ICES WGISDAA REPORT 2017

ICES CM 2017/SSGIEOM: 16

REF ACOM AND SCICOM

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

11-13 July 2017

Copenhagen, Denmark

International Council for the Exploration of the Sea

Conseil International pour l'Exploration de la Mer

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

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

Recommended format for purposes of citation:

ICES. 2018. Report of the Working group on Improving use of Survey Data for Assessment and Advice (WGISDAA) 11-13 July 2017. Copenhagen, Denmark. 38 pp.
https://doi.org/10.17895/ices.pub.8680

Contents

Executive summary

Over its three-year term the WG has been providing *ad hoc* advise to survey and advice working groups as requested. Predominantly these relate to survey WGs in this period as changes in the distribution of species and budget constraints have meant that the current survey protocols had developed deficiencies in relation to the assessment needs or survey effort required to deliver consistent information exceeded the available budgets, while assessments concerns were dominated by issues other than survey data.

Beyond addressing the specific questions, the WG also tried to characterise the methods used evaluate the survey information content relevant for assessments to provide some generalisation, as to how this could or should be used in other cases and where the weaknesses or assumptions behind underlying these methods were.

Model based solutions generally based on generalized regression approaches proved useful in dealing with catastrophic dataloss (missing year area combination) and other enforced changes such as small changes in survey execution or behaviour of fish (changes in the vertical distribution from year to year affecting catchability).

Design based approaches are favoured to deal with planning for survey design changes and how existing index series can be maintained despite some changes to survey protocols. One general concern with these methods are that it is difficult to use these to make decisions on future surveys not because they are not appropriate, but because the different survey objectives are not quantitatively prioritised.

1 Administrative details

Working Group name: working Group on Improving use of Survey Data for Assessment and Advice (WGISDAA)

Year of Appointment within the current three-year cycle 2015

Reporting year concluding the current three-year cycle 2017

Chair(s)

Sven Kupschus, UK

Meeting venue(s) and dates

20-22 January 2015, Copenhagen, Denmark (6 participants)

12-14 July 2016, Hamburg Germany, (5 participants)

11-13 July, Copenhagen, Denmark, (7 participants)

2 Terms of Reference a) – z)

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

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

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

3 Summary of work plan

Year 1: Initiate process eliciting advice requests from other elements of the ICES system; assessment, survey and benchmarking groups. Identify priorities within requests, and set up meeting and personnel accordingly.

Year 2: Continue and update process eliciting advice requests from other elements of the ICES system; assessment, survey and benchmarking groups. Identify priorities within requests, and set up meeting and personnel accordingly.

Year 3: As in year 2, plus appraisal of the success of the process, and make proposals for changes and any continuation.

4 Summary of Achievements of the WG during 3-year term

The WG operates more as a forum to assist survey and assessment scientists to confer and discuss specific issues. However, the working group has over the three years been developing a focus on methodologies for investigating and evaluating surveys where they contribute to stock or ecological assessments. Most of the evaluation work for surveys has taken a multi-year iterative approach with bespoke solutions being developed. But it seems that our surveys are not regularly evaluated in terms of their performance so there is a real need for survey scientists to have an evaluation toolbox. The WG has tried to generalise the solutions developed and to clarify their benefits and weaknesses so over time this work should provide a useful compendium to assist future survey managers. Survey WGs that have been directly supported by WGISDAA in their science are WGMEGS, IBTSWG, WGBEAM, WGALES.

5 Final report on ToRs, workplan and Science Implementation Plan

The WG addressed its work bus subject rather than by TOR with different elements of the individual studies addressing a number of TORs. To maintain the coherence in the studies the work here is presented as separating the studies by conclusions in terms of TORs would risk.

5.1 Resolving conflicting acoustic and trawl survey indices by understanding the impact of inter-annually variable fish behaviour

Indices of abundance from multiple surveys are often used in stock assessments of species for which one survey method cannot cover entire extent of stock distribution. Some demersal species spend periods high in the water column and are inaccessible to the bottom trawls (BT; e.g. Atlantic cod, haddock). On the other hand pelagic species at times are present in the near bottom zone (e.g. spawning herring). In the case of such species abundance indices are often estimated using acoustic or bottom trawl (BT) surveys, both of which sample a fraction of the water column. Acoustic instruments are effective at sampling the water column, but they have a near-bottom acoustic dead zone (ADZ), where fish near the seafloor cannot be detected. Bottom trawl surveys cannot account for fish that are located above the effective fishing height (EFH) of the trawl. Currently indices of abundance from these two surveys are commonly used independently in stock assessments. The alternative approach is to use just one of these indices in stock assessment models. Each of these approaches leads to introduction of additional uncertainty into stock assessments. This additional uncertainty arises from year to year variation in catchability of both indices of abundance. Variation in catchability can result from variable efficiency of the BT (e.g. Kotwicki *et al.*, 2014), or from variable availability of fish to the survey gear (Kotwicki *et al.*, in press).

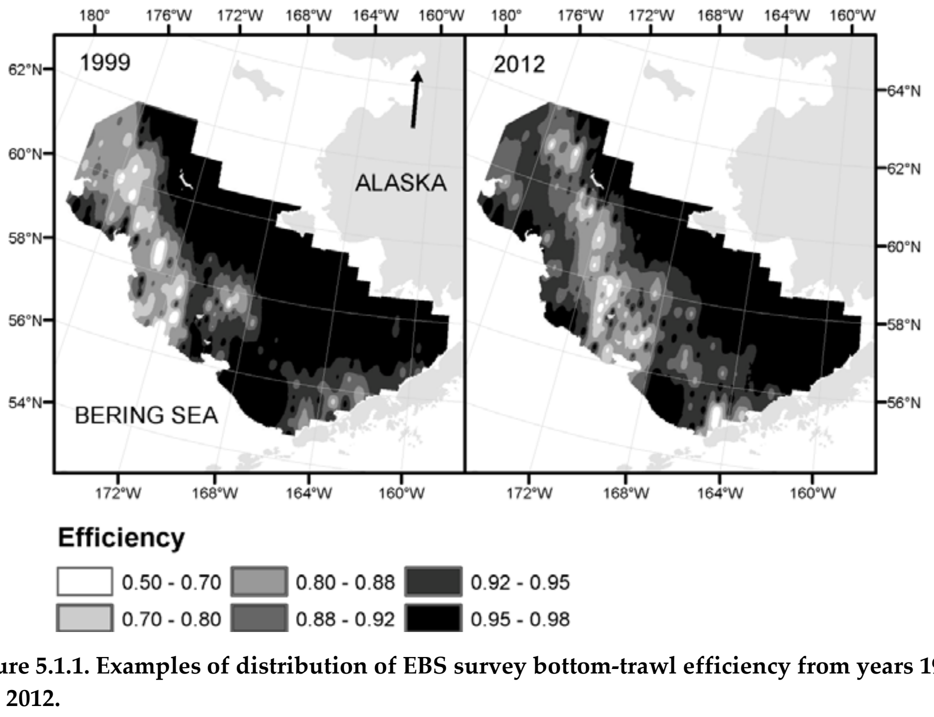


Figure 5.1.1. Examples of distribution of EBS survey bottom-trawl efficiency from years 1999 and 2012.

Variable catchability in the index of abundance is of concern because it is not usually represented in the main structure of the stock assessment model and can lead to bias in the estimates of depletion levels and contribute substantially to the uncertainty in stock assessment results and estimates of management quantities (e.g. Maunder and Piner, 2014). Kotwicki *et al.* (2014) showed that density dependence in BT efficiency can lead to variable catchability of walleye pollock in BT surveys (Figures 3 and 4). Additionally, it has been shown that availability of walleye Pollock to the BT and acoustic surveys changes in response to environmental factors (Kotwicki *et al.*, in press).

To address problems associated with variable catchability of the BT and acoustic surveys work has been undertaken at Alaska Fishery Science Centre on methods to combine BT and acoustic data to improve abundance estimates. Semipelagic walleye pollock (*Gadus chalcogrammus*) was chosen as a case study for this work because they are a dominant species with important commercial and ecological roles in the North Pacific.

