Black dots are the sampling positions. Blue dots are the mapping grid points of the GAM (Equation 1).

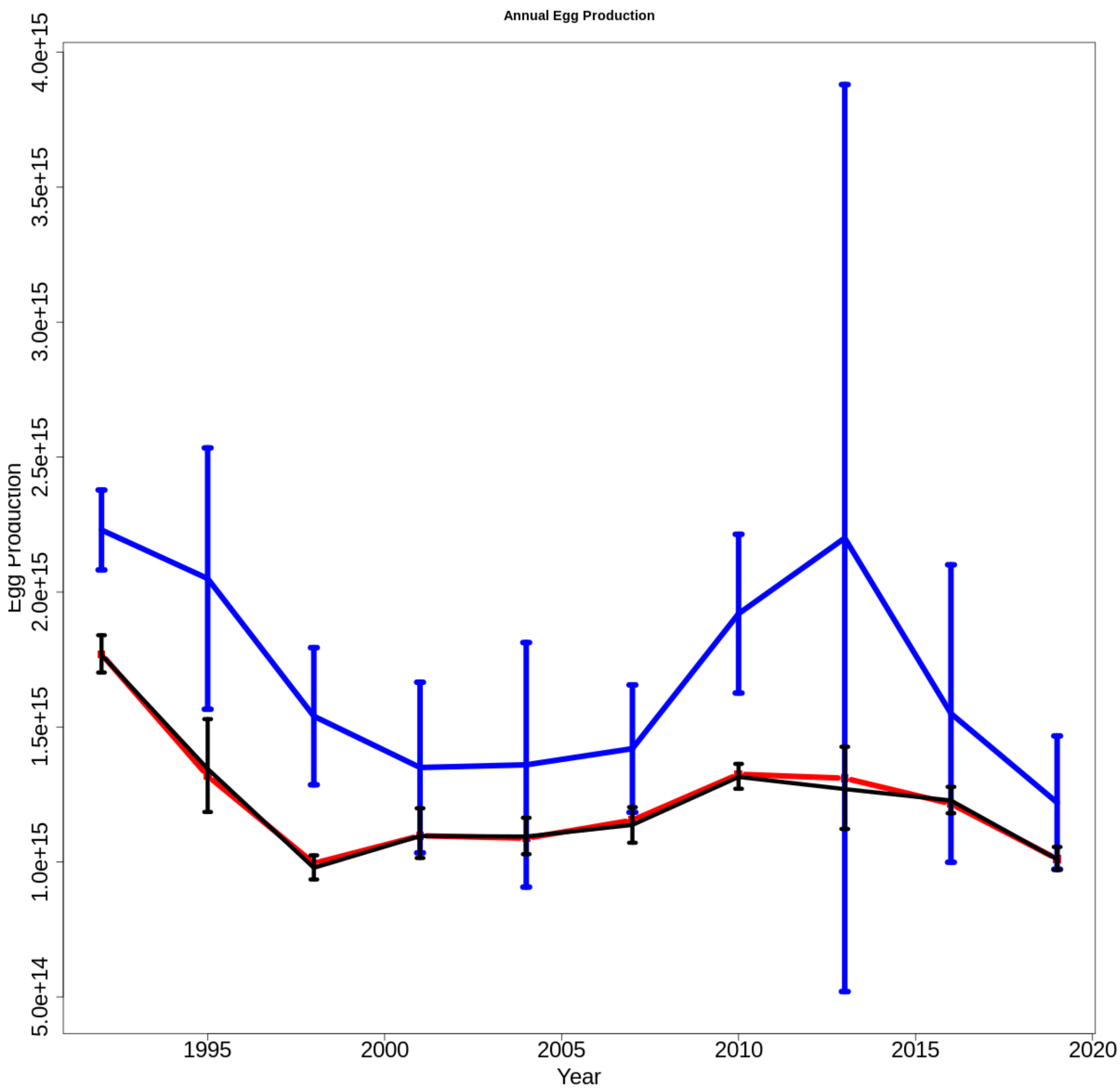


Figure 2.7 Total production of Mackerel eggs per year as calculated with the new interpolation method (red curve). Median egg production, 5% and 95% quantiles from 500 bootstrapped datasets are shown with a black curve and black error bars, correspondingly. The results from the traditional method are shown with the blue curve and blue error bars (the standard deviation, i.e. corresponding to 68% interval).

The total EP in the i^{th} grid cell is:

$$EP_i^c = EP_i \times dA_i \,, \qquad (2)$$

where the upper script c emphasizes that this is EP per grid *cell*. Therefore, the total EP for the whole mapping region and sampling period P is the sum over all grid cells:

$$EP_t^p = \sum EP_i^c \qquad (3)$$

Finally, the total EP for the present year is obtained by integrating each of the individual contributions EP_t^p over all 6 sampling periods:

$$EP_t = \sum EP_t^p \times dD_p \qquad (4)$$

where dD_p is the length of the p^th^ sampling period (see Table 2.1). The resulting total EP is shown in Figure 2.7.

2.3 Confidence limits for egg production

Confidence limits for the TAEP have been calculated following a non-parametric bootstrapping approach (Efron, 1979). Bootstrapping has previously been applied in fisheries science to estimate the uncertainty of key variables of stock assessments, like recruitment and fishing mortality (Mohn, 1993; Haddon, 2011), but also to estimate the coefficient of variation of Mackerel egg abundance (Borchers et al 1997). The idea is to repeat the process described above a large number of times with data resampled with replacement from a characteristic or "universal" dataset. This allows obtaining a distribution of pseudo-observations around each EP_t and, with it, error bars with a particular confidence. Therefore, we replace each observation EP_i with a randomly generated estimate EP_i' drawn from a probability distribution ϵ(lon, lat, time). The dependence of ϵ on position and time follows the notion of reproducing the heteroscedasticity of the EP, i.e. regions or years typically presenting large mean EP have also large EP variances.

The error estimation ϵ(lon, lat, time) is obtained by randomizing the EP observations inside spatial and temporal bins. For the choice of temporal bins three options were tested: the complete time-series, annual time bins and the individual sampling periods per year of Table 2.1. For the case of the spatial bins, we tested four sizes: 0.2, 0.5, 1.0 and 2.0° in both longitudinal and latitudinal directions. A too small spatio-temporal bin constrains the size of the universal dataset from which the random samples are drawn to few members (to a single one in an extreme situation not tested here). This obviously leads to underestimation of the natural variability and unrealistically narrow error bars. On the other hand, randomization of data on too large spatial or too long temporal bins yields biases in the bootstrapped total egg production curves. These biases indicate that the average total egg production from the bootstraps strongly depends on estimating the natural spatial and temporal autocorrelation of EP. EP observations of similar value occur near each other in space and time. Once those similar EP observations are separated by the randomization, their contribution to the total EP seems weaker than when they are close to each other. Therefore, the average total EP curve from the bootstraps being biased compared with the observed one is an indication that the bootstraps are drawn from data distributed over spatial or temporal scales larger than the natural scales of variability.

Thus, our main criterion to choose the sizes of the bins was to obtain a bootstrapped average total annual EP curve similar to the one observed with the real observations. This warrants that the randomized data are inside the range of natural variability and observational error. Based on trial and error with small bootstrapped sets (20 iterations), we found that the best bins for the bootstraps were the time bins of the sampling periods (Table 2.1) and spatial bins of 0.5 (the native survey's resolution). The procedure of generating random pseudo-observations of EP was repeated 500 times and total annual EP were calculated each time following the procedure of Section 2.2. above. This led to a distribution of error estimates centered on each EP_t), which in turn allowed to calculate 95% confidence limits with their empirical histograms (black error bars in Figure 5.7).

The total EP obtained with the new method is somewhat smaller on average (0.4 x 10^{15} eggs) than the traditional estimates. An explanation might be related to the spatial average of the traditional method, which implicitly assumes a normal distribution of EP. Local EP variations are, however, not normally distributed but strongly skewed towards large values. The real mean EP is thus not symmetrically centred on half way between small and large EP values, but is closer to the small than to the large values. Such asymmetry in the EP distribution is better represented by the Tweedie GAM, evidencing a positive bias in the traditional method. A first comparison with survey results by year and sampling period, however, also suggests that at least some of the differences between the GAM and the traditional method may lie in the fixed period dates chosen for the GAM. These fixed period dates often result in data from two survey periods being moved into a single model period, creating gaps in one of the neighboring model periods. These spatial gaps could be closed either increasing the allowed distances in the Delaunay triangulation or by including time (i.e. Julian day) as a third predictor in the GAM.

2.4 Future work on GAM models

Once the above mentioned problems have been solved, the chosen GAM shall be applied to the southern component as well, in order to achieve a TAEP time-series for the entire Northeast Atlantic stock, excluding the North Sea component. The model will also be used to evaluate the survey effort by gradually removing selected groups of station (e.g. whole transects) and recalculate TAEP with the reduced dataset.

2.5 Sensitivity of SSB calculation to variations of parameters in fecundity estimation

One of the aims of the mackerel and horse mackerel egg surveys (MEGS) is to provide an index of spawning-stock biomass (SSB) for the NEA mackerel. This index feeds into stock assessments for NEA mackerel as fishery-independent information. This SSB index is estimated applying the Annual Egg Production Method (AEPM, Lockwood et al., 1981).

The total annual egg production (TAEP), i.e. the total number of eggs produced over the entire spawning season in the spawning area is estimated from abundance values of freshly spawned mackerel eggs. TAEP is divided by the numbers of eggs produced per gramme female (realized fecundity, F_r). Accounting for the female to male ratio R in mackerel spawning stock and applying a correction factor C, which adjusts for prespawning to average spawning fish weight difference, its biomass (the SSB) can be estimated as follows:

$$SSB = \frac{TAEP}{F_r} \times \frac{1}{R} \times C \quad (1)$$

Currently, $R = 0.5$ and $C = 1.08$

Although being a simple method (Bernal et al., 2012), it requires both an accurate estimate of the total numbers of eggs spawned over the entire spawning area and season, as well as an accurate realized fecundity (total number of eggs spawned per g female in one spawning season) of a female.

A sensitivity analysis was performed applying a percentage increase or decrease change of mean realized fecundity (F_r) estimated during period 2001-2019 of MEGS temporal series. A percentage change from 50% to 150% (equivalent to a factor of 0.5 – 1.5) in F_r was applied to a range of egg production estimates within the observed maximum and minimum TAEP values between 2001 and 2019 (ICES, 2020). For the evaluation of the impact on SSB estimates the mean realized fecundity F_r for the survey years 2001-2019 (1087 oocytes/g) was used.

Figure 2.8 shows the effect of changes in F_r using the TAEP estimates of the survey years 2001-2019. The baseline represents the SSB estimation using the mean $F_{r,2001\text{-}2019}$ = 1087 eggs per g female. An asymmetric impact can be noted depending on whether F_r is increasing or decreasing. Results show that decrease in F_r has a greater impact in SSB estimates than an increase in F_r (Table 2.2). A 20 % reduction of the mean F_r results in a 25% increase of the SSB estimate, while an increase of the mean F_r of 20% will result in a reduction of only 17% in the SSB estimate (Table 2.2).

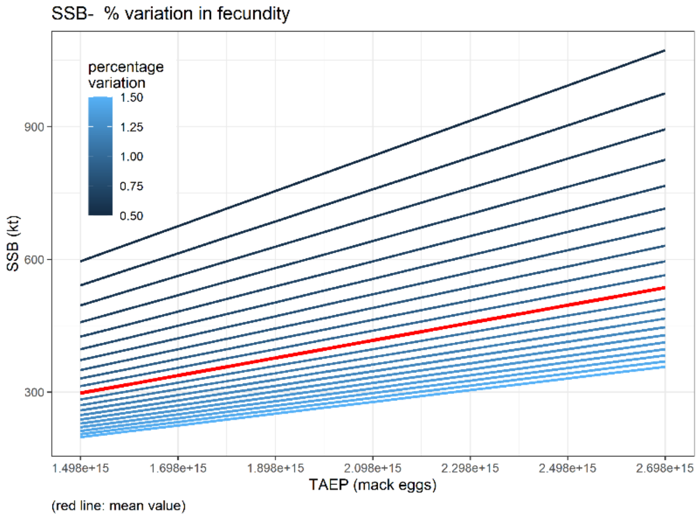


Figure 2.8 Effect in SSB calculated according to equation (1). Egg production values (AEPM) when is applied a percentage change in mean realized fecundity (from 50% to 150%). TAEP values used were within the range between the lowest and the highest estimates of mackerel TAEP ($1.49*10^{15}$ to $2.7*10^{15}$) in the period of the 2001-2019 from MEGS surveys.

Table 2.2. SSB estimate from equation (1) using a percentage change (50% to 150%) in mean realized fecundity (1087 oocytes/g). In this table, as baseline for comparison, the mean mackerel TAEP ($1.89*10^{15}$) of the MEGS between 2001-2019 has been used.

Mean TAEP	Mean Fr	Mean SSB	Variation Fr	Fr estimate	SSB	SSB variation
1.898E+15	1087	377	50%	543	754	+100%
1.898E+15	1087	377	55%	598	686	+82%
1.898E+15	1087	377	60%	652	629	+67%
1.898E+15	1087	377	65%	706	580	+54%
1.898E+15	1087	377	70%	761	539	+43%
1.898E+15	1087	377	75%	815	503	+33%
1.898E+15	1087	377	80%	869	471	+25%
1.898E+15	1087	377	85%	924	444	+18%
1.898E+15	1087	377	90%	978	419	+11%
1.898E+15	1087	377	95%	1032	397	+05%
1.898E+15	**1087**	**377**	**100%**	**1087**	**377**	**0%**
1.898E+15	1087	377	105%	1141	359	-05%
1.898E+15	1087	377	110%	1195	343	-09%
1.898E+15	1087	377	115%	1250	328	-13%

1.898E+15	1087	377	120%	1304	314	-17%
1.898E+15	1087	377	125%	1359	302	-20%
1.898E+15	1087	377	130%	1413	290	-23%
1.898E+15	1087	377	135%	1467	279	-26%
1.898E+15	1087	377	140%	1522	269	-29%
1.898E+15	1087	377	145%	1576	260	-31%
1.898E+15	1087	377	150%	1630	251	-33%

Sensitivity of SSB estimation to changing parameters of the atretic loss calculation

In order to calculate realized fecundity (F_r), the estimate of potential fecundity needs to be corrected for the loss through atresia (i.e. the number of oocytes, which are resorbed, ICES, 2014e). Accordingly, realized fecundity is derived by subtracting mean atretic loss (A_r), i.e. the loss of developing oocytes through atresia per gramme female during spawning, from the relative potential annual fecundity (F_p, the mean number of vitellogenic oocytes before spawning per gramme female).

$$F_r = F_p - A_r \quad (2)$$

A_r is estimated in the population using the following equation:

$$A_r = F_{atr} \times P_{rev} \times \frac{SD}{D} \quad (3)$$

Where

- F_{atr} = Mean number of atretic oocytes per gramme female showing atresia during spawning.
- P_{rev} = Proportion of females in spawning condition with atresia.
- SD = spawning duration of an individual female mackerel, assumed as 60 days (ICES, 1993).
- D = duration of early alpha atresia in an individual female mackerel, assumed as 7.5 days (ICES, 1993).

Therefore, mean atretic loss is a function of the F_{atr}, the prevalence of atresia (P_{rev}), spawning duration (SD) and the duration of atresia (D). While the F_{atr} is calculated from counts of atretic oocytes in histological samples, SD and D are constant values, which were set in 1993 based on a very limited data. Though the values of 60 days for SD and 7.5 days for D were set as provisional, they were used unaltered since. A sensitivity analysis (OAT) was run to assess how the SSB estimate would be affected by changes of these values.

Figure 2.9 shows the effect in variation of the assumption of spawning duration (60 days) on SSB estimate based on varying TAEP values. Table 2.3 lists the impact of the varying SD value on the mean TAEP values of the period 1998 – 2019. The baseline was set allowing TAEP to vary between the min and max of the period 1998 - 2019 ($1.49*10^{15}$ to $2.7*10^{15}$ mackerel eggs, ICES, 2020), and using the mean annual potential fecundity from the same years (1159 oocytes/g), a mean relative atresia (31 atretic oocytes/g), the assumed duration of alpha atresia (7.5 days) and mean prevalence of atresia (32%) from the same years. Neither increasing nor decreasing SD by factors between 0.5 and 1.5 had a large impact on the SSB estimate (from 30 d to 90 d, Table 2.3).