A model combining a subset of acoustic and BT data was developed to estimate ADZ correction and BT efficiency parameters (Kotwicki *et al.*, 2013). Fitting this model to the data provided estimates of the bias ratio between BT and acoustic data, the effective fishing height (EFH) of the BT, and the density dependent efficiency of the BT. Estimates of experimentally-derived ADZ correction and BT efficiency parameters were then used to develop a model predicting BT efficiency as a function of BT catch rate. It was found that BT efficiency decreased with increasing bottom trawl catches resulting in hyperstability of the abundance index derived from BT survey. Density-dependent BT efficiency resulted in spatially and temporarily variable bias in survey CPUE (Figure 3) and biased population age structure derived from survey data. Logistic regression models were developed to predict the availability of pollock to both acoustic and BT gears using environmental predictors and fish length (Kotwicki *et al.,* in press). Findings indicated that on average, availability of pollock in the EBS to the BT was larger than to the acoustics. Availability to both gears depended mostly on bottom depth, light conditions, and fish length, and to a lesser extent on sediment size. Availability to the acoustic gear also depended on surface temperature. Presently a new method is being developed for combining pollock abundance estimates from BT and acoustic data using estimates of bias ratio and overlap between BT and acoustic data. Preliminary results indicate that combined estimates provide relatively precise (CV~. 0.15 – 0.20) index of abundance corrected for variable catchabilty of the BT and acoustic gear. This methodology differs from previous attempts to combine BT and acoustic data (i.e. CATEFA project; Bouleau *et al.,* 2004) by incorporating specific processes associated with BT and acoustic sampling (i.e. bottom trawl efficiency parameters and existence of ADZ) into modelling process.

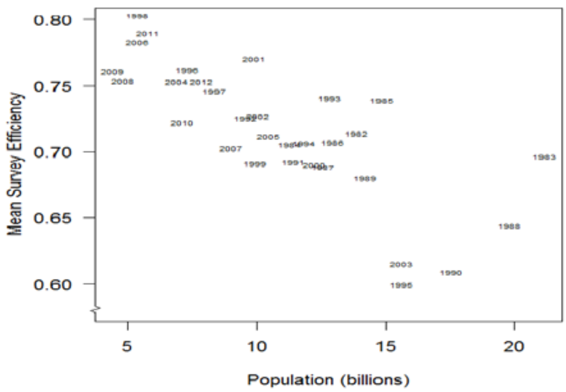


Figure 5.1.2. Negative relationship of mean BT survey efficiency vs. total abundance indicates hyperstability of the BT survey index of abundance.

Although this work specifically addresses the assessment of walleye pollock in the EBS it has general applicability in the assessment of other species that can be enumerated using both BT and acoustic gear. Methods being developed can be applied to other species to obtain estimates of ADZ correction or BT efficiency parameters. Similarly, methods for estimating density dependence of the BT, availability to the BT and acoustic gears, and combining BT and acoustic abundance estimates can be applied to other species.

5.2 Correcting density-dependent effects in abundance estimates from bottom-trawl surveys

Indices of abundance are important for estimating population trends in stock assessment and ideally should be based on fishery-independent surveys to avoid problems associated with the hyperstability of the commercial catch-per-unit-effort (cue) data. However, recent studies indicate that the efficiency of the survey bottom trawl for some species can be density-dependent, which could potentially affect reliability of survey-derived indices of abundance. Density-dependent effects of the BT have been identified as factors that may affect reliability of abundance estimates from BT surveys (Godø *et al.*, 1990; Godø and Wespestad, 1993; Godø, 1994; Aglen *et al.*, 1997; Kotwicki *et al.*, 2013). For example, survey trawl capture efficiency for Atlantic cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinis*) increases with fish density (Godø *et al.*, 1999). The opposite effect was observed for capelin *(Mallotus villosus*; O'Driscoll *et al.*, 2002), Atlantic croakers (*Micropogonias undulates*) and white perch (*Morone americana*; Hoffman *et al.*, 2009), and walleye pollock (*Gadus chalcogrammus*; Kotwicki *et al.*, 2013). Despite these findings, evaluations of the spatial and temporal variability in density-dependent BT survey efficiency are lacking, and methods which correct time-series of survey abundance indices are unavailable. Kotwicki *et al.* (2014) proposes to use a function $q_e \sim f(u)$ developed at the sample level, where q_e is bottom-trawl efficiency and u is a catch rate, obtained using experimentally-derived acoustic dead-zone correction and bottom-trawl efficiency parameters obtained from combining a subset of bottom-trawl catch data with synchronously collected acoustic data from walleye pollock (*Gadus chalcogramma*) in the eastern Bering Sea (EBS). We found that q_e decreased with increasing bottom-trawl catches resulting in hyperstability of the index of abundance derived from bottom-trawl survey. Density-dependent q_e resulted in spatially and temporarily variable bias in survey cpue (Figure 4) and biased population age structure derived from survey data (Figure 5). We used the relationship $q_e \sim f(u)$ to correct the EBS trawl survey index of abundance for density-dependence. We also obtained a variance–covariance matrix for a new index that accounted for sampling variability and the uncertainty associated with the q_e. We found that incorporating estimates of the new index of abundance changed outputs from the walleye pollock stock assessment model. Although changes were minor, we advocate incorporating estimates of density-dependent q_e into the walleye pollock stock assessment as a precautionary measure that should be undertaken to avoid negative consequences of the density-dependent q_e.

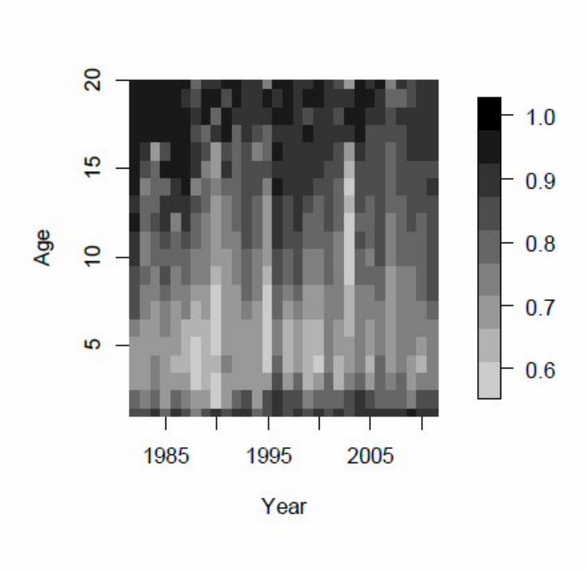


Figure 5.2.1. Bottom-trawl survey efficiency by year and age.

This method can be directly applicable for estimating efficiency of the BT for other species given availability of acoustic data collected during trawling. The advantage of this method is that in contrast to the survey independent experiments it can be easily incorporated into regular BT survey protocol and can cover the entire spatial extent covered by the survey. In Kotwicki *et al.* (2014) it was found that BT efficiency was density dependent. However for some pelagic species it is known that the abundance estimates from acoustic gear can be density dependent due to shadowing effect. In such case it should possible to account for shadowing effect by including density dependent effect in the acoustic data model. More over similar models are also possible for any situation where two independent enumeration methods are used simultaneously, given that appropriate model is proposed.

5.3 Geostatistical modelling to compensate for spatial and temporal variability of distribution of fish, survey gear differences, and sampling effort

Frequently multiple fisheries independent surveys cover or partially cover the extent of a stock of interest. Usually, the primary purpose of these surveys vary so that they differ in spatial and/or temporal coverage, in gear and/or vessel and potentially survey design. As a consequence, indices of abundance from different surveys can show differing trends and worse, they may conflict within a stock assessment model that does not appropriately account for these differences.