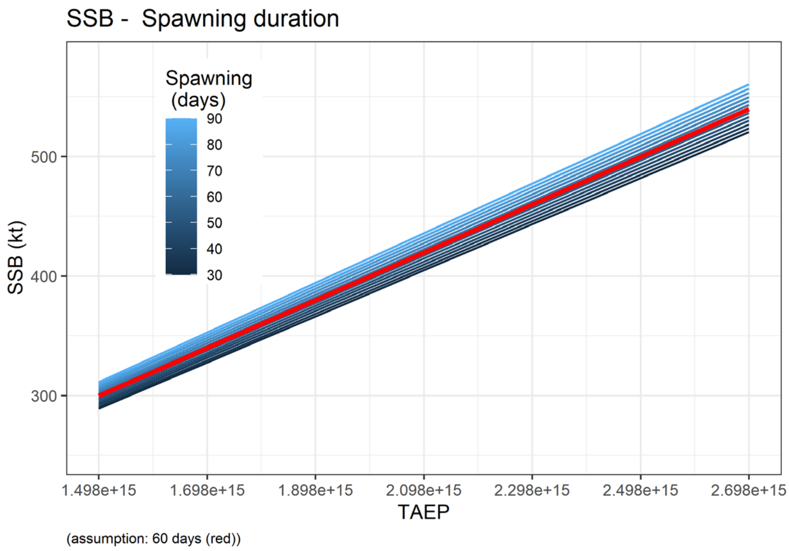


Figure 2.9 Effect of varying spawning duration (30 to 90 days) on the SSB estimate applying AEPM. The baseline (red line) was set using SD of 60 days (ICES 1993). TAEP values used were within the range between maximum and minimum of mackerel TAEP ($1.49*10^{15}$ to $2.7*10^{15}$) in the period 2001-2019 from MEGS surveys

Table 2.3. SSB estimates from AEPM using a range of values (30 to 90 days) as spawning duration. As SSB baseline estimate (bold) for comparison, the mean TAEP, potential fecundity, relative atresia, duration of alpha atresia and prevalence of atresia of from 1998 - 2019 MEGS was applied.

TAEP	Spawning duration (SD)	Mean F_r	Mean prevalence	Mean F_p	Mean SSB	F_r reestimate	SSB	variation SSB
1.90E+15	30	1080	0.32	1159	380	1119	366	-04%
1.90E+15	35	1080	0.32	1159	380	1113	368	-03%
1.90E+15	40	1080	0.32	1159	380	1106	371	-02%
1.90E+15	45	1080	0.32	1159	380	1099	373	-02%
1.90E+15	50	1080	0.32	1159	380	1093	375	-01%
1.90E+15	55	1080	0.32	1159	380	1086	377	-01%
1.90E+15	60	1080	0.32	1159	380	1080	380	0%
1.90E+15	65	1080	0.32	1159	380	1073	382	+01%
1.90E+15	70	1080	0.32	1159	380	1066	384	+01%
1.90E+15	75	1080	0.32	1159	380	1060	387	+02%
1.90E+15	80	1080	0.32	1159	380	1053	389	+03%
1.90E+15	85	1080	0.32	1159	380	1046	392	+03%
1.90E+15	90	1080	0.32	1159	380	1040	394	+04%

Figure 2.10 shows the effect of changing the assumed early alpha atresia duration (7.5 days) on the SSB estimate, using the same values of TAEP, F_{atr} and F_p. As baseline for comparison, the SSB has been estimated using a range of TAEP to vary between the minimum and maximum of the period 1998 - 2019 ($1.49*10^{15}$ to $2.7*10^{15}$), as well as the mean potential fecundity (1159 oocytes/g), mean relative atresia (31 atretic oocytes/g), assumed spawning duration (60 days) and prevalence of atresia (32%).

An asymmetric impact on the SSB estimates was detected depending on whether D is increased or is decreased (from 2 to 10 days). Results show that increase in D causes lesser impact in SSB estimates that decrease in D, e.g. while decreasing y by 2.5 days (D = 5 days) causes an increase of a 4% in the estimated SSB, an increase of 2.5 days (D = 10 days) would result in a decrease of only 2% in the SSB estimate (Table 2.4).

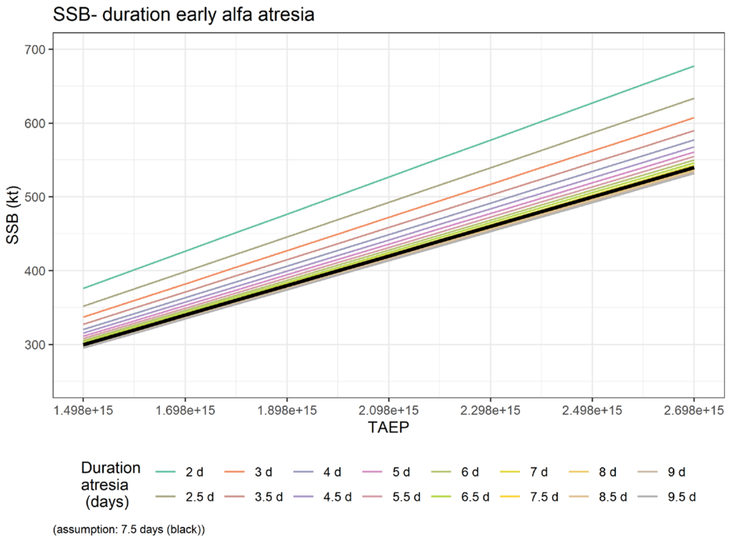


Figure 2.10. The effect different atresia durations (2 to 9.5 days) on the SSB estimate applying AEPM. As baseline (black line) a D of 7.5 days (ICES 1993) was used. TAEP values were chosen to vary in the range between minimum and maximum of mackerel TAEP ($1.49*10^{15}$ to $2.7*10^{15}$) of the 2001-2019 MEGS.

Table 2.4. SSB estimates from AEPM equation using a range of values (2 to 10 days) for early alpha atresia duration. As SSB baseline estimate (bold) for comparison the average of TAEP ($1.898*10^{15}$ mackerel eggs), potential fecundity, mean relative atresia, spawning duration and prevalence of atresia in period 1998 - 2019 from MEGS surveys was used.

TAEP	Atresia duration (D)	Mean F_r	Mean prevalence	Mean F_p	Mean SSB	F_r re-estimate	SSB	variation SSB
1.90E+15	2	1080	0.32	1159	380	860	476	+25%
1.90E+15	2.5	1080	0.32	1159	380	920	445	+17%
1.90E+15	3	1080	0.32	1159	380	960	427	+12%
1.90E+15	3.5	1080	0.32	1159	380	988	415	+09%
1.90E+15	4	1080	0.32	1159	380	1010	406	+07%
1.90E+15	4.5	1080	0.32	1159	380	1026	399	+05%
1.90E+15	5	1080	0.32	1159	380	1040	394	+04%
1.90E+15	5.5	1080	0.32	1159	380	1051	390	+03%
1.90E+15	6	1080	0.32	1159	380	1060	387	+02%

1.90E+15	6.5	1080	0.32	1159	380	1067	384	+01%
1.90E+15	7	1080	0.32	1159	380	1074	382	+01%
1.90E+15	7.5	1080	0.32	1159	380	1080	380	0%
1.90E+15	8	1080	0.32	1159	380	1085	378	0%
1.90E+15	8.5	1080	0.32	1159	380	1089	376	-01%
1.90E+15	9	1080	0.32	1159	380	1093	375	-01%
1.90E+15	9.5	1080	0.32	1159	380	1096	374	-02%
1.90E+15	10	1080	0.32	1159	380	1099	373	-02%

Results of these analyses show that changes in the fecundity estimate can have a significant impact on the SSB estimate. Also, effects of changes in fecundity are asymmetric in that reduction has a higher impact than an increase by the same amount. The analyses also showed that single parameters in realized fecundity estimation, which were provisionally set and left unchanged, may increase uncertainty in SSB estimation if these assumptions prove to be false. As a consequence, accurate fecundity estimates as well as regular examinations of the assumptions on which it is based, are important in order to deliver a reliable trend in the AEPM based SSB index.

The sensitivity analysis shall be further carried out including multiple parameters at the same time to investigate how combined or cumulative effects in changes of those parameters affect fecundity and SSB estimation.

2.6 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 it appeared that there had 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 2.11 and Table 2.5). 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.

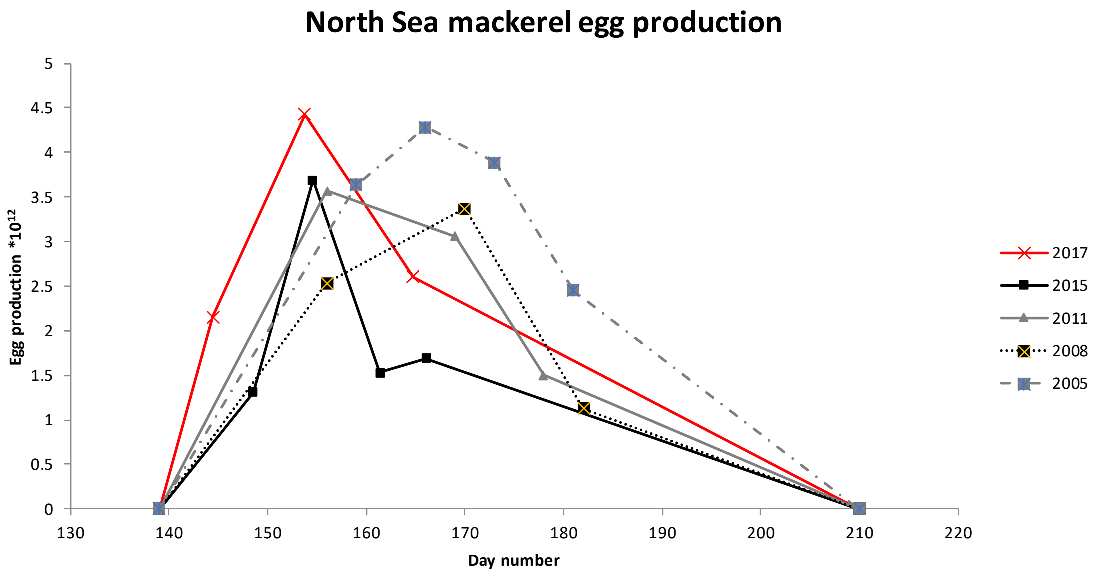


Figure 2.11 Annual egg production curves for North Sea mackerel (prior to 2015 the Lockwood egg development equation was used, since 2015 the Mendiola equation was used).

Table 2.5 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 period (ICES 2018). The remaining ship time plus the 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 first 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 was 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.

Only recently, however, the North Sea Mackerel Egg Survey received more attention through WGWIDE and the Workshop on a Research Roadmap for Mackerel (WKRRMAC) in that the North Sea could no longer be considered as an isolated spawning region for a discrete stock component. While genetic evidence on the discreteness of the southern, western and North Sea mackerel stock components is lacking, it was recommended that the North Sea should be included in the triennial Northeast Atlantic survey (ICES 2019). WGMEGS is currently investigating possibilities extending the triennial Northeast Atlantic survey area into the North Sea.

2.7 WGISDAA conclusion

The spawning-stock biomass estimates WGMEGS have been usefully challenged by the development of a swept-area index for mackerel using trawl data. This has been particularly important in relation to the expansion of spawning and reduction of survey coverage possible. The divergence between the indices was thought to be an artefact of the sample interpolation methodology used in conjunction with the reduced survey effort. The in-depth investigation carried out in cooperation between WGISDAA and WGMEGS resulted in an improved model for the spatial interpolation of egg abundance indices, which suggests that although SSB is likely to be lower than historically estimated the trends is not materially different and therefore is unlikely to resolve the conflict in the assessment where the index is used as a relative measure of SSB.

Egg abundance is only a part of the derivation of the SSB index. The potential importance of factors such as time varying fecundity or atresia were investigated through sensitivity analysis and suggest that potential variations in the assumed constant fecundity have the greatest potential to affect the temporal trend in SSB estimates.

Future cooperative work will focus on investigating the evidence of such changes in fecundity. WGISDAA did review the initial survey design of the swept-area survey and concluded that the design in the horizontal direction was sound but recently there has been some suggestion from acoustic and trawl data that increases the potential for changes in catchability in that survey too cause through differential behavior in the vertical dimension. This could also be a source of divergence between the two indices and should be reconsidered in light of the new findings.

3 Developing an effective strategy for making decisions on survey implementation integrated across multiple objectives and survey designs

3.1 Report back from the Workshop on Unavoidable Survey Effort Reduction

Stan Kotwicki presented summary of the outcomes from the Workshop on Unavoidable Survey Effort Reduction (WKUSER) which examined the extent of the issue and its effects on use of survey data by reviewing available research, evaluating current practices, and recommending future directions on four key topics:

- **TOR 1: Current processes.** The current processes dealing with unavoidable reductions (and often subsequent increase) in survey effort and examining the existing coping strategies (e.g. spatial coverage, survey frequency, or sampling density) and their qualitative consequences.
- **TOR 2: Survey uncertainty.** Develop key quality metrics that can be used to describe "total survey uncertainty" for common survey designs and indices of abundance.
- **TOR 3: Survey continuity.** Define "changes to survey designs" that require inter-survey calibration and what changes can be resolved by a model-based approach to index generation.
- **TOR 4: Decision-making tools.** Develop methods that can provide quantitative, decision-making tools describing impacts on the quality of survey data and advisory products.

WKUSER concluded that surveys effort reductions often present a reactive challenge to survey and assessment scientist and there is a little pre-event advice on the subject. Decision trees and tables were developed under each key topic to assist survey managers in decision-making on different time-scales. They also indicate key scientific knowledge needed to formulate coherent responses between survey and assessments. Together the trees can deliver best practice decision tools and provide assessments of the impact of survey effort reductions on data and advice quality through a series of questions linked to information tables.

Recommendations for best practices and future refinements of process are:

- Monitoring agencies are encouraged to routinely apply the developed approach to conduct survey evaluations. This will facilitate appropriate prioritization of monitoring tasks by examining its relation to objectives by exploring possible methods for gains in survey efficiencies (such as: reducing the number of biological samples, shortening tow duration, increase in catch subsampling while also considering station thinning, excluding areas, reducing survey frequency, or changing survey design).
- Continue further studies on estimation of total survey uncertainty by conducting research into the various subcomponents inherent in survey design and metric calculations, which include sampling design, sampling efficiency, spatial availability, density-dependence, vessel effects, timing, and environmental conditions. The interactions of these uncertainty components require studies to assess total survey uncertainty for appropriate weighting in likelihood-based assessments, provide greater insight into the impact of certain changes, and provide a long-term strategy for improved surveys.
- Develop and expand simulation studies and research on model-based capabilities that can be used to define methods for survey effort reduction, aid in estimations of total survey uncertainty, and help with inter-calibration studies.

- Survey groups and assessment groups together should develop quantitative applications that can be used for any survey and assessment combination to determine the impacts of different monitoring strategies in terms of inputs (cost) and outputs (uncertainty). They should include functions to process abundance data, and to incorporate ecosystem data for use in model-based estimation and in process studies, multispecies/multi-objective optimization, and evaluation of trade-offs between different survey and estimation approaches.
- Survey managers are recommended to intensify preparation for response to ecosystem changes, which are already underway in many areas. These preparations should include strategies for survey expansions into new areas (or reductions on other areas) to assure continued relevance of survey information to fisheries management and research.

WGISDAA reviewed and discussed these findings and supports the conclusions arrived at by WKUSER and recognizes the significant advancements provided by the workshop. More over it was felt that the direction and research suggestions developed by WKUSER are key to advancing the quality of ICES science generally but also essential to maintenance of quality for future advice in an 'uncertain survey' world.

WGISDAA considered the research priorities needed to make further progress in the field and considered. The list of topics was too extensive for WGISDAA to consider with its current membership and a further workshop is necessary to focus work and to bring in additional expertise. WGISDAA considered the ToRs for the new workshop should be developed from the following list of research topics.

Potential TORs were discussed and proposed for WKUSER2.

Proposed Subjects for WKUSER2 based on the discussions at WGISDAA:

Survey design, survey flexibility:

- Importance criterions for survey locations. Use information from the past. Densities, variance, spatio-temporal covariance to identify consistent patterns to allow for lower sampling (easier to do). Use of the information from the past surveys to allocate effort relatively to the expected information contents. Adaptive sampling based on previews of expected distribution. However, While adaptive sampling can be theoretically useful to get more information using the same or even less tows, this often implies that the analysis need to account for preferential sampling to avoid getting biased results. This complicates the analysis considerably. There is also a danger that areas with general low abundance have some unique/rare species that are not covered as well with an adaptive design. Careful evaluation of such issues should be performed before switching to an adaptive design.
- Easy calibration exercises during existing surveys to test different designs using the same gear.
- We want to be prepared to the ecosystem change. How to embed more ecosystem process studies to inform catchability and spatial dynamics studies. Technology can help. develop methods for coming up with design that allows options and strategies to allow flexibility in survey effort allocation between years areas to maximize information from surveys.
- Smarter stratification and sample allocation to achieve better precision from a given effort. For example: change sampling densities within strata of existing surveys in response to expected changes in distributions, which may require consideration of trade-offs among species with differing distributions and life histories that are indexed by the same survey
- How do expand surveys into the new areas where species are moving to? Stretch existing sampling effort? or do something different?