Effects of gear and vessel are adequately accommodated in most modern assessment models by estimating selectivity and catchability parameters independently for each survey. Developing spatially explicit models is now also possible, but efforts are often hampered by a lack of historic data at the appropriate spatial resolution. An alternative approach is to combine all survey data and construct an overall index of abundance that accounts for the differences in survey gear/boat (selectivity and catchability) and survey spatio-temporal coverage. This is often done through what is called a catch per unit of effort (cpue) standardization (Maunder and Punt, 2004). cpue standardization starts from the simple assumption that cpue is proportional (q) to abundance (B_t) with some observation error (e^{ε_t}).

$$cpue_t = qB_t e^{\varepsilon_t} \quad \text{with } e^{\varepsilon_t} \sim N(0, \sigma^2) \tag{1}$$

However, catchability is seldom constant over time and space and varies depending on the gear, location, time of the day, etc. (Wilberg *et al.*, 2010). Moreover, many environmental covariates can affect the underlying species distribution (biomass) (Whittaker *et al.*, 1973). Therefore, changes in catch rates are often modelled as a function of covariates that we think affect them.

$$cpue_t = q_t B_t e^{\varepsilon_t} \approx f(\text{covariates})\, e^{\varepsilon_t} \tag{2}$$

Spatio-temporal dependence has received much attention over the last decades as it is innate to many ecological processes as both exogenous (e.g. climate, habitat) and endogenous (e.g. dispersal, predation) drivers of species distributions vary through time and space (Legendre, 1993; Hoeting, 2009; Cressie and Wikle, 2011) which influence catch rates in the end. Traditionally, spatial dependence has been incorporated in cpue analysis in the form of spatial grids (Ono *et al.*, 2015). But the effects are rarely homogenous within grids (depending on the size of grid and species of interest). Therefore, spatial dependence is now preferably modelled as a continuous distribution over space by the use of spatial covariance structure, therefore allowing the analysis at the haul by haul level accounting for variation at appropriate spatial scales (Dormann, 2007). Temporal correlations, e.g. consistent distribution of a species over years, can be modelled by the use of autoregressive (AR), moving averages (MA) or autoregressive moving average (ARIMA) effects using techniques common to ecology, economics and engineering (Box *et al.*, 2016). Where such relationships exist, it is possible to compensate for missing data within the space-time array (DeYoung *et al.*, 2008; Perretti *et al.*, 2013).

There is however at least one remaining issue in cpue standardization and it is related to the problem of gear selectivity and catchability. If we ignore the age and/or size structure of the cpue data, there isn't an easy way to incorporate the differences in survey catchability and selectivity in the model as its effect is confounded with changes in underlying species size/age composition (moreover, gear selectivity can sometimes be quite different in shape). One approach to simplifying this problem is to divide the analysis by size groups (or age groups as in Berg and Kristensen, 2012 or Berg *et al.*, 2014). In doing so, differences in gear selectivity and catchability by group can be more easily modelled (by a constant through linear approximation) than over the whole size range, therefore increasing model accuracy i.e. $\overline{q'_{s,g}}$ in equation 3 is more accurate than $\overline{q''_{s}}$ in equation 4 as it is based on smaller size range.

$$CPUE_{s,g,t} = \left(\sum_{l \in g} q_{s,l} S_{s,l} B_{s,l}\right) e^{\varepsilon_{s,g,t}} \approx \overline{q'_{s,g}} B_{s,g} e^{\varepsilon_{s,g,t}} \tag{3}$$

$$CPUE_{s,t} = \left(\sum_{l=1}^{n} q_{s,l} S_{s,l} B_{s,l}\right) e^{\varepsilon_{s,t}} \approx \overline{q''_{s}} B_{s} e^{\varepsilon_{s,t}} \tag{4}$$

In the above equations, q is the catchability coefficient, S is the survey gear selectivity, B is the species biomass, and the subscript s, g, t, l stand for survey, size group, time, and fish size, respectively.

The final step in a cpue standardization is to produce an index of abundance. One approach is to calculate it based on the sum of standardized prediction over the assumed range of the species (Thorson *et al.*, 2015). An advantage of this approach is we can explore if there is any perceivable shift in the species distribution (e.g. center of gravity, distribution envelope) in addition to the variation in the index of abundance. Alternatively, the model could account for known shift in species distribution in calculating the final index of abundance. The resulting index of abundance by size (or age) group could then be readily incorporated into a stock assessment as a global index of abundance (or recruitment index if analysing small size groups).

To address the problem associated with combining multiple indices of abundance, works have been undertaken at the University of Washington on developing a cpue standardization model by size class that incorporates multiple survey data with different gears, space and time coverage to create an overall index of abundance. Pacific halibut (*Hippoglosomus stenolepis*) was chosen as a case study to illustrate the method.

Pacific halibut is one of the world's largest right-eye-flounder (up to 250 kg and 2.5 m) that is widely distributed along the Pacific Northeast from Northern California to the Alaska Bering Sea (Stewart *et al.*, 2016). The species has been exploited over a century and the average annual removals approximated 40 000 t (ranging between 20 000 to 60 000 t) (Stewart *et al.*, 2016) with a value exceeding 100 million $ in recent years (Fissel *et al.*, 2015). The species has been managed by the International Pacific Halibut Commission (IPHC) for nearly 100 years and stock assessment has been conducted annually since 1991. Many sources of data go into the assessment (both fishery-independent and dependent data) including some age and length composition data, survey indices of abundances, estimates of commercial and recreational catch and discard (Stewart and Monnahan, 2016). However, many challenges still remain in the assessment (Stewart *et al.*, 2016). Among others, recruitment variability is a significant source of uncertainty in estimating the stock status. One reason is that there is a lag of six to ten years in seeing any effect of recruitment in both the commercial fishery (as the legal capture size limit is 81.3 cm) and the setline survey abundance index and age composition data (which uses the same gear as the commercial fishery). The current abundance index from Alaska Bering Sea (BS), used in the assessment, provides some information about young halibut (thus recruitment) as it captures smaller fish than the setline survey but it is not representative at the coast-wide level and is based on an analysis that aggregated all size groups. By combining all bottom survey throughout Alaska (i.e. BS, Gulf of Alaska (GOA) and Aleutian Islands (AI)) and performing a cpue standardization focused on small size groups, one can develop an index of abundance specific to younger halibut (i.e. a recruitment index) that could be potentially useful in informing the coast-wide recruitment strength in the stock assessment model.

The cpue standardization model is based on a delta GLMM model by size group that account for spatio-temporal autocorrelation structure (spatial correlation is modelled via the Matérn covariance matrix while temporal correlation is modelled using a first order autoregressive process (AR1)) and the effect of other covariates such as year, survey region, depth and sea surface temperature. The final index of abundance by size group is determine by summing the standardized predictions (removing for the region specific difference in catchability) over a 1x1 km spatial grid overlaid across the survey region. Predictions were limited to the alpha shape (shape defining algorithm, a generalization of the convex-hull algorithm) of all survey points in order to prevent extrapolating much outside the area that have been surveyed in the past. In this sense, the derived index of abundance is representative of the young halibut abundance within the survey area. Nonetheless, sensitivity test to the choice of prediction area (one limited to the survey region and one extending beyond the survey region) did not show qualitative difference in the final index of abundance.

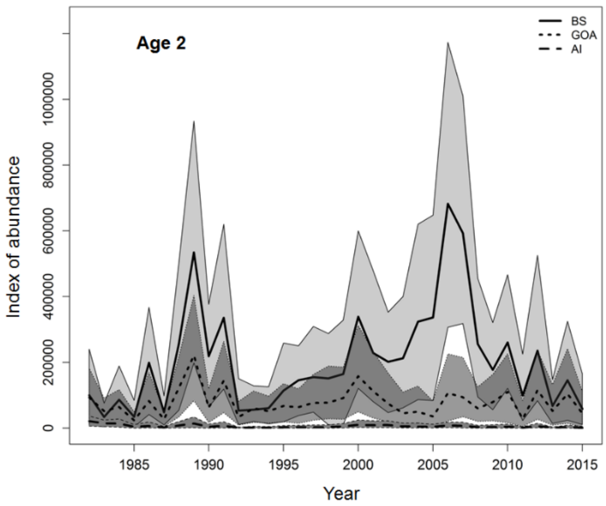


Figure 5.3.1: Index of abundance of young halibut (roughly age 2) across Alaska regions (i.e. BS, GOA, and AI) between 1982 and 2015 with the 75% credible interval.