Combining surveys/data sources to improve survey data products.

- Advice on methods to estimate how much overlap in time and space is needed in order to estimate calibration factors (mainly gear effects) between different surveys/data sources with sufficient precision?
- Identify problems and explore solutions to situations when parts of the survey area are not covered at all for some periods of time, since the combined time-series are often of different length leaving gaps to be filled out by the model. In such cases the results are often much more sensitive to the details of the model formulation, both with respect to the abundance estimates but also CVs.
- Developing methods for dealing with elimination of entire survey areas (e.g. due to area closures to trawling, wind farms, high commercial fishing densities)
- Recommend potential calibration studies (with the respect to the areas and sample sizes) across surveys in ICES countries. Such studies are often are easy and relatively inexpensive compared to survey costs.

Tools and technology development:

- Initiate processes to develop universal survey analysis tools to evaluate and process survey data. Focus on 2 types of evaluation tools: 1. independent from assessment models (new estimation methods would require development of new tools), 2. within the models that are used to create the indices. The output of this evaluation would be considered as feedback to the data collectors about survey utility and priorities. Assure that tools can be used across different survey types and databases (e.g. R packages, augment existing code from researchers and make it universal for the processing survey data and/or estimating reductions impacts. Make analytical studies easier. Continue to streamline the process of QC of the survey data.
- Develop new and improve existing packages for OMs and EMs to easily simulate surveys and assess performance of survey data products.
- Technological developments to improve ability to collect more data with higher accuracy (automated age reading),
- Decision trees for survey managers in areas and for survey types which were not covered in subgroup meetings – The discussion can concentrate mostly the fixed station design and index stations survey. The options could include building models to estimate areal availability to correct indices. There are some tools we can use to assess the leverage of a particular position on the index which can help provide some idea of how to prioritize stations to maintain contrast and consistency of surveys. Incorporating these in the decision trees developed during WKUSER1 maybe preferable as at least the impacts and the surveys responses fall into the same category. It just you would use different tools to do the evaluation.
- Use simulation studies to explore causes of additional variability estimated within assessment models.
- Develop technological methods to obtain absolute estimates of biomass or abundance to calibrate existing surveys and obtain estimates of survey catchability and variation in catchability - (comment from Sven: There is a sizeable literature on this the Scotts have attempted to estimate catchability experimentally and they use this in their monkfish survey. However it seems despite this there are issues with catchability estimates at least our beam trawl survey has a higher catchability than uncorrected that their corrected CPUE.)
- Open codend tows for some stations. Considerations of technologies to reduce number of tows to be processed by hand.

Modelling and simulations:

- Increase ecosystem data collection during survey for use in model-based estimation and process studies

- Continue exploration of model-based estimation in challenging situations. Consider all survey data streams (i.e. impacts on age/length comps in addition to just abundance indices).
- Explore the benefits and risks of the resampling and simulation studies for surveys. Ideally they would provide similar results. If this is the case, then either is an option, if not under what circumstances would one be preferable over the other?
- Explore the multivariate approach (redundancy analysis)vs.the' multivariable' VAST approach. One is much simple and quicker and can easily handle all the species, not just the ones currently used in assessment, but may lack the necessary detail, while a good vast model will be more precise, but more time consuming and less inclusive.
- Propagate uncertainty through the full assessment process to assess tangible impacts on stock status and harvest recommendations/reference points.
- Multispecies/multi-objective optimization, trade-offs, approaches. How to agree on a common currency for what you are optimizing (what is the composite metric)? How do you get stakeholders to agree on weights for different species? Need of increased/appropriate feedback from stakeholders (e.g. year to year). Look at the multivariate matrix to weigh between environmental and biological objectives. How do we balance survey objectives (e.g. focus particularly on less abundant species, environmental information, etc) For example: approach could consider: 3 axes of optimization: 1. Risk – what aspects of life history may require more precise information for sustainable management (e.g. some of our more K type long-lived species are at greater risk of overexploitation compared with more r type and periodic species.). 2. Value – The value of a fishery has always been a consideration, but perhaps it does not need to be the most important. (e.g. The most valuable species are (in the US) the species with the most robust management (i.e. pollock), and improving data streams for poorly understood or indexed choke species might be more important.). 3. Current status – If a stock is well-managed and harvested at a sustainable level, we probably don't need to put a premium on reducing variance for that species.
- Propose objective functions to optimize survey design. Objective functions and design preferably should be flexible enough to allow for changes in priorities of all important aspects of a multispecies and ecosystem survey. One option would be to suggest formulating multiple objective functions with emphasis on different uses of the survey data. Sampling the survey area more or less evenly across space may avoid many problems and pitfalls when the data are to be analysed, so adaptive sampling should be used with great care in my opinion
- Improve evaluation of value of ecosystem data collected during surveys. Assessing alternative survey designs in the context of ecosystem information is a formidable challenge. On the stock assessment survey end of the spectrum, where most of our analysis has been focused, we actually have it quite easy. We know how the data are used to inform management through the specifics of the stock assessment and how it treats uncertainty. However, we don't have great metrics for calculating the value of ecosystem survey information, or how reductions in effort or change in design impact our ability to describe marine ecosystems and climates. While some work has been done, this is an area where we need to expand our simulation/analysis efforts. Another fundamental challenge is that unlike stock assessment data inputs, we don't always know the ecosystem-based questions we will be asking in future.
- Test model-based flexibility to different sample allocation strategies. Flexible designs should provide the same expected values but can differ in CVs.
- Testing AI or other multivariate approaches to discover predictable spatial relations and estimation for advisory products.
- Look for correlations between bottom type covariates and densities (occurrence) or catchability. Collect easy to collect data. Pictures of the bottom. Process on the spot automatically.

Logistics, admin

- Identify decision-makers to get feedback from to clarify objectives and priorities in conjunction with development of methods to evaluate survey changes across multiple objectives, augment a structure of advisory process to assure feedback to survey groups (e.g. SSC feedback provided to AFSC survey group in 2018) on priorities with respect to importance of the areas and species to direct survey effort to accordingly to the needs.
- Limited number of staff to do the work. This can affect survey groups ability to complete full surveys.
- Organizational divide between data collections and assessment programs. Organizations should create mechanism for the survey and stock assessment scientist to have productive communication and collaboration.
- Logistical challenges with moving to new areas. Developing new surveys with very limited information on the distribution of species in previously unsampled areas and habitats.
- Problems with fish moving across survey and country boundaries creates logistical challenge because not all surveys can be easily modified due to jurisdictional issues.
- Limited number of research boats and boats available for charter. Boats bidding for charters are old.

3.2 WGISDAA conclusions on WKUSER

The group concluded the workshop on unavoidable survey effort reduction had been a great success laying a foundation for supporting the integration of survey decision-making with the advisory processes. The result of the workshop was the establishment of a process by which these objectives are to be achieved. Much work still remains in the full implementation and therefore WGISDAA strongly supports the development of a future workshop to compile the outcome of work to be carried out in response to the first workshop with a greater detail on the tools to be applied to the processes. A recommendation for a future workshop will be developed based on the reported options in conjunction with the next WGISDAA chair and submitted separately once finalized.

4 Solving the retrospective issue; Working towards a benchmark for North Sea Cod

4.1 Background

In recent years, assessments of North Sea cod have shown a persistent downward revision of SSB and upward revision of fishing mortality (*F*) from year to year (Figure 4.1). This retrospective pattern is caused by lower catch rates of older fish in the NS-IBTS surveys compared to commercial catches and, due to SSB falling below biomass reference points, has resulted in large cuts to the TAC and advice. The problem first presented itself in the 2018 assessment of North Sea cod (ICES WGNSSK, 2018), with the stock put forward for benchmark when the issues remained in 2019 (ICES WGNSSK, 2019). A benchmark will take place November 2020–February 2021 (WKNSEA 2021).

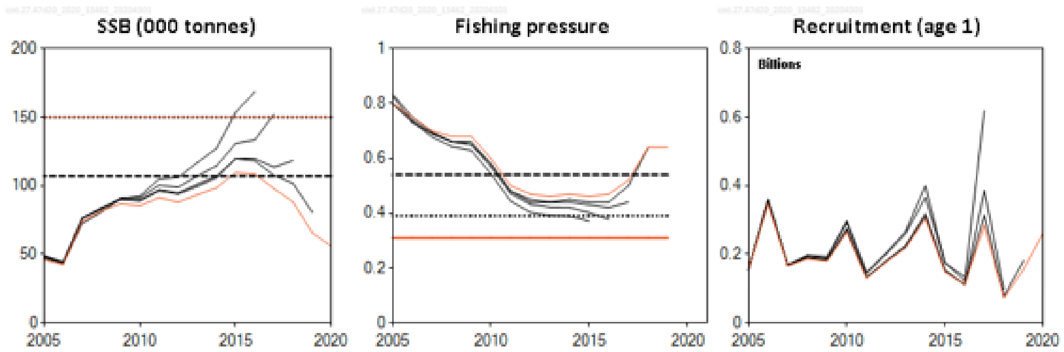


Figure 4.1. North Sea cod assessment results from the last five years. The downward revision of SSB from 2016 to 2017 is a consequence of re-estimating maturity-at-age. From 2017, revisions are caused by the addition of new data.

WKNSSK have conducted several analyses that suggest the retrospective pattern in the assessment of North Sea cod relates to the survey and commercial catch data (ICES WGNSSK, 2018, 2019, 2020). The evidence includes:

- Sequential removal of new survey index points and commercial catch data returns SSB to the level of the previous assessment. The 2018–2019 assessments showed the revisions to be caused primarily by the survey data, while it was primarily the catch data in 2020.
- A loss of cohort signal between the 2012- and 2013-year classes common to both fishery dependent and independent data sources.
- A discrepancy between the negative gradients of catch curves from fishery dependent and independent data sources, with recent commercial gradients representing some of the lowest values in the time-series while survey gradients have shown substantial increases (Figures 4.2,4.3).
- SAM assessment residuals for the IBTS-Q1 in 2018–2019 (bar age 1 in 2019) and the IBTS-Q3 in 2017–2018 (bar age 1 in 2017) are all negative (Figure 4.4).

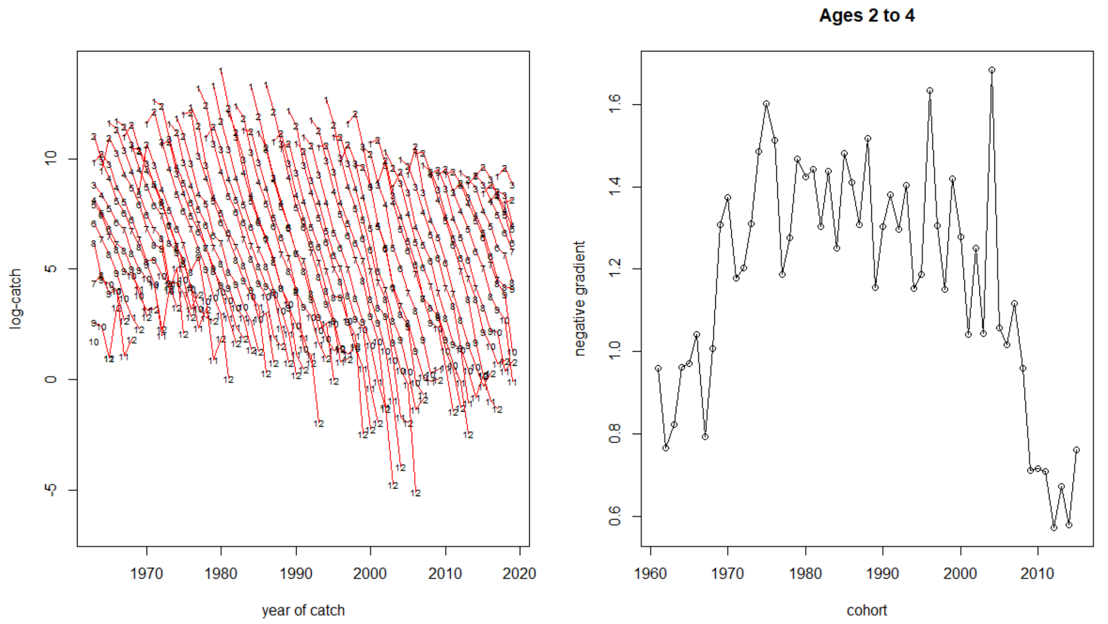


Figure 4.2. Log-catch cohort curves (left) and associated negative gradients for each cohort across the reference fishing mortality age of 2–4 (right).

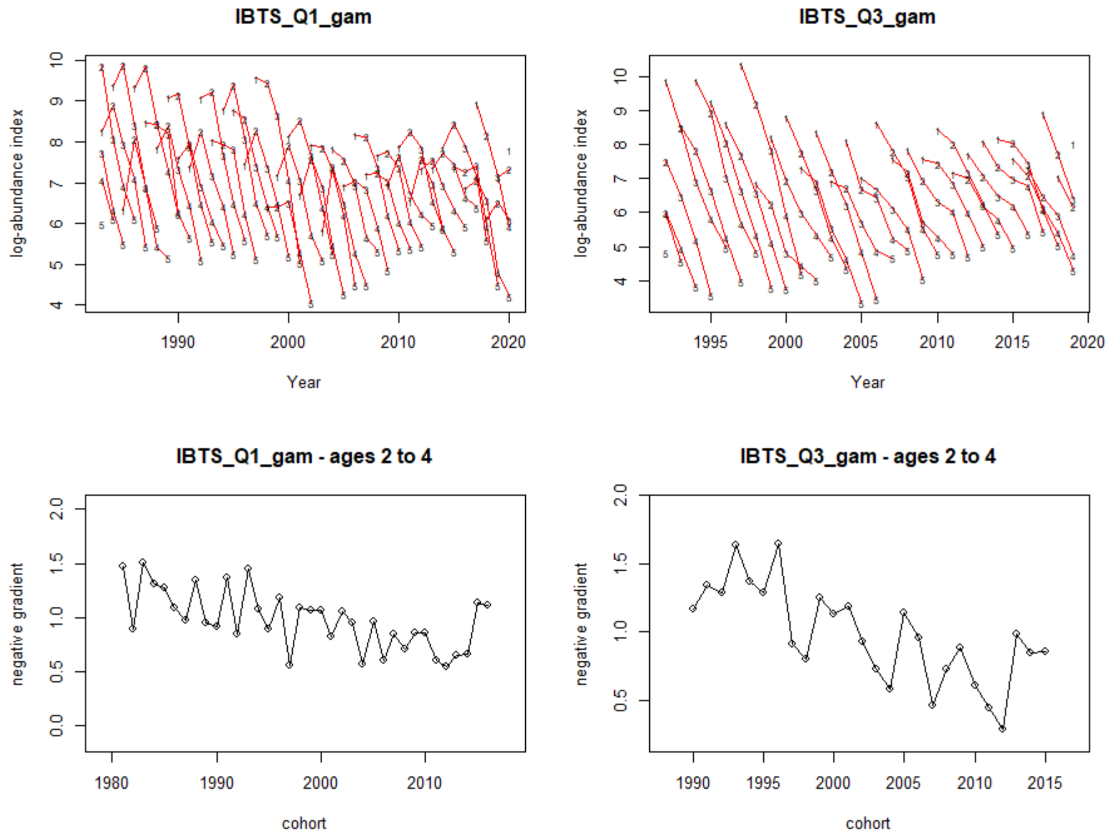


Figure 4.3. Log-abundance curves (top) and associated negative gradients for each cohort across the reference fishing mortality age of 2–4 (bottom) based on the IBTS-Q1 (left) and IBTS-Q3 (right) survey indices.