The results of the analysis suggest that the young halibut (approximately age 2) is mostly found in the BS region and that the AI has a minimal contribution to the young halibut stock (Figure 6.1.1). Moreover, BS shows the largest fluctuation (in magnitude) in young halibut index over the years while GOA has a more stable index. In addition, there are two noticeable peaks in young halibut abundance across Alaska: 1990 (which is equivalent to a 1988 age-0 recruit) and 2007–2008 (equivalent to a 2005–2006 age-0 recruit). The 1988 recruitment peak is a well-known event that is also estimated in the stock assessment. However, the 2005–2006 peak is not apparent in the assessment. Looking at the regional contribution on the overall young halibut abundance in Alaska, the 1988 peak was observed both in the BS and in the GOA. However, the 2005–2006 peak was mostly evident in BS but not in GOA. When calculating the correlation between the scaled recruitment estimates from the latest assessment (Stewart and Martell, 2016) and the lagged scaled young halibut abundance index by region, correlation was as high as 0.6 when using the index within GOA but reduced to lower than 0.4 when using the index from the other regions or Alaska-wide. This suggests that the current assessment might be missing some source of recruitment that is coming from or generated within the BS. While this might be true, the current analysis is not able to inform about the contribution of the young halibut population across Alaska to the halibut spawning biomass that is the main target of the commercial fishery in Alaska.

Finally, WGISDAA stresses the need to explore and keep developing modelling approaches to analyse the combined haul-by-haul data from multiple surveys to detect spatio-temporal changes in distribution-at-size (or age). Such analysis could be useful in designing future survey (e.g. decision on extending or not survey area) but also lead to the development of stock-wide indices of abundance that are less biased and sensitive to the differences in survey spatio-temporal asynchrony and gear selectivity. Such development could help reducing subjective decisions in stock assessment whereby a lack of understanding of survey differences (such as spatio-temporal coverage and gear selectivity) frequently lead to the exclusion of one index vs. another (as in the case of NS cod) or down-weighting an index of abundance based on its divergence from other indices or assessment estimates.

5.4 Two-meter Midwater Ring trawl (MIK) sampling for 0-winter-ring herring larvae in the North Sea during the first Quarter of the year in conjunction with the International Bottom-trawl Survey (IBTS).

Update on the MIK herring larvae survey presented during the meeting.

The MIK survey provides an abundance index for large herring larvae (around and > 20 mm SL) that is used as a recruitment index for North Sea herring. It takes place during first quarter IBTS. Catches are standardized to abundance of herring larvae per m^2 and from those values an index for larval abundance is calculated for the entire survey area. Only the offspring of the autumn spawning components (Orkney/Shetlands, Buchan, and Banks) are considered for the index because those components have dominated the North Sea stock in the past and larvae of the winter spawning Downs component have not passed the stage of high and variable mortality. When the index calculation algorithm was formulated, it was assumed that small Downs larvae are only abundant south of 54°N. Consequently, only for stations south of that latitude, an exception rule is implemented. The mean larval length for each of those stations is calculated, and if that value is < 20 mm all data from that station are excluded from the index calculation (for a more detailed description, see last year's WGISDAA report, ICES 2015).

With the increasing importance of the Downs component in total North Sea herring SSB it became apparent that the current algorithm for calculating the MIK herring larvae index became more likely to produce biased results. In addition, small larvae drifting beyond that boundary may cause problems in the index calculation. This was particularly true for the 2014 MIK survey when large numbers of small herring larvae, assumed to be originating from the Downs component resulted in an extraordinarily high MIK index (Figure 6.3.1.1).

That problem has again been dramatically highlighted during the 2016 Q1 IBTS. A large advection of herring larvae originating from the Downs component to waters west of Denmark again led to another apparently severely biased MIK index.

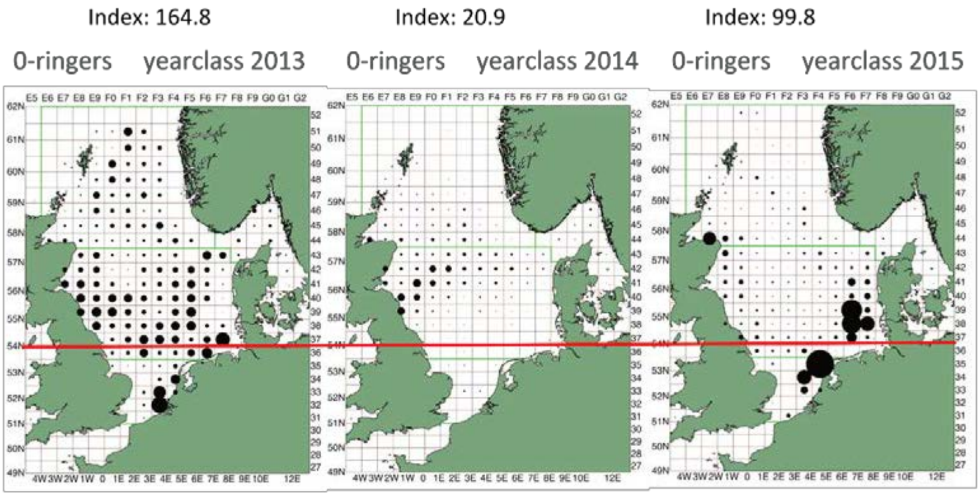


Figure 5.5.1: Examples of 3 consecutive years (2014–2016) of MIK index results for North Sea herring larvae. The distribution of the mean herring larvae abundance per each ICES rectangle is shown. Note the high abundances of the 2015-year class north 54°N (red line) in the eastern part of the North Sea. All of those correspond to the high advection of small larvae from the Downs component.

During first year's meeting (2015) of WGISDAA, these issues were thoroughly discussed and the working group recommended that in the first place the Herring Assessment Working Group (HAWG) should evaluate the importance of the MIK index for the assessment of the North Sea herring stock. HAWG discussed the recommendation and concluded that the index is still needed for recruitment forecasts and that HAWG will not refrain from using the index in future. Therefore, HAWG also recommended the implementation of a workshop especially dedicated to resolving the issues with the MIK survey. That workshop (Workshop on Herring Larvae Surveys, data needs and execution, WKHERLARS) was held in November 2016 at Wageningen Marine Research, IJmuiden, the Netherlands.

A new index calculation algorithm was recommended by WGISDAA during its second meeting in 2016. The algorithm should follow a 2-phase approach: in a first phase, each year in the short time between the MIK survey and annual meeting of the Herring Assessment Working Group, a preliminary MIK index should be calculated based on simple, fixed rules. Larvae of Downs origin should be excluded. The samples of all stations should be used while an exclusion rule is only applied to length classes within each sample. That exclusion rule should only apply at stations in an area where the occurrence of Downs larvae is most likely. For a preliminary index, the boundary shall be a fixed, but not necessarily straight line. In a second phase and prior to the HAWG meeting in the following year, the index will be refined based on modelling results. Accordingly, the index calculation will follow 3 steps:

1) Redefine the area where exclusion rules for larvae of Downs origin apply. In the preliminary index the exclusion rule should apply in the area between 6° E and the Danish and German coasts and south of 57° N as well as south of 54° N in the rest of the North Sea.

2) Inclusion of larvae greater than 18 mm in areas where currently the rule is if the mean length of larvae at a station is < 20 mm all data from that station are to be excluded from inclusion in to the index value. The 18 mm minimum size, should exclude any 'Downs' larvae but will allow larvae from the other spawning grounds to be included

in the index calculation. The drift modelling indicates that larvae originating in spawning grounds other than the Downs can occur south of 54°N.

3) Utilize the drift modelling results to inform any decisions in the southeastern part of the North Sea as to the likelihood of areas being significantly influenced by winter spawned (Downs) larvae. These data will provide additional support for any decision made in 1) and 2) above, either for fixing cut off values for both, larval size and geographic boundary, in calculating a preliminary index, but also for their revision while providing a final index in each year. Further refinements in the modelling framework e.g. more realistic spawning times, growth rates and mortality rates will provide more conclusive support for the decisions concerning inclusion in to the index and origins of larvae.

The newly defined algorithm for the preliminary index had to be applied on the MIK herring larvae data for the time series back to 1992. It was, therefore, necessary to check and correct the data for consistency. This was done in collaboration with all survey participants. It turned out that the French MIK data could only be used back to 2008 while before that year there were serious problems with misreporting the small larvae < 20 mm and confusion of herring with other clupeoid larvae like sardine and sprat. All French records before 2008 had therefore been excluded from the data. The results in contrast those of the old algorithm are shown in figure 6.3.1.2, and for the time series back to 2008 in table 6.3.1.1. It becomes apparent, that the new calculation method brings down the conspicuous indices for the 2013 and 2015 yearclasses while it doesn't alter the trajectory of the old index much. Particularly in the early 2000s when the contribution of the Downs component to the North Sea herring stock was stronger, the correspondence with the 1-ringer IBTS index remained very weak.

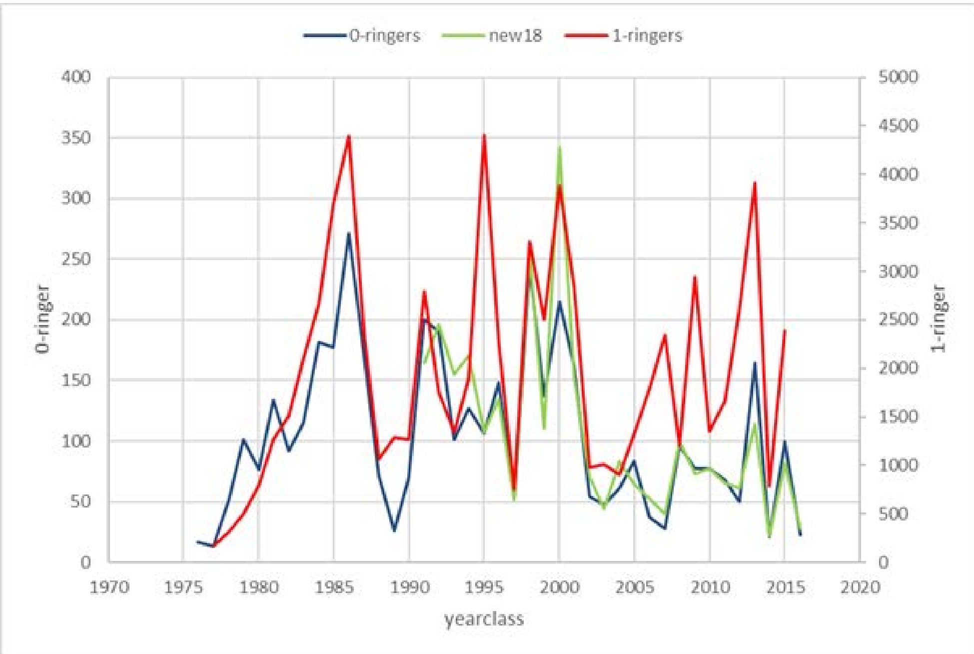


Figure 5.5.2 Time series of 0-wr and 1-wr indices. Year classes 1976 to 2016 for 0-wr fish, year classes 1977-2015 for 1-wr fish. The green line shows the results of the new MIK index algorithm applied to the data 1992 – 2017 (year classes 1991 – 2016)

Table 5.5.1 Time series of 0-ringer indices year classes 2008 – 2016 for traditional and new18 index algorithm.

yearclass	traditional	new18
2008	95.8	99.2
2009	77.1	73.5
2010	77.0	77.6
2011	68.0	65.1
2012	50.4	61.2
2013	164.8	113.8
2014	20.9	21.7
2015	99.8	81.2
2016	22.8	27.8

5.4.1 Adaptions to the 2016 MEGS-survey

5.5 Update on the 2016 MEGS-survey

During its 2014 meeting, the Working Group on Mackerel and horse mackerel Egg Surveys (WGMEGS) asked WGISDAA for support in conducting the Atlantic mackerel and horse mackerel eggs survey while taking account of the extension of both spawning area and season at limited survey resources without creating a biased SSB index.

The mackerel egg survey (MEGS) is carried out triennially and delivers the only fishery-independent data for the assessment of Northeast Atlantic mackerel and horse mackerel. Total annual egg production 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. during the first half of each cruise the assigned survey area is sampled on every other transect and the remaining transects filled in during the second half. In addition, survey participants are requested to follow an adaptive strategy while following their transects, i.e. each transect should only be finished after encountering zero counts of freshly spawned mackerel eggs. Total annual egg production (TAEP) is then calculated from stage 1 egg abundance data. With the fecundity values estimated during the same survey, the TAEP of mackerel is then converted into an SSB value for mackerel, which is used an index in the assessment. For horse mackerel, the TAEP is used directly as an index for SSB in the assessment (for more details see first interim report of this WGISDAA term, ICES 2015a).

Extension of the spawning time and area at limited survey resources necessitated leaving out of every other transects in order to achieve a full coverage. Additionally, it also became increasingly difficult to represent the annual egg production for both target species of the survey, mackerel and horse mackerel, as their time of peak spawning appears to drift further apart. This raised concern that the current survey design will not be able to provide reliable and defendable estimates of TAEP and SSB for mackerel and TAEP for horse mackerel, in future.

During its 2015 meeting, WGISDAA recommended that WGMEGS should consider investigating temporal and spatial variability of mackerel and horse mackerel egg production and to estimate the contribution of northwestern spawning extension to total annual egg production. This should indicate areas where effort savings could be made that have minimal impact on the precision and accuracy of the index. WGMEGS should also replace the double zero rule for transect termination by a more meaningful one.

WGMEGS was in the middle of survey preparation for 2016 when WGISDAA's recommendations were published. The last recommendation, however, was adopted for this year's survey. No participating country would, thus, be forced to conduct an unnecessarily long transect at the cost of having less survey time available for the core spawning area. Thorough sampling in the core area was considered by WGISDAA being of higher importance for an unbiased SSB index than a full coverage of the total spawning area. WGMEGS was also able to recruit new survey participants for a better coverage of both survey area and time. The necessary ship time will be provided by the pelagic fishing industry at the cost of scientific quota. Need to discuss the efficacy and utility of this once results are available.

The 2016 MEGS was carried out after recruiting more ship's time from the fishing industry, aided by the modelled predictions of egg production from Mark Payne's (DTU-Aqua, Denmark) spatio-temporal mackerel egg production model, and following the advice of WGISDAA. All this enabled the survey to cover the entire prospected mackerel spawning season between February and July as well as the enlarged core spawning area between Portugal and the Faroese.

Contrary to the expected early spawning peak in early March, the 2016 MEGS revealed that not only was the major mackerel egg production observed in May but also in the most northern and northwestern part of the survey area (Figure 5.4.2.1). The western and northwestern boundaries in the Rockall Plateau area and southwest of Iceland, respectively, could not be sufficiently established.

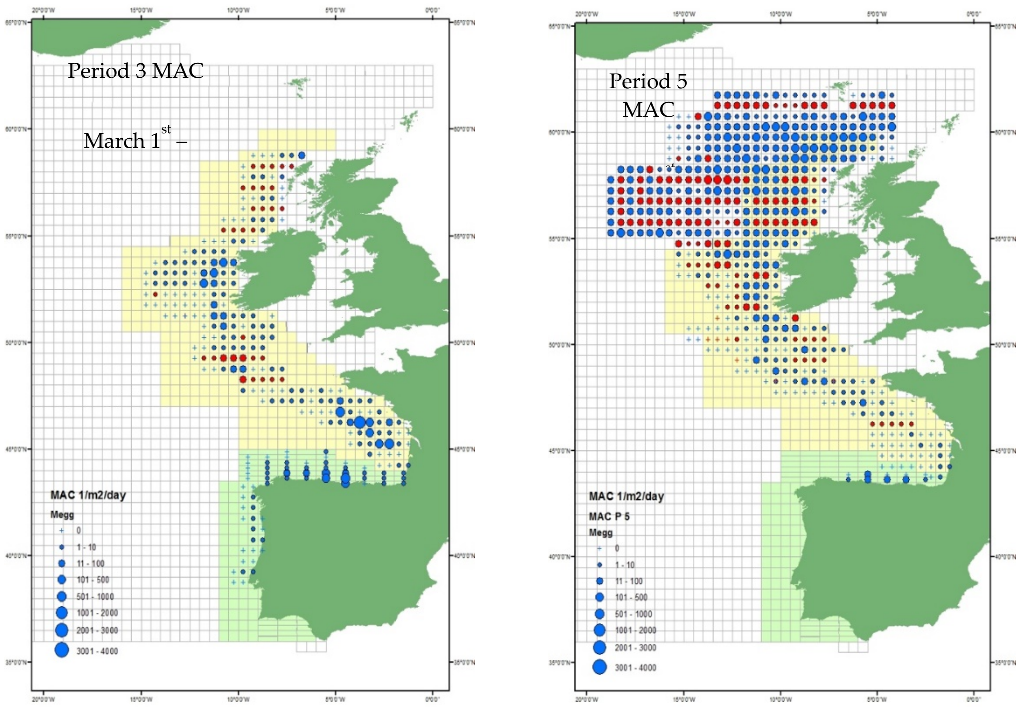


Figure 5.4.2.1: The daily mackerel egg stage 1 production per m2 per day in each half rectangle in MEGS survey periods 3 (where peak spawning was expected) and period 5. Note that in period 5 the western boundary (in the area of the Rockall Plateau) and northwestern boundary (in the area southeast of Iceland) of mackerel egg production could not be established.