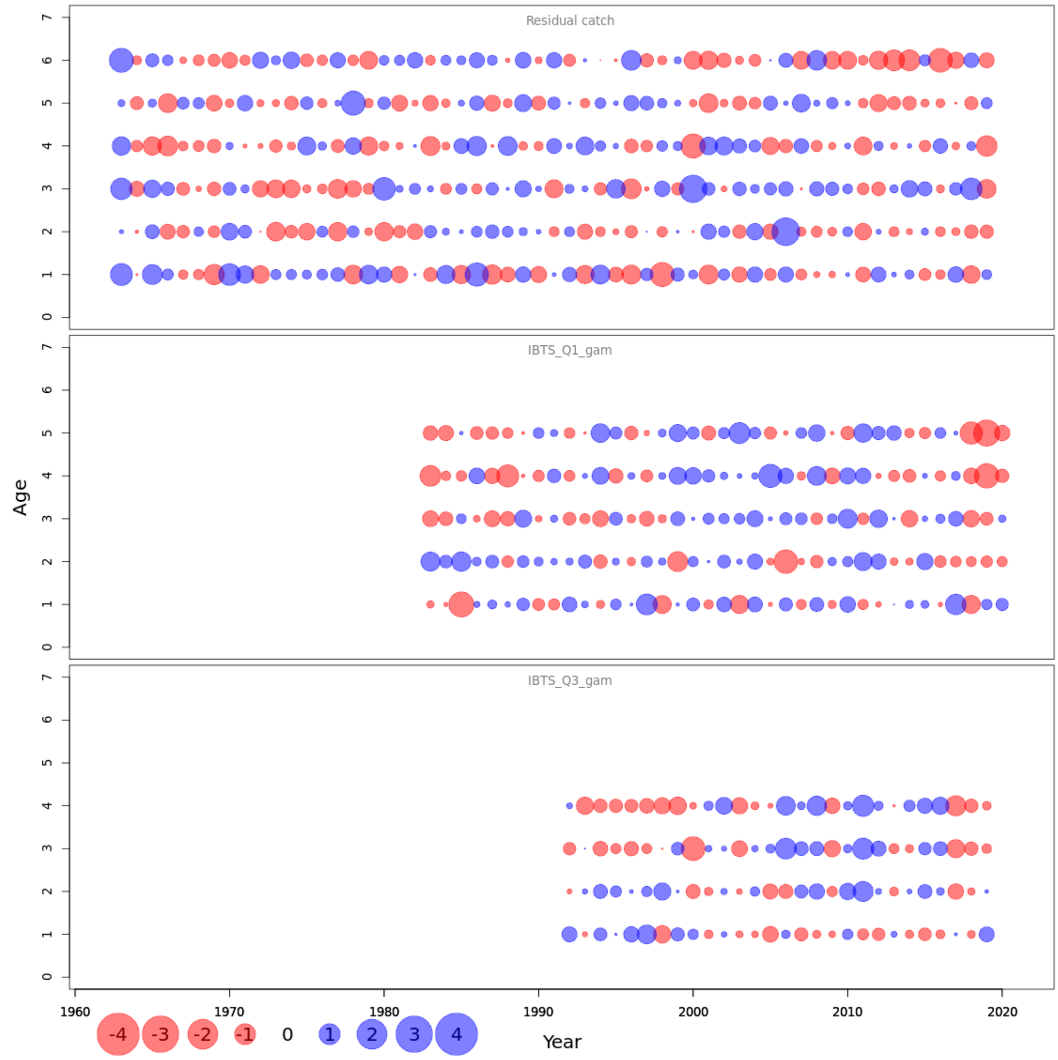


Figure 4.4. Normalized residuals for the SAM assessment for total catch and the IBTS-Q1 and IBTS-Q3 survey indices.

4.2 Summary of WGISDAA contributions

Issues in the North Sea cod assessment relating to the discrepancy between survey and commercial data were first presented to WGISDAA in 2018 (ICES WGISDAA, 2018). Since then then there has been a two-way communication between WGISDAA and WGNSSK where the latest developments in the assessment are presented to WGISDAA and WGISDAA offers suggestions and exploratory analyses that feed back into the assessment and benchmark investigations. Analyses conducted, or stemming from discussions, at WGISDAA have fed into several related workshops including the Workshop on North Sea Stocks Management Strategy Evaluation (ICES WKNSMSE, 2019), the Workshop on Catch Forecasts from Biased Assessments (ICES WKFORBIAS, 2020) and the Workshop on Stock Identification of North Sea Cod (ICES WKNSCodID, 2020).

During the 2018 meeting, WGISDAA reviewed the assessment and carried out further work to investigate the reported issues focusing primarily on the survey data. This work is reported in detail in the WG report (ICES WGISDAA, 2018) and is summarized in Table 4.1 but included further examination of survey data and indices and exploratory assessment runs. These analyses found year effects in the recent survey indices and, although the underlying mechanisms remained unresolved, were thought likely to be caused by changes to survey catchability. This conclusion, and an exploratory assessment configuration that correlates errors between age classes to simulate year effects in the surveys, was taken forward as an alternative operating model for management strategy evaluations with the result that more precautionary management was needed in this case (ICES WKNSMSE, 2019).

The 2019 meeting of WGISDAA focused more on the commercial catch data and the potential for changes to fishing practices and selectivity; which may, for example, have occurred following introduction of the landings obligation and discontinuation of the fishing effort regime and cod specific fully documented fisheries (FDF) scheme. Removal of the conflicting catch data (2017–2018) from SAM inputs made little difference to assessment results, suggesting that the SAM model may be too constrained in *F*. Two exploratory assessment configurations that relaxed the assumptions on *F* both resulted in more variable *F* processes but with neither resolving the retrospective pattern or residual problems. Several configuration options and new functionalities of both the surveyIndex package (used to produce indices for the assessment) and SAM assessment model were highlighted and will be considered in the forthcoming benchmark. In addition, preliminary work in preparation for the Workshop on Catch Forecasts from Biased Assessments was presented to WGISDAA and benefited from the suggestions of the WG (ICES WKFORBIAS, 2020). A full list of investigations and suggestions from the 2019 meeting of WGISDAA is given in Table 4.1.

The Workshop on Stock Identification of North Sea Cod (WKNSCodID) in 2020 presented an opportunity to revisit some survey-based analyses suggested by WGISDAA in 2018. The analysis is described in full in the workshop report (ICES WKNSCodID, 2020) but, by combining data from three surveys in a delta-GAM model, suggests considerable migrations of older cod from the North Sea to the West of Scotland. In addition, some exploratory assessments conducted by WGNSSK were presented to WGISDAA in 2020. These assessments followed from the comments of WGISDAA the previous year (that correlations in the year effects model are weak) and the work of WKFORBIAS which found the year effects model unable to sustain the observed retrospective pattern in MSE simulations (ICES WKFORBIAS, 2020). The exploratory assessments split each of the survey indices into two periods resulting in a more parsimonious assessment, reduced Mohn's rho (retrospective pattern) and more pronounced correlations for the later index series, with some significant differences in estimated catchability parameters between the two survey periods. Together, these results suggest that there has been a change in survey catchability and that this may have been caused by migrations of older cod outside the assessment area. Based on these results, WGISDAA concluded that it would be preferable to get the stock definition right by potentially including part of catches in division 6.a in the assessment rather than fixing the SAM configuration to attempt to account for missing fish via correlated survey errors.

4.3 WGISDAA conclusion

WGISDAA welcomed the opportunity to contribute to the analysis through comments and information requests but recognizes that it does not have the role to take decisions in the advisory processes so takes no formal opinion on the appropriate form of a future assessment. We are pleased to note that the analyses conducted in conjunction with WGNSSK will be considered in the upcoming benchmark.

It is however noted that the situation here is far more complex than simply suggesting the index is flawed. The assessment and index solutions to the retrospective problem tested all find different ways to resolve the conflict in the data, with those options that reflect the true underlying processes likely to prove most robust for future advice. It is vital to examine these processes external to the assessment as carried out in this cooperation. For example, a spatial change in the distribution of the stock beyond the management area will result in changes in catchability in the survey but is also likely to do the same for the fishery. Adding time varying catchability to both elements will reduce the retrospective pattern, but unless the relative strength of these effects can be determined reliably it is unlikely to resolve the issue long term. Taking a more process-oriented approach by redefining the biological units at least for the index and potentially for the fishery is likely to provide longer term solutions.

Table 4.1. A summary of WGISDAA discussions and outcomes from the 2018–2020 meetings.

Investigation or suggestion	Results, conclusions, and further analyses
WGISDAA 2018 (ICES WGISDAA, 2018)	
Examination of survey CPUE plots.	The highest concentrations of older cod are found near the border of the assessment area towards the west of Scotland. There are no clear outliers suggesting the results are not driven by a single observation or nation.
Combine NS-IBTS data with SWC-IBTS data to verify if migrations in and out of the assessment area are causing year effects in the survey indices.	The required SWC-IBTS data for 2017–2018 were not available in DATRAS at the time. The analysis was conducted in 2020 and showed increased connectivity with 6.a (ICES WKNSCodID, 2020).
Examination of log-mean standardized indices by year and cohort.	Possible year effects, particularly for the IBTS Q1 which shows a peak for most ages in 2017 followed by a subsequent decline. Aligning by cohort shows some loss of signal between the 2012 and 2013 cohorts.
SURBAR survey-based assessments (i.e. without commercial catch data).	Residuals from the 2018 SURBAR assessment suggest the IBTS-Q1 2017 data are over-optimistic (positive residuals) and the IBTS Q3 2017 pessimistic (mostly negative residuals), suggesting year effects may be present. Subsequent SURBAR assessments conducted by WGNSSK continue to find positive residuals in the IBTS Q1 in 2017 and negative residuals in the IBTS Q3 from 2017 onwards (ICES WGNSSK, 2019, 2020).
Conduct a separable-VPA as an additional check of the commercial catch data.	The analysis presented no large step changes in fishing mortality.
Re-run the assessment leaving out (1) the IBTS Q3 (2016)–Q1 (2017) data (high survey catch rates) and (2) the IBTS Q3 (2017)–Q1 (2018) data (low survey catch rates).	Both options resulted in slight improvements to the residuals (although patterns were still present), with the first option producing a more pessimistic assessment and the second option a more optimistic assessment in line with the assessment from 2017.
Explore assessments more robust to year effects by correlating errors on the survey indices of older age classes.	Effect shown to be significant and results in a more parsimonious model but makes little to no improvement to the survey residuals. WKNSMSE took a similar assessment configuration forward as an alternative operating model for management strategy evaluations (ICES WKNSMSE, 2019).

Performance of a robust assessment to check restrictiveness of smoothing in the accepted (May 2018) assessment. Mixture distributions consisting of 10% t3-distribution and 90% normal distribution were assumed for log F, N and survey and catch observations, allowing for jumps in these processes and observations.	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.
WGISDAA 2019	
Examine time-series of mean size at age in the surveys to detect potential aging issues.	Unresolved.
Investigate fitting the delta-GAM, used to derive survey indices, with a non-stationary spatial distribution.	Specifications of the delta-GAM are currently being investigated for the benchmark and some preliminary results are presented below.
Examination of commercial log-catch cohort curves and consistencies.	There is strong and significant cohort consistency in the catch when considering the full time-series (from 1963). This becomes weaker when considering only a recent period (from 2007).
Procedures to raise commercial catch data may need revisiting to account for potential changes to discarding practices since introduction of the landings obligation and discontinuation of the fishing effort regime and FDF schemes.	Fishing mortalities, split between landings and discards, derived from the assessment show landings to have increased while discards have remained constant or decreased. Raising procedures will be considered during the benchmark.
Re-run the (May 2019) assessment without commercial catch data for 2017–2018.	This made very little difference to the assessment results and did not solve the residual problem, suggesting that the assessment model is too constrained in *F* as it cannot jump even when the conflicting catch data are removed.
Re-run the assessment with independent random walks for fishing mortality (i.e. removing correlations between ages so the assessment can potentially react quicker to changes in selectivity).	This worsened the model diagnostics but did show more variation in selectivity compared to the accepted (May 2019) assessment. Residual patterns were still present.
Approximate an XSA assessment with SAM, by setting the catch variance to zero and the F random walk variance to be large, to see if F jumps.	This resulted in a more variable F process and a notable upward revision of SSB and downward revision of F in recent years compared to the accepted assessment. The XSA approximation still had retrospective and residual patterns but less pronounced than in the accepted assessment.

Explore new SAM functionalities including the option to input index uncertainty and new variance and covariance options in the configuration.	The option of unstructured catch correlations was attempted but failed to run. An update of the survey-year effects model of WKNSMSE was presented but did not estimate strong correlations between ages. SAM configuration options will be explored as part of the benchmark process.
Survey index correlations are likely weak because they are estimated across the full survey time-series while we have only noticed potential year effects recently. It could be considered to implement selectivity blocks within SAM.	An exploratory assessment conducted at WGNSSK in 2020 split each of the IBTS indices into two time-series (pre- and post-2010). This resulted in a more parsimonious assessment, reduced Mohn's rho and more pronounced correlations for the later index series, with the estimated parameters suggesting some change in survey catchability.
Use the MSE framework of WKNSMSE to simulate process changes and evaluate how the assessment performs when these processes are ignored in the assessment. For example, *M* is updated only every three years so any sudden changes to natural mortality not accounted for in the assessment may contribute to the retrospective pattern.	Unresolved.
Discussions on work being prepared for WKFORBIAS.	Analytic estimates of Mohn's rho calculated from raw data (i.e. accounting for re-smoothing of maturity, re-running of delta-GAM indices and periodic revision of multispecies-derived Ms) are higher than those based on a peel of the assessment inputs, suggesting that the true level of retrospective bias in the assessment procedure is currently underestimated (ICES WKFORBIAS, 2020).
WGISDAA 2020	
Investigate the distribution of the fishery in relation to survey distributions, as accessibility to the fishery may affect catchability.	Survey and commercial distributions were explored by the recent Workshop on Stock Identification of North Sea cod (ICES WKNSCodID, 2020).
Investigate/modify SAMs variance assumptions around the commercial catch data.	Unresolved / SAM configuration options will be explored as part of the benchmark process.
Given delta-GAM derived survey distributions suggest connectivity of the North Sea stock with the west of Scotland stock, it would be preferable to include part of 6.a in the assessment definition rather than fix with correlated survey errors.	WKNSCodID recommended that the connectivity of the 'northern inshore and offshore components' of cod in 6.aNorth and 4.aWest should be considered in a future benchmark assessment workshop.

5 A combined index for whiting in Division 6a from the Scottish and Irish Q4 surveys

5.1 Introduction

The assessment of whiting in Division 6a is carried out yearly with catch and survey data (ICES, 2020a). 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 Scottish surveys have been conducted since 2011 using a stratified random sampling design. The data from the three surveys were used until recently (prior to 2020) to produce three indices for the stock.

During the ICES Inter-Benchmark Protocol for West of Scotland Roundfish (IBPWSRound) 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 5.1). It was hoped that a combined index could provide a more robust and precise index. A simple 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 individual indicies suffered from low internal consistency and high variability observed resulting in low influence in the assessment. Eventually, the IBPWSRound concluded that the Scottish and Irish Q4 surveys should be retained as separate indices.

With more data collected in subsequent years and being available for further analysis, the concept of using a combined index was revived within the Working Group on Improving Use of Survey Data for Assessment and Advice (WGISDAA) in 2018 (ICES, 2018). This issue became important in view of a planned benchmark. During the Benchmark Workshop for Demersal Species (WKDEM) that followed two years later (ICES, 2020b), it was agreed to use the combined Q4 index. The index was also used by the Working Group for Celtic Seas Ecoregion (WGCSE) in 2020.

This document presents the results of the analysis of the Scottish and Irish Q4 surveys. The analysis delivers a combined index to be used in annual assessments of the whiting stock.

5.2 Data and analysis

Data

Data for the analysis were downloaded from DATRAS for the period 2011–2019. They included haul data (date, vessel, haul time, tow duration, depth, latitude and longitude), fish numbers-at-length and age-length keys (ALKs). Two vessels were operating in the area at that time, the Scottish and Irish one, both using the GOV trawl. Common ALKs were assumed for the two surveys as the confidence in precise age determination in both was high, and the increase in the number of samples available improved precision.

GAM analysis

For the GAM analysis, hauls were selected that were taken in the southern part of Division 6a (Figure 5.1). In this area, comprising of 16 ICES rectangles, both surveys were operating and it is referred to here as the "common area". In 2013, the Scottish survey was not fully conducted and it covered only the northern half of the division. Consequently, this year was excluded from the analysis. This resulted in 497 valid hauls that were subsequently subjected to modelling. The following variables were included in the model: year, vessel, haul position (longitude and latitude), depth (in meters), haul time (Coordinated Universal Time, UTC) 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).

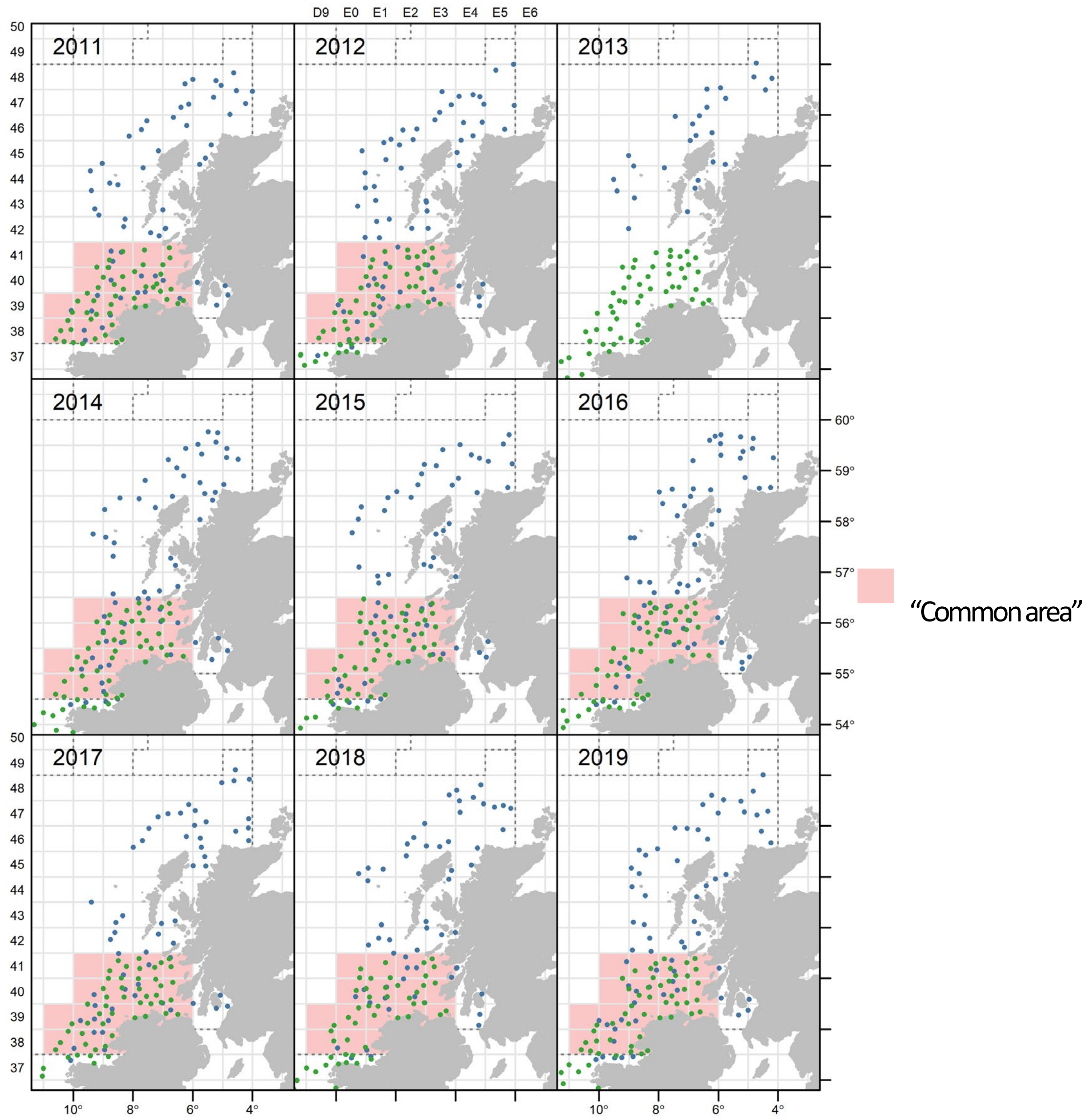


Figure 5.1. The location of hauls in the two Q4 surveys: UK-SCOWCGFS-Q4 (in blue) and IGFS-WIBTS-Q4 (in green), in 2011–2019. The shaded red area marks the statistical rectangles where both surveys were operating ("common area").

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 distribution was found automatically during the model optimization. The model was run separately for age groups (0, …, 6 and 7+), as well as for an additional aggregate age group (1+). The age-0 group was omitted in this aggregation as it usually shows distinct characteristics before the settling transition.

Theestimates surevy catch numbers (N_i) were the response variable in the model. To account for the differences in tow duration, numbers were 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, time of day ($Time_i$), depth, longitude and latitude:

$$N_i \sim NB(\mu_i, k) \quad (1)$$

$$E(N_i) = \mu_i = e^{g(x_i)} \quad \text{and} \quad \text{var}(N_i) = \mu_i + \frac{\mu_i^2}{k} \quad (2 \text{ and } 3)$$

where

$$g(x_i) = \log(Td_i / 30) + \alpha + \beta_1 \times Vessel_i + \beta_2 \times Year_i + f_1(Time_i) + f_2(\log(Depth_i)) + f_3(Lon_i, Lat_i) \quad (4)$$

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 $Time_i$; f_2 is the smoothing function of the log-transformed depth, $Depth_i$; f_3 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 is done by maximizing the likelihood (delivering maximum likelihood, ML; Wood, 2006).

In the next step, a Chi-squared 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 differences were constructed by applying a bootstrap procedure.

Index calculation

The indices in the Irish survey were provided by the Marine Institute in Ireland. The index in the Scottish Q4 survey was calculated using the following procedure (also being applied for UK-SCOWCGFS-Q1).

Numbers at length (the length frequencies, LF) per haul are standardized to numbers per hour towing. In the past (prior to 2011), all otoliths from all hauls in a given demersal sampling area were combined to create an age length key for that area (Holmes, 2008). With the new survey design, all otoliths taken within each of the eleven strata are combined to form an ALK. This ALK is applied to all LFs in the stratum individually to produce age frequencies for each haul. Then, for each stratum, the age frequencies are summed and the values divided by the number of valid hauls to provide numbers at age per hour. This procedure can be summarized as

$$CPUE_{i,a} = \frac{\sum_{h=1}^{H_i} \sum_{l=l_{\min}}^{l=l_{\max}} N_{i,a,l,h}}{H_i} \quad (5)$$

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

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

$$I_a = \frac{\sum_{i=1}^{S} CPUE_{i,a} A_i}{\sum_{i=1}^{S} A_i} \tag{6}$$

where A_i = area (m^2) of stratum i and S = number of strata

To combine the Scottish and Irish Q4 surveys, the hauls therein were initially taken together. They were considered to form one dataset. The age frequencies $N_{i,a,h}$ in Equation (5) were calculated in the same way for each haul, irrespective of the survey. Then, the frequencies in the Irish hauls were modified by using the ratios of catch rates established through the GAM analysis. As a result, the final modified index could be derived as in Equation (6).

5.3 Results

CPUE and length distributions

Some differences in the whiting CPUE with the observed data can be noted between the two surveys in the "common area". For fish at age 0–1 and to a lesser degree for age 2, the CPUE tended to be higher in the Irish survey (Figure 5.2). Nonetheless, the overall trends were rather similar in the two surveys for all age groups. Also, no considerable difference was observed for the aggregate age group 1+.

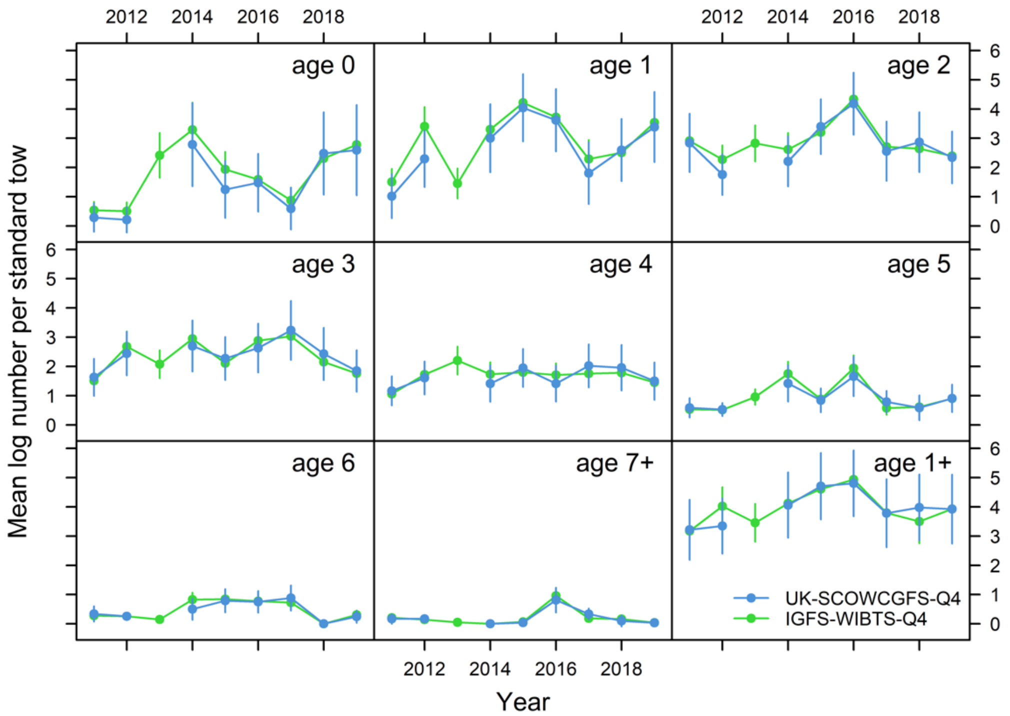


Figure 5.2. The observed CPUE of whiting (log values with 95 % confidence limits) by age in the two surveys in the "common area" in 2011–2019.

An examination of the observed length distributions showed a shift in time between the Irish and Scottish surveys, with the former taking place somewhat earlier (Figure 5.3). It was most pronounced for the smallest fish. For the bigger fish (over 20 cm in length), it was noticeable only in some years. In general, the length distributions were rather similar in shape. The difference in catch rate between the two surveys was most pronounced in the first few years of the time-series.

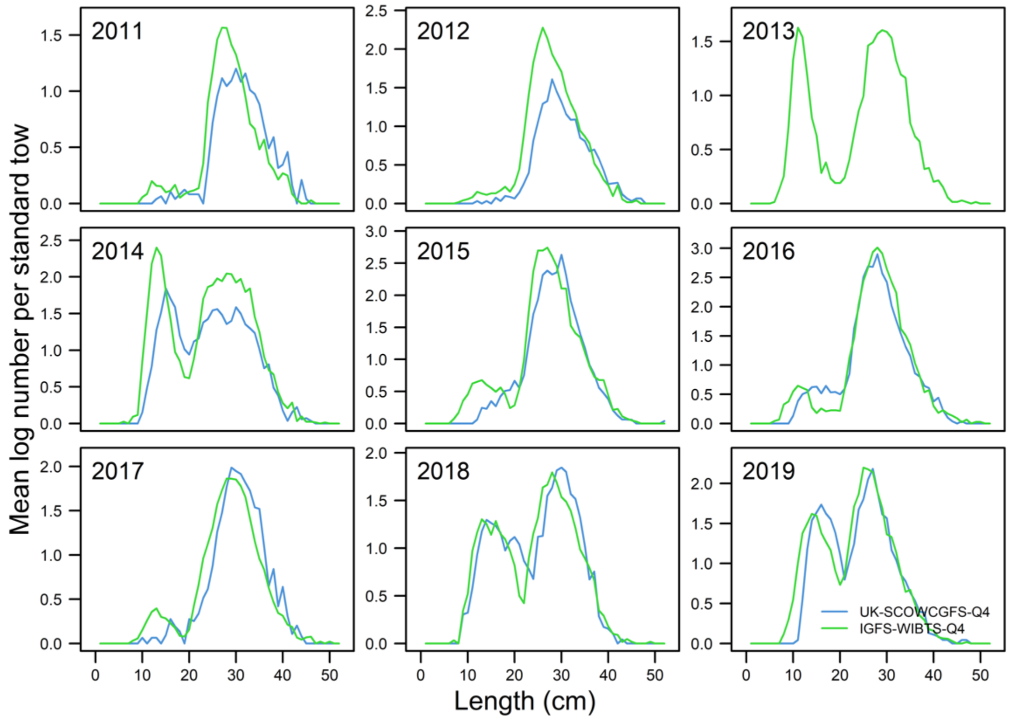


Figure 5.3. The length frequency in the two surveys in the "common area" in 2011–2019.

GAM analysis

The GAM analysis provided more insight into the actual difference between the two surveys. The dispersion parameter in the negative binomial distribution varied within a wide range from 0.28 to 1.59 (Table 5.1). Overall, the 0-group tended to form aggregations which is demonstrated by a low value of the dispersion parameter. This tendency clearly 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 for them was that whiting densities gradually decreased along the western edge of the survey area (not shown). The effect of *Time*was generally not significant except for age 6.

The observed mean difference in catch rate between the Scottish and Irish surveys was negative for ages 0–2 (meaning higher catch rates in the Irish tows) and showing no difference for older fish (Figure 5.3). The model-estimated catch rates were found to be consistently higher in the Irish survey for all age groups. The modelled effect of vessel (the difference in catch rate between the Scottish and Irish surveys) varied across the age groups, being the lowest for age 0, increasing for ages 1–3 and changing little for older fish. In relative terms, the proportion of catch rates in the Scottish survey to those in the Irish survey varied between 0.5 and 0.8, depending on the age group (Table 5.1). For the aggregate age group 1+, it was 0.7. The model explained 53–73% of the deviance, with decreasing (with fish age) ability of the model to explain the variation in the data.

Table 5.1. Summary of the GAM analysis.

Age group	Dispersion parameter	*Time* significance	log(*Depth*) significance	*Lon* × *Lat* significance	*Vessel* effect	*Vessel* effect significance	exp(*Vessel* effect)	Deviance explained (%)
0	0.279	0.795	<0.001	<0.001	-0.782	0.800	0.46	73.4
1	0.526	0.402	<0.001	<0.001	-0.504	0.086	0.60	65.0
2	0.605	0.412	<0.001	<0.001	-0.283	0.501	0.75	60.6
3	0.649	0.382	<0.001	<0.001	-0.212	0.770	0.81	57.7
4	0.732	0.337	<0.001	<0.001	-0.187	0.774	0.83	52.8
5	0.820	0.403	<0.001	<0.001	-0.262	0.887	0.77	56.1
6	1.286	0.035	<0.001	<0.001	-0.188	0.978	0.83	57.4
7+	1.589	0.475	<0.001	<0.001	-0.240	0.808	0.79	62.0
1+	0.605	0.444	<0.001	<0.001	-0.300	0.434	0.74	60.1

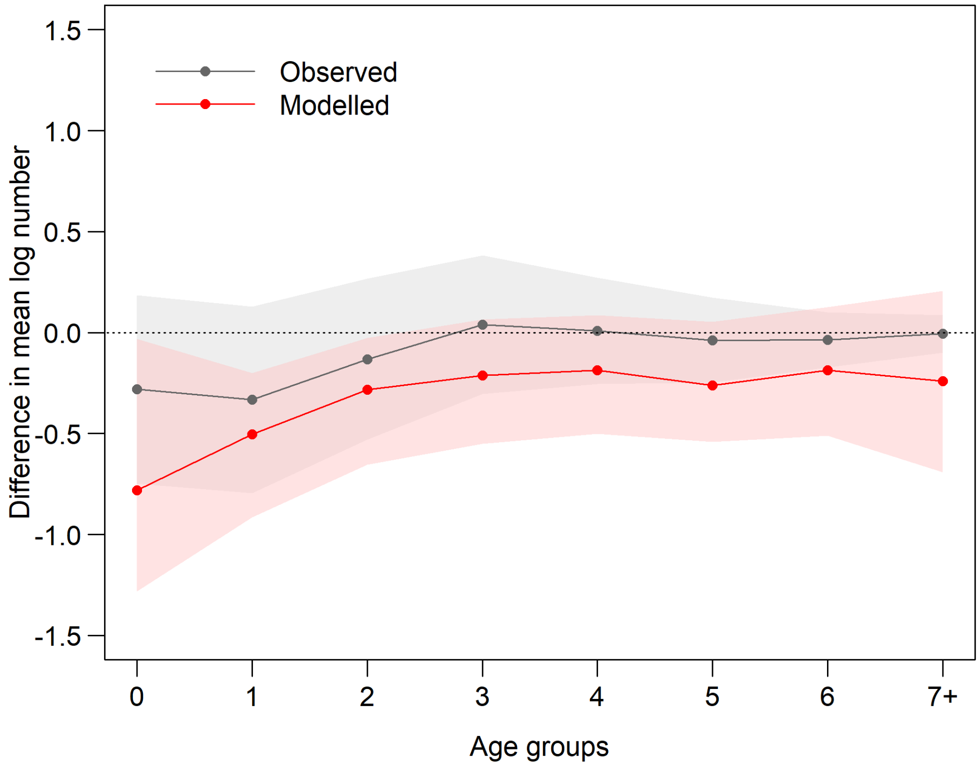


Figure 5.4. The observed and modelled difference in CPUE of whiting (with 95 % confidence limits) by age between the two surveys in the "common area".