In order to test whether the present survey and its current design continue to adequately capture the spawning activity of NEA mackerel, the contribution of the sampled ICES half-rectangles to 50 and 90 % of the total annual egg production (TAEP) was calculated and explored (Figure 5.4.2.2).

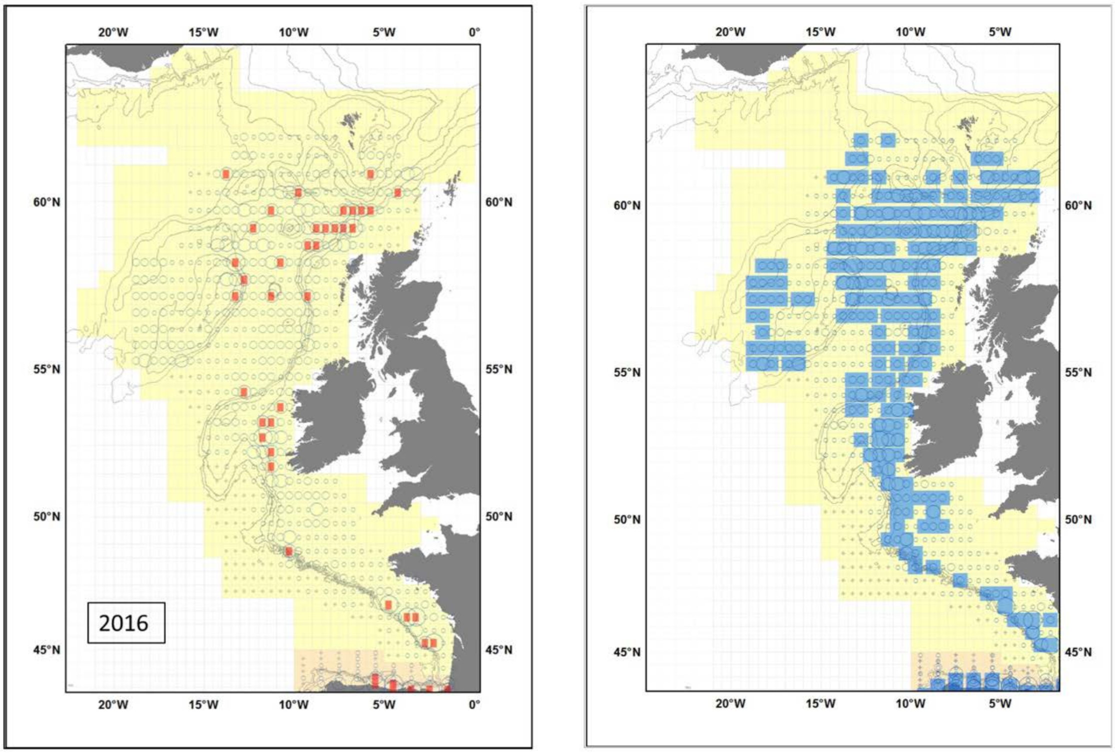


Figure 5.4.2.2: The ICES half-rectangles that cumulatively contributed to 50 % (red) and 90 % of total annual egg production.

The half-rectangles that contributed to 50 % of mackerel TAEP lie all well within the surveyed area as does the majority of all half-rectangles that contributed to 90 % TAEP. Only in the Rockall Plateau area the delineation of the spawning area does not appear to be well established. However, the WGMEGS concluded that despite the unexpected distribution of mackerel egg production, it was representatively captured with the 2016 survey.

In order to eliminate or at least minimize the effects of such unexpected spawning events on the 2019 MEGS, the WGMEGS has undertaken and will undertake several measures:

- Explore potential for utilising existing surveys within Northern region to collect plankton samples during interim period: e.g. Iceland's Q2 Herring Survey/International Ecosystem Summer Survey in the Nordic Seas
- The Marine Institute (IE) and Marine Scotland Science (SCO) will carry out limited surveys to undertake transects within Northern/Western areas in May/June of 2017/2018 (MI/MSS)
- Mackerel egg data from SAHFOS CPR samples, of which the monthly Scotland – Iceland transects are of particular interest.
- Mackerel egg data from the IEO monthly transects in spring
- Environmental Niche Modelling of Mark Payne (DTU-Aqua) to continue in 2017/18 (see ICES 2016b for a description of the modelling)
- Fecundity sampling on commercial trawlers – 2017/2018

Those activities should monitor the spawning behavior of mackerel during the between survey years in 2017 and 2018, and should help with planning of the 2019 survey design (see also ICES 2017a for more details).

Even though in some of the northwestern area the survey was not able to delimit the survey area, WGISDAA concluded that its advice to concentrate the effort on the core spawning area was still valid for an unbiased estimate of mackerel TAEP. This was corroborated by the fact that almost all of the rectangles contributing to 90 % TAEP lie clearly in the surveyed area. For the between survey years, WGISDAA suggested to not only investigate the potentially expanded spawning area in the Northwest, but also roughly monitor the spawning activity in the traditional areas west of Ireland and in the Celtic Sea. Other than the Scotland – Iceland CPR transect should also be considered for monitoring the spawning behavior of mackerel between the survey years. That would give a more complete picture of mackerel spawning phenology enabling for a better planning of the survey.

It was also suggested that WGMEGS considers using the newly developed PIA system to continuously sample fish eggs between standard sampling stations. The PIA is an automated plankton analysis device that could be attached to the ship's water intake and could be trained to automatically identify and count mackerel eggs between stations. That way it could be determined whether the detected by station egg distribution gives a good representation of the true spawning activity.

5.6 Various strands of work on the IBTS survey

5.6.1 Potential effect of a Q3-IBTS redesign by WKPIMP on commercial fisheries indices for stock assessment.

Background:

WGISUR requested information regarding methods that would make it possible to maintain timeseries of important commercial species assessed in ICES stock assessments while altering the survey design and effort levels in the North Sea Q3 IBTS to develop a more ecosystem integrated approach. The WG was unable to offer general advice on this question because of the extremely wide range of options for change. Without further specific information as to the desired changes the response would need to be too extensive and time consuming. Instead the WG decided to use the draft proposal developed by WKPIMP 2016 as a starting basis for assessing the appropriateness of the design and the possible consequences on the delivery of survey information for stock assessments.

Survey designs:

Current design: The current Q3-IBTS design is based on a systematic design based on ICES rectangles in which sampling is conducted randomly, with most rectangles having replicate samples from different vessels. Practically at least sampling within rectangles has been less than random with several countries / ships returning to the identical tow positions year after year due to concerns over gear damage at other possible locations.

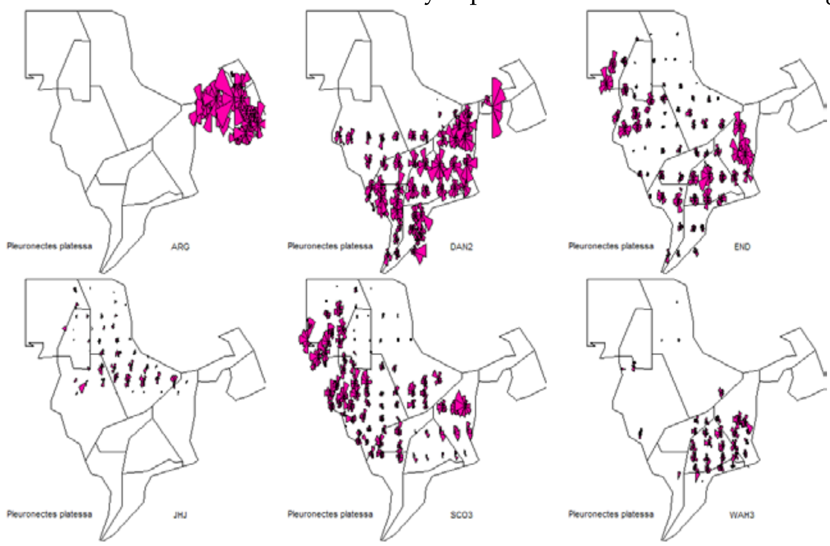
Proposed design: The new survey design is planning to use the same GOV gear currently used and envisions no change to the spatial extent of the survey area, simplifying the maintenance of time series significantly. The intention is to stratify the area into ecologically relevant strata abandoning the current rectangle based systematic design for a random design for a substantial number of possible safe towing positions.