The vessel effect modelled with GAM was found non-significant for all age groups (Table 5.1). However, the confidence intervals obtained with the bootstrap method indicated significance of the effect for ages 0–2 (Figure 5.4).

Index diagnostics

The combined index improves internal consistency and avoids the apparent conflict between the separate indicies likely caused by spatial shifts in the population. It is therefore and can be considered more informative of the population densities. This is understandable as the combined dataset was larger and covered a large sampling area within the division. Also, a combined index for 2013 could be produced as the coverage of the sampling area was almost complete in that year (Figure 5.1).

Figure 5.5 compares the four tuning series: IGFS-WIBTS-Q4, UK-SCOWCGFS-Q1, UK-SCOWCGFS-Q4 and IGFS-UK-SCOWCGFS-Q4. For the 0-group, the index was similar for the three Q4 series regardless of the area being surveyed. For older fish, the Irish index tended to be higher compared to the indices for the whole survey area. There was a general consistency in indices among the surveys for the whole survey area, with the Scottish Q1 index being overall the highest.

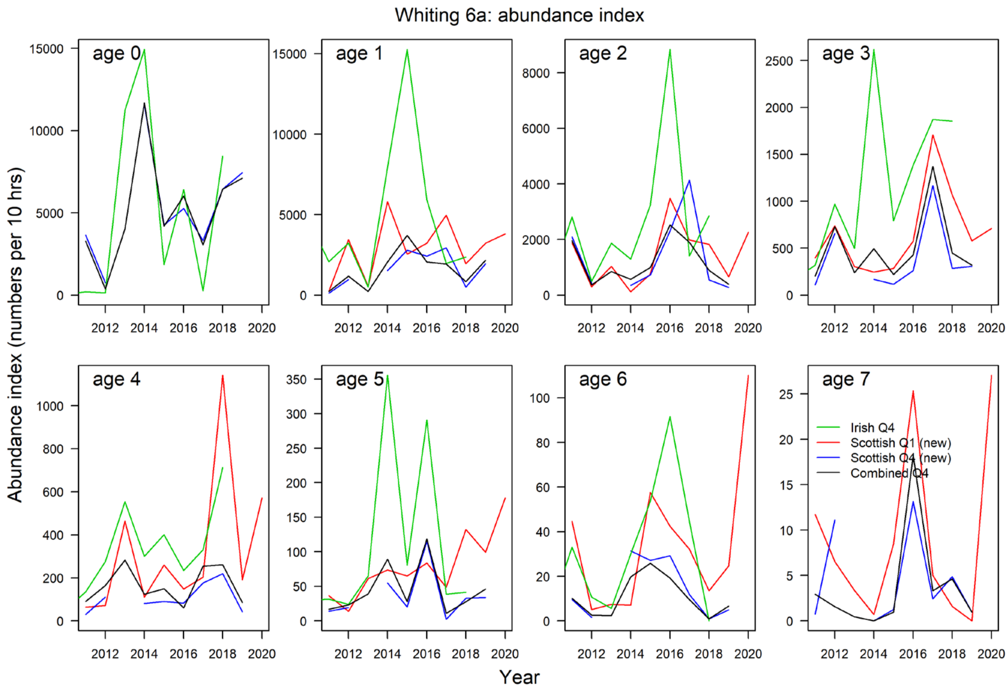


Figure 5.5. Abundance index in the four tuning series: IGFS-WIBTS-Q4, UK-SCOWCGFS-Q1, UK-SCOWCGFS-Q4 and IGFS-UK-SCOWCGFS-Q4.

The mean standardized catch proportions at age per year showed some similarities among the three Q4 tuning series (Figure 5.6). The plots indicate strong year classes (2009 and 2014 year-classes), but also consistently weak year classes (2012 year-class). In most cases, year class tracking was reasonably consistent.

Figure 5.7 shows the log mean standardized indices in the three survey series (IGFS-WIBTS-Q4, UK-SCOWCGFS-Q4 and IGFS-UK-SCOWCGFS-Q4) by year class and year. There were noticeable differences among the survey series with the Irish survey tending to be noisier and strong/weak year-classes were more difficult to identify. The combined index performed best – with the year classes being tracked relatively well and with a discernible year effect.

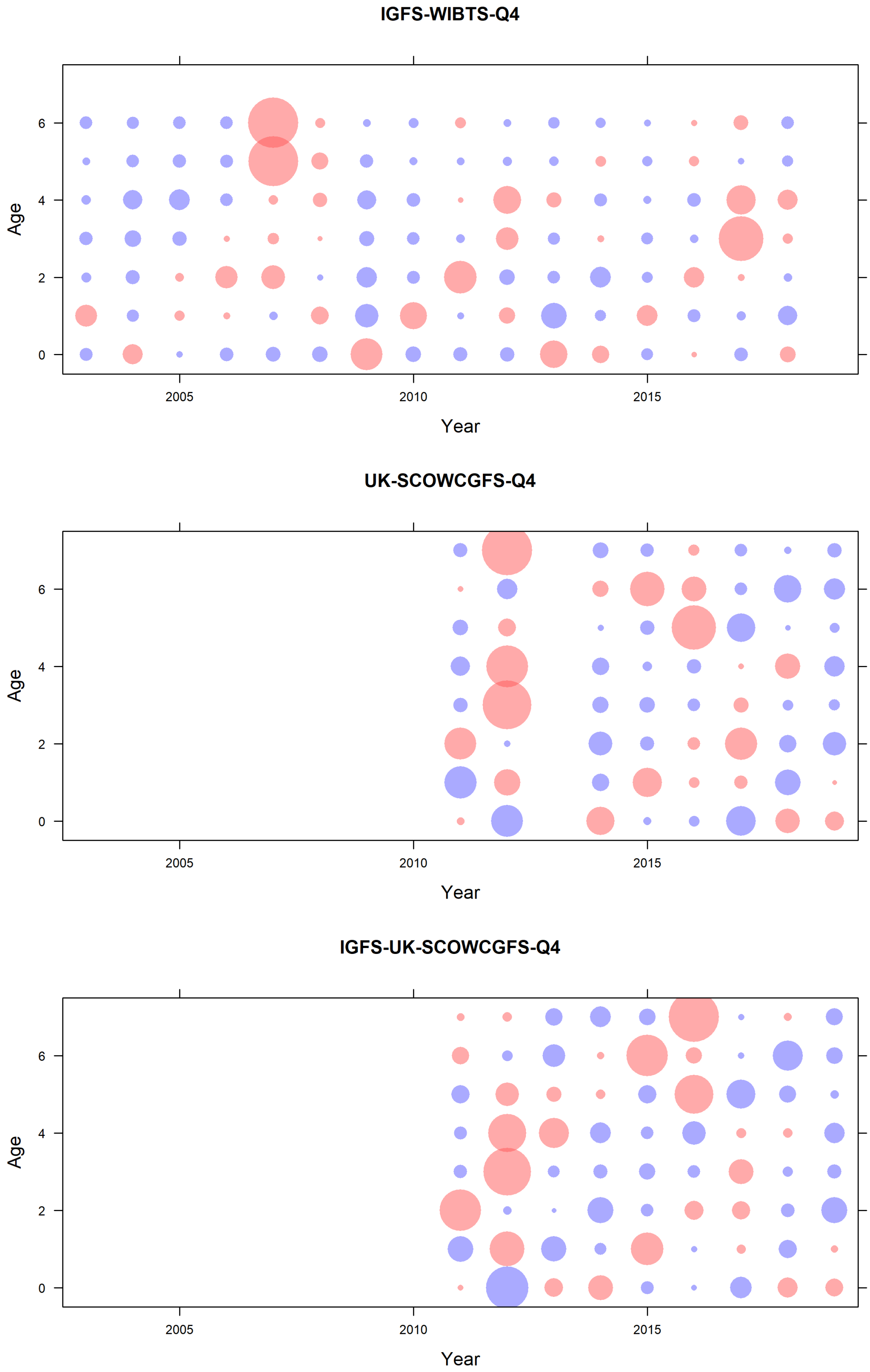


Figure 5.6. Standardized proportions at age per year ("spay") for three Q4 survey series. The positive values are shown in red, the negative values are shown in blue.

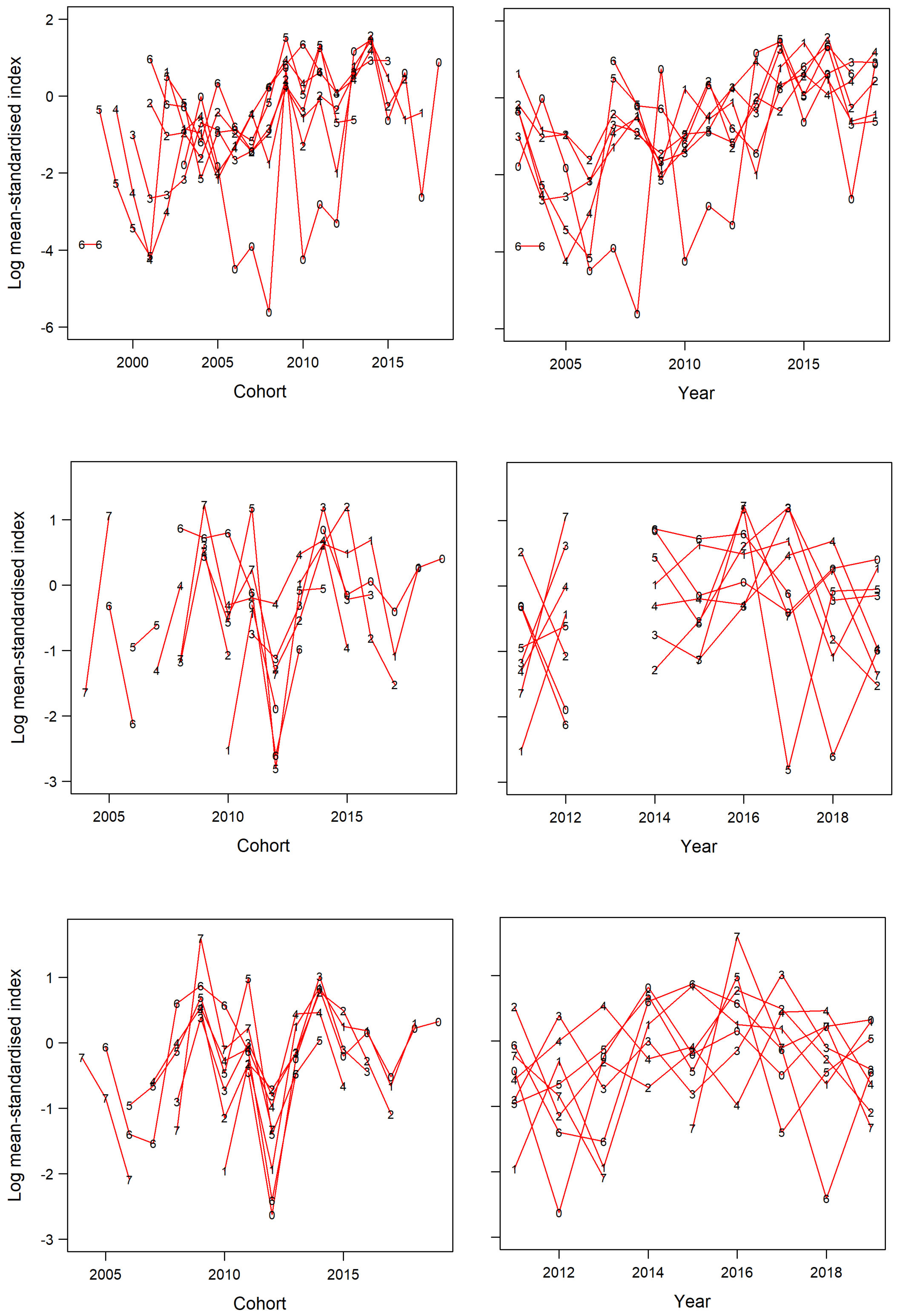


Figure 5.7. Log mean standardized survey index for each age by cohort (left panels) and year (right panels) in the three Q4 survey series.

The log catch curves for the three tuning series are shown in Figure 5.8. The curves for the Scottish survey and combined survey are relatively linear and not very noisy. They also show a fairly steep and consistent drop in abundance. With the Irish survey, the patterns are less clear.

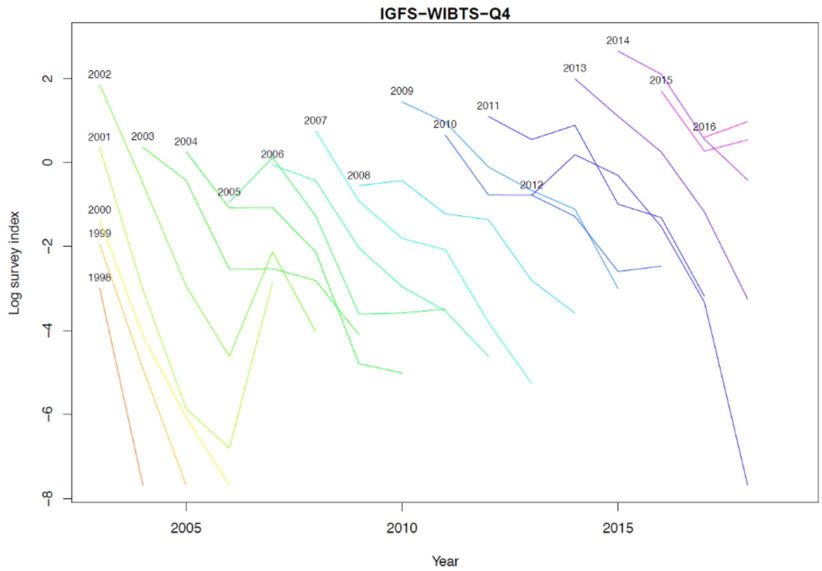
Figure 5.9 shows the survey scatterplots. For the three tuning series, there is a general consistency in the estimates of year-class strength across age groups, but the points are more scattered for the older age groups. The second and third time-series are less consistent, but they are shorter than the first one. For the third tuning series, the index values show relatively high consistency.

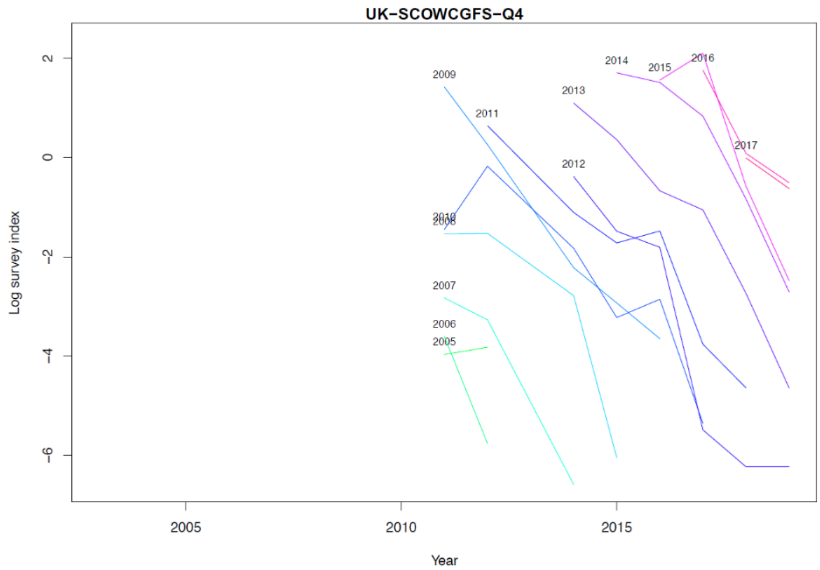
5.4 Discussion

This analysis demonstrates some differences between the two Q4 surveys in terms of catchability. The catch rates were found higher in the Irish survey, particularly for the young age groups. The reason for this discrepancy remains unclear, but it should be subject to further investigation or assessment. Nonetheless, one of the main goals of the analysis was to quantify the differences for age groups based on the available information on explanatory variables. This goal was achieved, which subsequently allowed to calibrate CPUE in the two surveys to produce a combined index.

While the observed differences were previously considered to result from the difference in time of the year of the surveys (IBPWSRound; ICES, 2015), this effect could not be ascertained with exploratory analyses that preceded the main analysis. As a result, the effect was omitted from the current model. The effect of haul time was rather weak or, in most cases, absent. It is likely that more data would be needed to detect an effect with either factor. The haul location and depth are clearly the most influential explanatory variables and any index calibration should take them into account.

There are several advantages of using a combined index for assessments of fish stocks. In this particular case, the combined index for the Scottish and Irish surveys provides a more complete representation of the population compared to the respective indices used on their own. This is supported by the diagnostics shown for the different survey series. The combined index simplifies, to some extent, the modelling procedure in the annual assessments of the stock (with two rather than three indices in the following years). It is possible to combine the index from the two surveys in situations where the Scottish survey can be completed only in the northern part of the survey area as it was the case in 2013.


IGFS-WIBTS-Q4
Log survey index
Year
2005
2010
2015


UK-SCOWCGFS-Q4
Log survey index
Year
2005
2010
2015

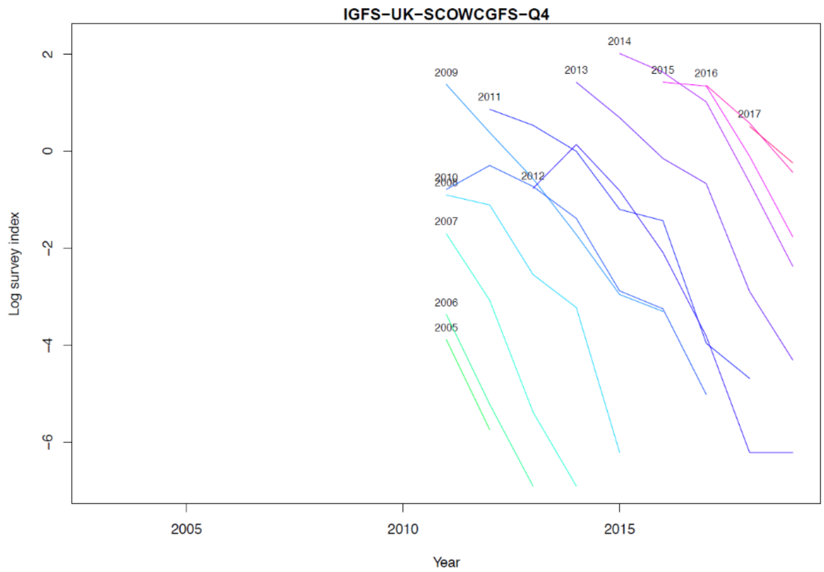

IGFS-UK-SCOWCGFS-Q4
Log survey index
Year
2005
2010
2015

Figure 5.8. Log abundance indices, by year with a line for each cohort, for each of the three Q4 survey series. The spawning date of each cohort is indicated at the start of each line.

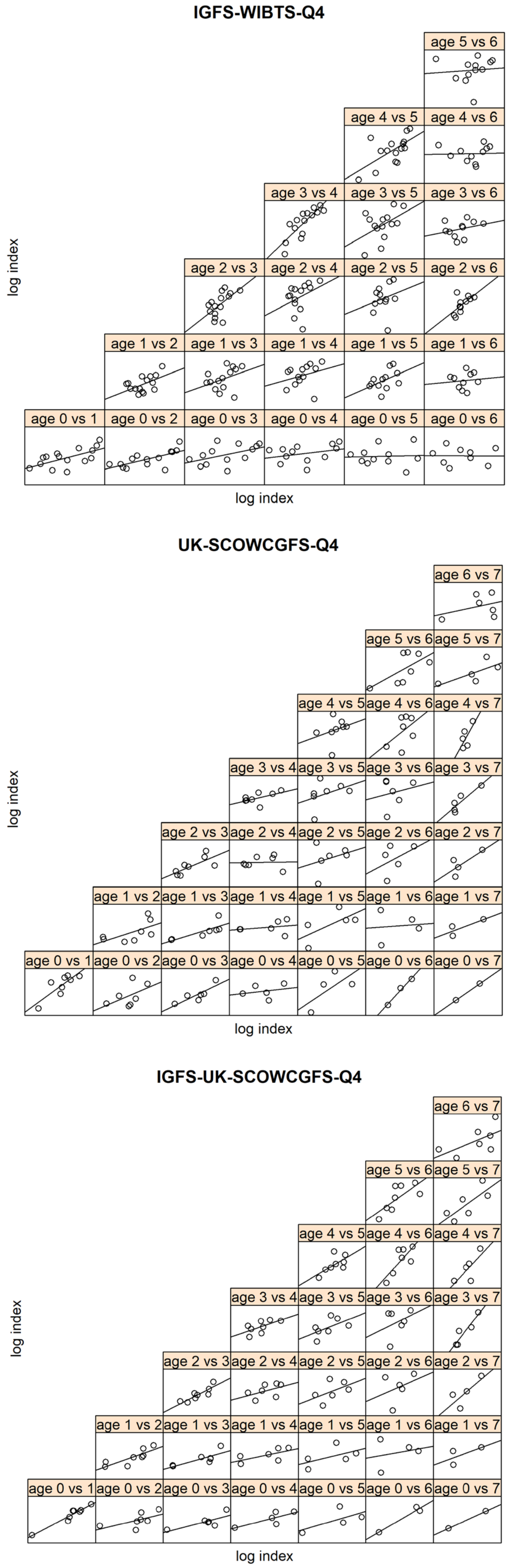


Figure 5.9. Within-survey correlations comparing index values at different ages for the same year classes for the three Q4 series. The straight line is a linear regression.

5.5 WGISDAA conclusion

WGISDAA welcomed the opportunity to contribute to the analysis through comments and information requests but recognizes that it does not have the role to take decisions in the advisory processes so takes no formal opinion on the appropriate form of a future assessment. We are pleased to note that the analyses conducted in conjunction with WGCSE will be considered in the upcoming benchmark.

The evidence presented here does strongly suggest that there are spatial differences (north v south) linked to changes in the distribution of the stock. The resulting current indices therefore contain conflicting signals which are reducing the effectiveness of the assessment. The design-based solution to creating a single index based on a fixed catchability conversion demonstrates a significant improvement in the internal consistency of the index. The analysis is not yet exhaustive and in the long run it would be useful to explore options for a model-based index able to monitor / improve the estimate of relative catchability based on areas of existing and future survey overlap but in the short term such changes are likely to be less relevant. Right now It would seem sensible to prioritize the full diagnosis of the behaviour of the combined index in the assessment.

6 Evaluating the potential for model-based index methods to deal with interrupted or conflicting survey signals; towards a benchmark for Celtic Sea gadoids

6.1 Report back from the Workshop on Evaluating Survey Information on Celtic Sea Gadoids

Dave Stokes (MI, Ireland) presented a summary of the WGISDAA instigated Workshop on Evaluating Survey Information on Celtic Sea Gadoids (WKESIG, Galway, Ireland 4-5 February 2019). WKESIG was established to provide expertise in the review and preparation of survey indices for the Celtic Sea Benchmark meeting WKCELTIC 2019/2020.

Cod 7e_k, Haddock 7b_k and Whiting 7b_k are three of the largest stocks assessed by the Working Group for the Celtic Seas Ecoregion (WGCSE) and form part of a significant mixed demersal fishery in the Celtic Sea. These fisheries rely heavily on recruitment and therefore survey indices, which are all combined survey indices in this case between two IBTS surveys, namely Ireland (IE-IGFS) and France (EVHOE). The WGCSE assessment group highlighted a number of specific survey data issues and it was felt that WGISDAA would be well placed to contribute to this benchmark review. Key issues were how survey data are standardized and combined varies across the indices. In addition, a number of unavoidable data gaps in survey coverage have occurred in recent years. Finally, estimates of uncertainty are not routinely calculated as part of these index calculations. Particularly where unavoidable effort reduction becomes an issue it was felt desirable to agree on appropriate methods to do this going forward.

Having explored variability of survey effort during the workshop it was agreed data gaps would be addressed and this was achieved in time to allow production of indices by swept-area for the benchmark. While this was not considered a primary cause of noise in the indices, it is a useful added quality check that gear parameters, vessel speed and trawl positions for example were 'normal' and recorded correctly. It was agreed that recent modelling approaches, such as those being considered by the workshop, provide good opportunities to evaluate and account for survey observation error and should be explored further.

Migration patterns were evident from tagging data presented at the workshop (Figure 6.1) and should be explored as supplementary information in the benchmark. It was not clear how best that might be integrated into assessments yet in a quantitative way, but should be discussed at WKCELTIC.

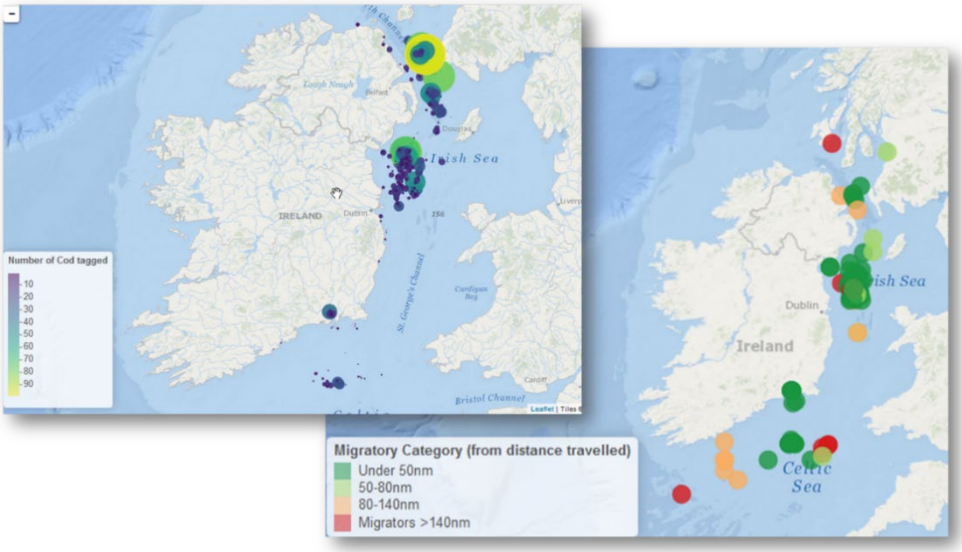


Figure 6.1. Release (left panel) and recapture sites (right panel) for Irish Sea cod tagging project.

Cefas had been working towards a framework for evaluating the impact of survey design changes on the advisory processes. So far efforts have focused on three aspects, first implementing different ways of subsampling, second implementing different methods of index calculation, both model and design based indices and third automating the assessment process in an effective way to allow the evaluation of large numbers of simulations. Preliminary analysis comparing two methods of survey effort reduction was presented to evaluate the impact of reducing tow duration from 30 minutes to 15 minutes (Figure 6.2) as well as the alternative of reducing tow number by 50% (Fig 6.3).

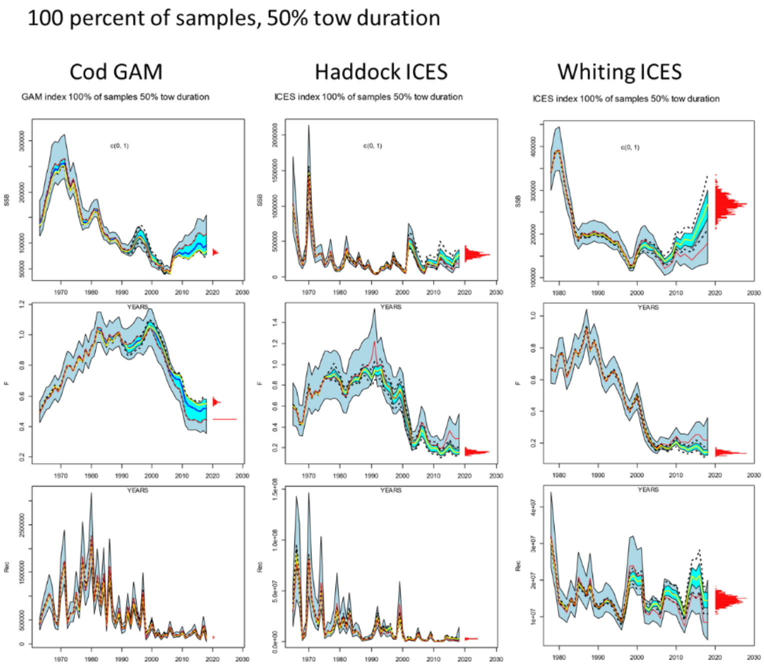


Figure 6.2. Stock assessment output for three species (NS cod, haddock, whiting left to right, SSB, F, recruitment top to bottom) compared to the current stock assessment as performed by WGNSSK 2018 using the approved index method for each assessment for all of stations at 50% tow duration.

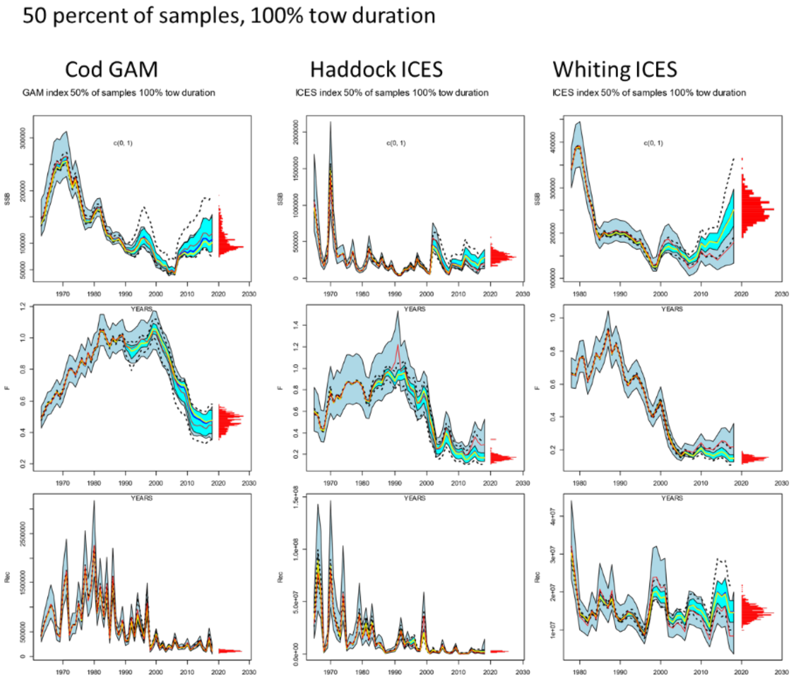


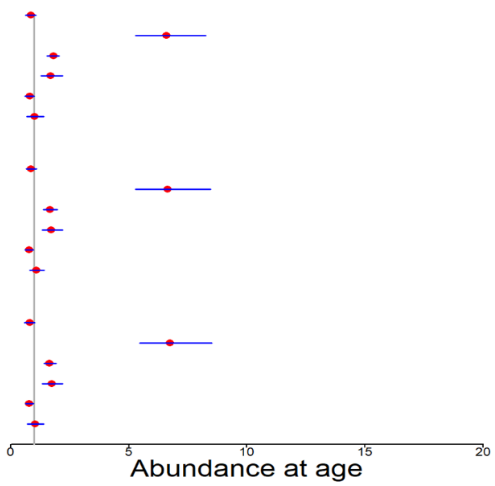
Fig 6.3. Stock assessment output for three species (NS cod, haddock, whiting left to right, SSB, F, recruitment top to bottom) compared to the current stock assessment as performed by WGNSSK 2018 using the approved index method for each assessment for 50% of stations at full tow duration.

Neither method of effort reduction had a major effect on the variability of assessment outputs, but unexpectedly some biases developed compared to the current assessment based in both cases. However, the impact of using different methods for survey index calculation had a bigger effect on the assessment outcomes than did the reduction in sampling effort. From an efficiency perspective a 50% reduction in the number of stations appeared preferable to a 50% reduction in tow duration.

Modelling of ALK data highlighted the precision gains achievable (Table 6.1) when spatial variability is accounted for. The work is a progression of the methods currently used for ALK 'fill-ins' for current indices Gerritsen *et al.,* 2006), but still developing so of interest to the group to follow up.

Table 6.1. Estimates of mean CPUE at age (with 95% confidence intervals CI_{95}) for cod in the North Sea using 3 alternative ALK derivation methods. The reduced confidence intervals of both haul based and model based can be largely attributed to their ability to account for spatial variation in the data.