Qualitative determination of effects and means to quantify likely impact on assessments:

Most indices used in stock assessments based on the Q3-IBTS survey simply use the mean abundance at length in a rectangle adjusted to an hourly tow and converted to the mean abundance at age using age length keys for the associated species specific ALK areas. The resultant rectangle abundance is then summed by age over the species specific survey area to derive the annual age based index. This method assumes that all rectangles are sampled, so where intended effort levels are insufficient to cover all rectangles in all years the new design will fail to produce values consistent with the historic time series. WGISDAA qualitatively examined the data available in the current design to examine if there may be potential alternate methods of index calculation that would be consistent across both designs, with the following conclusions:

Plots of CPUE at each station over the last 11 years for the important stocks assessment species in which the Q3-IBTS data is or could reasonably be used were examined for coherence with the proposed stratification scheme (some examples shown in Figure 5.5.1). Variability in total abundance within strata generally seems to be lower than variability between strata. The stratification scheme does not appear to be ideal for all species at current abundances. The differences in the distribution of individual species suggest that no one stratification would be ideal for all 10 species at reasonable effort levels, but the proposed scheme with possible modifications is likely to be considerable more efficient than a completely random design. If YEAR and Stratum variance components are orthogonal and station abundances within a stratum are more or less random, then it should be possible to derive continuous time series of abundance using the GLM approach despite the change in survey design. WGISDAA should derive quantitative descriptions of the variance components for each species for the proposed design are required for the abundance estimates.

Considering abundance alone may be insufficient to evaluate the effect of design changes, because some indices are used at age. Since the major change in the survey design, at least on an initial examination appears to be the overall change in effort, it is necessary to examine if such reduced effort would still provide sufficient randomly collected otolith samples to reflect the relative abundance of age-at-length. For those stocks where the current index merely represents the sum of the rectangle abundance



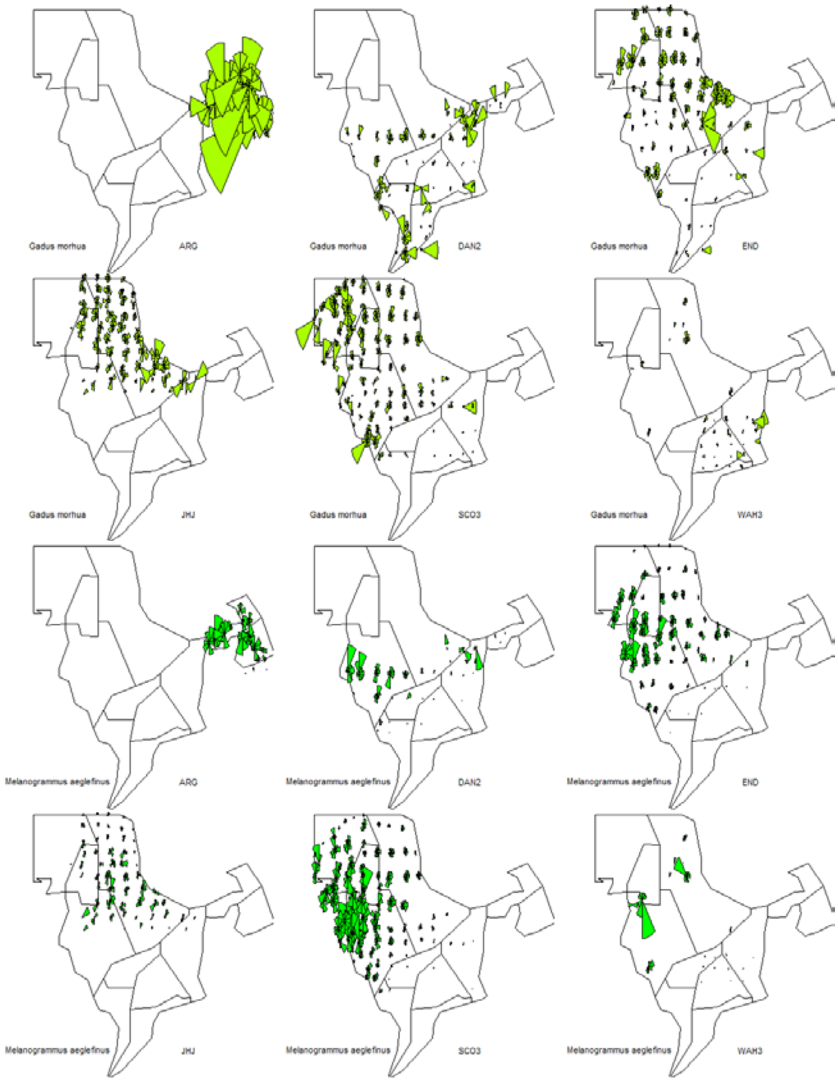


Figure 5.4.1: 11-year IBTS survey abundance plots by vessel, overlayed on the stratification scheme recommended by WKPIMP. Pie-slices represent the 11 years (2005-2015 clockwise starting at 12:00) with the radius of the slice proportional to the square root of CPUE. Plaice (top) indicate a temporally stable pattern with low variability within strata. Cod (middle also indicate a reasonable coherence with the strata, but also indicate some temporally random large catches spread across the coastal strata. Haddock (bottom) are consistently encountered in greatest abundance in the NW also showing occasional large catches, but temporally consistent suggesting cohort effects.

it is likely that a decrease in effort will merely decrease the precision of estimates, particularly for older rare ages. Some of this could be compensated by larger subsamples of otoliths at length. However, where lengths themselves become rarer due to fewer samples increases in stochasticity are unavoidable. Fewer larger samples (longer duration) may be an option if but may lack the independence of otolith samples. Generally, we would expect a decrease in precision with a substantial decrease in effort. Dependent on the appropriateness of the current roundfish ALK strata versus the new ecosystem strata historic estimates of trend may change and precision of estimates under the new design may be slightly lower.

The Q3-IBTS cod index already uses annual regression methods to derive the index. Unlike the other index method, it does not assume orthogonality between distribution in time and space. Instead it assumes that samples are collected randomly, and that covariates used (location and ship, for example) are uncorrelated. These assumptions may be violated in historic data and the effect of this should be tested. In terms of future collections, the proposed plan is much more consistent with the way the data is analysed currently specifically attempting to reduce the correlation between different variables in the design and maintaining randomness where ever possible. Potential disadvantages of the new design would be that the distance between stations is likely to be much more variable than in the systematic design so that the spatial smoothing may be more imprecise. The problem may be exacerbated for the age information particularly for rarer ages. WGISDAA attempted to bootstrap the current data set using the index methodology at different effort levels to try to identify the contribution of these effects to the uncertainty in the index. There were insufficient resources available to complete this work prior to the WG, but the necessary tasks were identified and discussed at the WG. It is hoped that this work can be completed inter-sessionally.

5.6.2 Effect of reduced tow duration on abundance indices and MSFD indicators in the 3Q NS-IBTS in 2015 and 2016

IBTSWG 2015 agreed to conduct an experiment on tow duration during the NS-IBTS Q3 2015. Evidence exists for other surveys that benefitted from changing to shorter tow duration (see IBTSWG Report 2015 section 10.3. for a thorough discussion on the pros and cons; ICES 2015). The majority of the IBTSWG considered the risk that the experiment would impair the quality of the long-term data to be outweighed by the potential advantages in survey coverage and efficiency. The sample numbers for comparison in 2015 proved to be inconclusive to determine significant effects so the experiment was continued through 2016. Since this was largely supported by preliminary analyses conducted on the data of the 2015 experiment, the IBTSWG decided during its 2016 meeting to continue using a mixture of 30 and 15 min tows during the NS-IBTS Q3 2016 (ICES 2016c).