Age	Species	SE	RSE%	Width CI-95	Estimate (CI_{95})
Area based					
1	cod	0.124	14.49	0.428	0.86 (0.65-1.07)
2	cod	0.872	13.21	2.959	6.60 (5.29-8.25)
3	cod	0.174	9.62	0.491	1.81 (1.55-2.04)
4	cod	0.281	16.50	0.890	1.70 (1.30-2.19)
5	cod	0.124	15.39	0.371	0.81 (0.62-0.99)
6+	cod	0.244	23.81	0.695	1.02 (0.69-1.39)
Haul based					
1	cod	0.126	14.57	0.437	0.87 (0.65-1.09)
2	cod	0.936	14.07	3.173	6.65 (5.28-8.46)
3	cod	0.162	9.68	0.587	1.67 (1.39-1.97)
4	cod	0.283	16.42	0.855	1.72 (1.34-2.20)
5	cod	0.117	14.88	0.337	0.79 (0.61-0.95)
6+	cod	0.267	24.30	0.610	1.10 (0.81-1.42)
Model based					
1	cod	0.124	15.21	0.433	0.82 (0.59-1.02)
2	cod	0.918	13.60	3.042	6.75 (5.46-8.50)
3	cod	0.178	10.77	0.501	1.65 (1.42-1.92)
4	cod	0.292	16.83	0.853	1.74 (1.35-2.20)
5	cod	0.115	14.59	0.319	0.79 (0.63-0.95)
6+	cod	0.252	24.01	0.675	1.05 (0.71-1.39)



Dealing with overall data gaps was addressed by presentation of a case study on Celtic Sea whiting index calculation using a spatio-temporal model, VAST. It was shown that the model had good stability and predictive accuracy even when large amounts of input data were removed from a given year, to simulate vessel breakdown for example. However, the ability to model the missing values was heavily influenced by how 'average' a survey year was. In unusually high or low abundance years the capacity to 'model your way out' of data gaps is reduced significantly (Figure 6.4).

Overall, the modelling approach offered potential to address spatial variability of the data, estimate uncertainty in terms of observation and process error separately and also reduce the impact of unavoidable data gaps. The result at the benchmark was for a significant increase in internal consistency in the indices for whiting in particular, but for the other species in general also (Figure 6.5.).

6.2 WGISDAA conclusion

WGISDAA welcomed the opportunity to contribute to the analysis through comments and information requests but recognizes that it does not have the role to take decisions in the advisory processes so takes no formal opinion on the appropriate form of a future assessment. We are pleased to note that the analyses conducted in conjunction with WGCSE will be considered in the upcoming benchmark.

The group was pleased to cooperate with WGCSE and although much of the work was done after the actual meeting and in other groups felt the workshop provided a useful opportunity to discuss the benefits and risks of design- and model-based indices on the basis of some new case studies. Because, data collection and index calculation methodologies are not independent recommendation can only be made on a case-by-case basis but it is possible to identify generality in conditions which could lead to more effective decision-making in an integrated management system as suggested by WKUSER (see above). It is suggested that this work continue in conjunction or as part of the proposed future work of WKUSER.

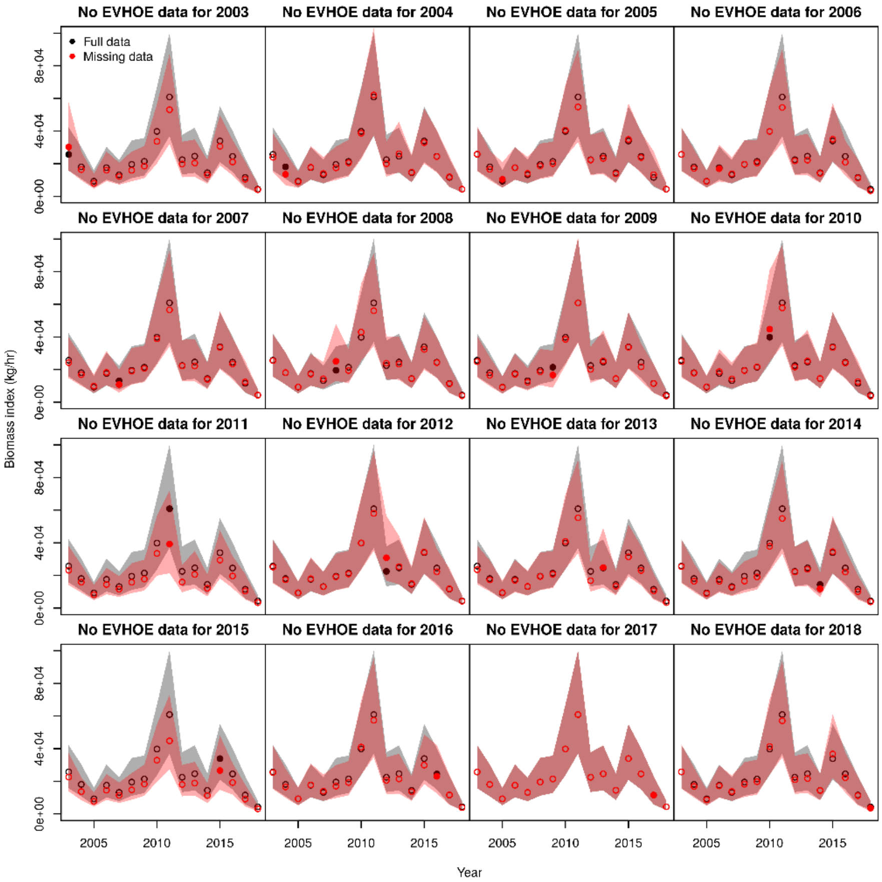


Fig 6.4. Whiting in area 7b-ce-k age 2: The impact of missing data were investigated by sequentially removing the EVHOE survey data year on year to mimic survey failure for that year and impact on the time-series. The index based full dataset is given as black circles with grey confidence intervals. The recalculated index with a missing year is given as red empty circles with the red solid circle indicating the annual data point where EVHOE survey has been removed. Confidnce bounds are shaded light red for this recalculated index. The VAST model produced very similar biomass estimates for the full dataset and the partial data, suggesting that the survey coverage in 2017 was sufficient for the model to accurately estimate the index. It is noteworthy that when data are removed from a year with an unusually high index that the re-estimated index shows the greatest difference from the original.).

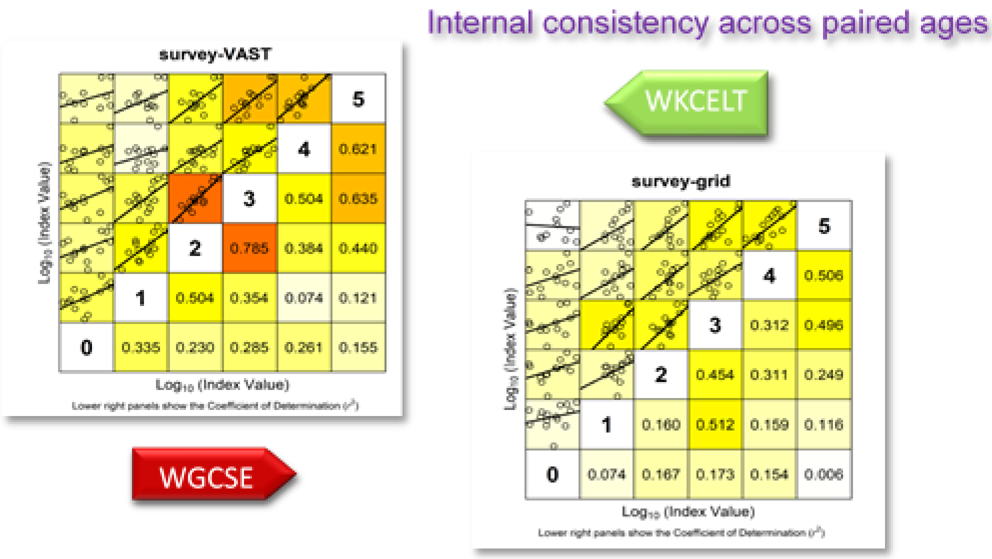


Fig 6.5. Internal consistency of Log index for Whiting in area 7b-ce-k age 0-5: plots of Agevs.Age+1 for the time-series. Top left panel in VAST based survey index, lower right panel is original grid based index from previous WGCSE 2019 assessment.

7 References

Augustin, N. H., Trenkel, V. M., Wood, S. N. and Lorance, P., 2013. Space-time modelling of blue ling for fisheries stock management. Environmetrics. 2. 109-119.

Bernal, M., Somarakis, S., Witthames, P.R., Damme, C.J.G. van, Uriarte, A., Lo, N.C.H. and Dickey-Collas, M. 2012. Egg production methods in marine fisheries: An introduction. Fisheries Research: 117-118, 1-5.

Borchers D.L., Buckland S.T., Priede I.G., and Ahmadi S.. Improving the precision of the daily egg production method using generalized additive Models Can. J. Fish. Aquat. Sci. 54: 2727–2742 (1997)

Efron, 1979. Bootstrap Methods: Another look at the Jacknife. The annals of Statistics. Vol 7, No 1, 1-26.

Gerritsen, H.D., McGrath D., and Lordan C.. 2006. 'A Simple Method for Comparing Age–Length Keys Reveals Significant Regional Differences within a Single Stock of Haddock (Melanogrammus Aeglefinus)'. ICES Journal of Marine Science: Journal Du Conseil 63(6):1096–1100. doi: 10.1016/j.icesjms.2006.04.008.

Hastie, T., Tibshirani, R. 1986. Generalized Additive Models. Statistical Science. 3. 297-310.

Haddon, M. 2011 Modelling and Quantitative Methods in Fisheries; Second Edition. Chapman and Hall/CRC, 449 pp.

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

ICES. 1993. Report of the Mackerel/Horse Mackerel Egg Production Workshop. ICES CM 1993: H4.

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. 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 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

ICES. 2018. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK), 24 April - 3 May 2018, Oostende, Belgium. ICES CM 2018/ACOM:22. pp.

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. 29 pp.

ICES. 2018. Report of the Working group on improving use of Survey Data for As-sessment and Advice (WGISDAA). 3–5 July 2018. Copenhagen, Denmark. ICES CM 2018/EOSG: 15. 30 pp.

ICES. 2019. Working Group on Mackerel and Horse Mackerel Egg Surveys (WGMEGS). ICES Scientific Reports. 1:66. 233 pp. http://doi.org/10.17895/ices.pub.5605

ICES. 2019. Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). ICES Scientific Reports. 1:7. 1271 pp. http://doi.org/10.17895/ices.pub.5402.

ICES. 2019. WORKSHOP ON NORTH SEA STOCKS MANAGEMENT STRATEGY EVALUATION (WKNSMSE). ICES Scientific Reports. 1:12. 378 pp. http://doi.org/10.17895/ices.pub.5090.

ICES. 2020. Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). ICES Scientific Reports. 2:61. 1140 pp. http://doi.org/10.17895/ices.pub.6092.

ICES. 2020. Workshop on Catch Forecast from Biased Assessments (WKFORBIAS; outputs from 2019 meeting). ICES Scientific Reports. 2:28. 38 pp. http://doi.org/10.17895/ices.pub.5997.

ICES. 2020. ICES Workshop on evaluating survey information Celtic Sea gadoids (WKESIG).

ICES Scientific Reports. 2:107. 26 pp. http://doi.org/10.17895/ices.pub.7574.

ICES. 2020. Workshop on Stock Identification of North Sea Cod (WKNSCodID). ICES Scientific Reports. 2:89. 82 pp. http://doi.org/10.17895/ices.pub.7499.

ICES. 2020. Report of the Working Group on Mackerel and Horse Mackerel Egg Surveys (WGMEGS), 28–29 April 2020.

ICES. 2020a. Working Group for the Celtic Seas Ecoregion (WGCSE). ICES Scientific Reports. 2:40. 924 pp. http://doi.org/10.17895/ices.pub.5978

ICES. 2020b. Benchmark Workshop for Demersal Species (WKDEM). ICES Scientific Reports. 2:31. 136 pp. http://doi.org/10.17895/ices.pub.5548

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

Lockwood, S.J., Nichols, J.H., Dawson, W.A. 1981. The estimation of a mackerel (*Scomber scombrus* L.) spawning stock size by plankton survey. Journal of Plankton Research: 3, 217-233.

Peel, D., Bravington, M. V., Kelly, N., Wood, S. N. and Knuckey, I., 2013. A Model-Based Approach to Designing a Fishery-Independent Survey. Journal of Agricultural, Biological, and Environmental Statistics. 1. 1-21.

Shono, H., 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research. 1. 154-162.

Smith, S.J., Hunt, J.J., Rivard, D. (eds) 1993 - Risk Evaluation and Biological Reference Points for Fisheries Management. Canadian Special Publication of Fisheries and Aquatic Sciences 120.

Swan A. R. H. and Sandilands, M., 1995. Introduction to Geological Data Analysis. Blackwell. 446 pp.

Wood, S., 2017. Generalized Additive Models: An Introduction with R (2nd ed).. Chapman & Hall/CRC Texts in Statistical Science. 476.

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.

Annex 1: List of participants

Sven Kupschus (chair)		**UK**	**sven@kupschus.net**
Stan Kotwicki	AFSC, Seattle	USA	stan.kotwicki@noaa.gov
Kotaro Ono	IMR, Bergen	Norway	kotaro.ono@hi.no
Casper Berg	DTU Aqua	Denmark	cbe@aqua.dtu.dk
Gersom Costas	IEO	Spain	gersom.costas@ieo.es
Matthias Kloppmann	TI-SF, Bremerhaven	Germany	matthias.kloppmann@thuenen.de
Andrzj Jaworski	Marin Scottland	UK	a.jaworski@marlab.ac.uk
Kai Wieland	DTU Aqua	Denmark	kw@aqua.dtu.dk
Nicola Walker	Cefas	UK	nicola.walker@cefas.co.uk
Richard Nash	Cefas	UK	Richard.nash@cefas.co.uk
Esther Beukhof	Wageningen Marine Research	Netherlands	esther.beukhof@wur.nl
Curry Cunningham	University of Alaska, Fairbanks	USA	ccunningham@alaskapacific.edu
David Stokes	Marine Institute	IR	david.stokes@marine.ie
Ismael Núñez-Riboni	TI-SF, Bremerhaven	Germany	ismael.nunez-riboni@thuenen.de
Ioannis Thasitis	DEPARTMENT OF FISHERIES AND MARINE RESEARCH	Cyprus	ithasitis@dfmr.moa.gov.cy
Patrik Börjesson	SLU Aqua, Lysekil	Sweden	patrik.borjesson@slu.se

Annex 2: Resolutions

WGISDAA – Working Group on Improving use of Survey Data for Assessment and Advice

2017/2/EOSG06 A **Working Group on Improving use of Survey Data for Assessment and Advice** (WGISDAA), chaired by Sven Kupschus, UK, will work on ToRs and generate deliverables as listed in the Table below.

	Meeting dates	Venue	Reporting details	Comments (change in Chair, etc.)
Year 2018	3-5 July	Copenhagen, Denmark	Interim report by 20 September to ACOM/SCICOM	
Year 2019	8-10 October	Copenhagen, Denmark	Interim report by 15 November to ACOM/SCICOM	
Year 2020	6-8 October	By Correspondence/Online meeting	Final report by 5 November to ACOM/SCICOM	

ToR descriptors

ToR	Description	Background	Science Plan Codes	Duration	Expected Deliverables
a)	To work together with assessment working groups to provide resolution to assessment issues prioritized by the assessment working groups	Specific resolutions to individual assessment issues with a report to feedback into the assessment, or where necessary into the benchmark process. In addition, cataloguing and classification of issues and review of methods used to resolve problems in order to provide "self-help" options to resolve similar issues in other assessments.	5.1		
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.	Specific resolutions to individual survey issues with a report to feedback into the survey working group. In addition cataloguing and classification of issues and review of the methods used to resolve them in order to provide "self-help"	3.1, 3.2		

		options for survey working groups.			
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	Survey data issues, as in ToR a, are often critical in the benchmarking process. WGISDAA can advise best if involved in this process from the start, can collaborate with the operators and present conclusions at the benchmark	5.1	As required	Reports and presentations to the appropriate Benchmark workshop.
d	Review the output from the topic specific workshops initiated by WGISDAA and document more general principles learned from the specifc cases dealt with in TOR a and b	WGISDAA has had difficulty in attaining wider participation in its work	-		

Summary of the Work Plan

Year 1	Continue and update process elicitating 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 elicitating 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

Supporting information

Priority	This group will feed the results of its work directly into the assessment and hence advisory process. As such it should be considered central and of high priority
Resource requirements	The key additional resource requirement is the group needs particpation of the key players in the relevant assessment, survey or benchmark group. This would be in addition to work required for the normal operations of htese groups. Essentially, this would involve key personnel attending the relevant WGISDAA meeting, and where rquired, personnel from WGISDAA attending the relevant requesting EG
Participants	Dependant on information requests, but normally less than 10 core members
Secretariat facilities	None.
Financial	No financial implications.
Linkages to ACOM and groups under ACOM	ACOM, Benchmark process and assessment EGs as well as Survey EGs will be the key clients for the work of WGISDAA
Linkages to other committees or groups	WGISDAA will have strong links to to survey working groups under SSGIOMP, and in particular to the work of WGISUR. Given surveys as an important source of wider ecosystem data there will also be important links to groups under SSGIEA
Linkages to other organizations	None specific