In order to warrant a thorough comparison with the standard methodology, a plan was formulated whereby in each ICES rectangle one of the two assigned hauls would remain a 30 min tow whereas the duration of the second would be reduced to 15 minutes. England and Sweden decided to retain tow duration for all of their hauls at 30 min and thus no combinations of the two tow durations were planned for the Skagerrak, which is almost exclusively covered by Sweden. The 15 min tows were either placed in rectangles in which England was required to conduct a 30 min tow and in the remaining rectangles not allocated to England, a balanced share by rectangle of nominal 30 and 15 min tows were distributed amongst the other countries. This was achieved for most rectangles, but the overall proportion of 15 min tows conducted by Denmark, Germany, Norway and Scotland differed between 53 and 76 % in 2015 and between 54 and 76 % in 2016.

An update of the analyses including the data for both years which accommodated the suggestions from WGISDAA in 2016 (ICES 2016b) was presented to this meeting (WGISDAA 2017 WD's #1 (Devine and Pennington) and WD #3 (Wieland)). To provide a balanced overview WGISDAA considered all information presented to together with some ancillary information discussed during the IBTSWG meeting held earlier this year (IBTSWG 2017 WD's #4 (Jaworski et al.) , WD #6 (Hatton et al.) and WD #7 (Moriarty and Burns); ICES 2017b). A number of theoretical and applied statistical considerations were provided by different participants of the afore mentioned meeting which are summarized briefly here as:

Modelled index approach (age-based):

For cod, whiting and plaice GAM's with negative binominal error structure were compared with GAM's assuming a Poisson error distribution. In all cases, the negative binominal GAM's provided a more satisfactory fit performed better that the Poisson GAM's, and haul category (short or long) had no significant effect in the statistically most appropriate models for a given species and age group once the difference in swept area had been accounted for. In further analysis (WGIBTS WD4) no significant effects of haul category were found for negative binominal GAM's for the different age groups of haddock and Norway pout.

Modelling indicated that there was a significant effect of tow duration on the presence and absence of some size groups of the species. For an in-year comparison the model based approach is generally appropriate, but there seems to be insufficient power in the analysis to detect differences for rare individuals. Without an appropriate method of compensating for the difference in tow duration (off-set function in GAM) for the binomial case model, delta based models will suffer from inconsistency problems in detecting timeseries effects, because the magnitude of the effects is density dependent. Other likelihood based modelling approaches e.g. tweedy distributions are more appropriate for these cases and should be investigated.

Design based pairwise comparisons:

Pairwise (by rectangle) comparisons of catch rates and mean length within rectangles for nine species (Herring, mackerel, cod Norway pout, plaice, saithe, whiting, sprat and haddock) did not show any significant effects except for catch rates of mackerel and the mean length of whiting in 2016. Hence, the two complementary approaches do not indicate that the short tows are in general less efficient than the long tow. However, the application of both methods is not entirely unproblematic e.g. because the allocation of the short tows to the different countries was not balanced with regards to the ship and location effects. In addition, and the pairs within a rectangle may differ in depth and have not necessarily been conducted simultaneously as a result of ships availability within the extended survey period of about two months. The general lack of significance of an effect may therefore relate to a lack of power because the variation caused by the above effects may be too large to detect the differences at the current sample size.

Biodiversity – species richness:

NS-IBTS data has recently been used in the OSPAR assessment of biodiversity and food chains. Although currently based largely on the length composition of fishes. Further indicator development may attempt to examine changes in species richness. The mean number of species per haul was consistently lower for 15-minute tows compared to the standard tow. When aggregated over the entire survey species richness was highest in 2015 and 2016 when 15-minute tows were implemented. It should be investigated whether there are spatial effects discernible as the use of 15-minute tows permitted a wider spatial coverage than previously. Irrespective of the cause it presents a dilemma to biodiversity assessors in trying to trade off the consistency of timeseries provided by 30-minute tows with the greater species coverage possible with 15-minute tows.

Generally, one would expect the central tendency of a proportional indicator such as the large fish indicator (LFI) to be independent of tow duration. It has been proposed though that longer tows catch a greater proportion of large individuals because these will only be sufficiently exhausted to be caught over longer distances. Instead, the mean LFI for 30-minute tows on average was lower than that for the short tows in 2015-2016. This should be investigated more closely on a haul by haul basis to eliminate spatial effects, but it presents two problems, one with regards to the consistency of time series and two in respect to our understanding of the interpretation of LFI.

Theoretical considerations:

Generally, the lack of a significant effect is not an indication of the absence of an effect so caution may be appropriate. The power of the above analyses has not been investigated and without such further analysis it is not possible to disregard such concerns. A concurrent analysis of power should be conducted commensurate with all these analysis.

The degree of autocorrelation is another consideration highlighted in WD13. The example provided is not directly related to this dataset, nor is it evaluated against the metric of interest to stock assessment (numbers-at-age) it is therefore inappropriate to assume that auto correlation is universally high. However, if this is the case for the NS-IBTS then the statistical argument holds and the ecological consequences of killing a greater number of fish for no benefit would suggest that shorter tows would provide a desirable alternative.

Data on 0-minute tows as classified by WGIBTS indicate significant catches are possible when the net is set and immediately hauled. Shorter tows will proportionately be more affected by this additional variation compared to longer tows. Undoubtedly this causes a bias in the CPUE estimates, which should be largely irrelevant to stock assessments which treat the index as a relative abundance estimate. Whether the effects are random with regards to other parameters remains a question and whether it introduces significant additional variability remains to be investigated. Without a clear understanding of the bias caused by this effect it is not clear whether this is a significant concern that has not been tested on the specific dataset. In addition, the ‘end effect’ (catch outside the nominal tow duration) may cause problems when using a mixture of short and long tows, in particular if it is correlated with depth. Furthermore, an effect of haul category on species richness (higher number of species per km2 in the long tows) and on the Large Fish Indicator, LFI, (higher value for the short tows) was found. Based on these reservations it has been decided by the 3Q IBTS participants not to continue with the mixture of short and long tows in 2017.

The independent results presented by participants from the IBTSWG provide a broad-spectrum analysis of the issues to consider when making a decision to change the tow duration. The diversity of metrics compared (age based indices to species richness) and the methods used (simple means to complex likelihood models) makes it difficult to compare the results and come to specific recommendations in the absence of a clear prioritisation of the output metrics. The difference in conclusion from the various bits of work highlight the difficulty in making survey decisions when multiple objectives are considered and WGISDAA feels that this is highly beneficial.

When taking support of fisheries management (cat 1) as the primary purpose of the NS-IBTS then one has to conclude that there is little hard evidence to suggest that shorter tows will have a major impact on advice. IBTSWG has terminated the experiment for now due to concerns over inconsistency of timeseries so it is unlikely that experimental sample sizes will increase. However, failure to detect a difference in the GAM analysis with sample sizes equivalent to the annual sample size suggest that where present, these differences are smaller than the average annual variation in indices.

WGISDAA would suggest that with potential upcoming changes in the IBTS survey including new ships, new gear, and potentially a more ecosystem focused redesign such changes are likely to be more severe than the change in tow duration. More work will need to be done on these issues to maintain timeseries consistency and IBTSWG should consider making a change to shorter tow durations at that time. In principle, there seems to be little wrong with the idea of 15-minute tows if 'end effects' can be kept to a minimum and the major concern is over time series consistency issues. For most stock assessment purposes, it seems possible to minimise disruption to indices by using model based approaches to index development.

Despite some reservations, WGISDAA considered the conducted data analysis as sufficient appropriate and suggest that 15 min tow duration as standard for all tows could be introduced when also other changes in the 3Q NS-IBTS such as a change of the survey gear and the stratification as well as station allocation scheme may be implemented the future.

5.6.3 Stratification and post-stratification of IBTS survey data. Developing some tools to inform on options.

Options for (post)stratification of the IBTS

An initial analysis on options for stratification versus post-stratification of the North Sea IBTS has been discussed. This output from the IBTSWG 2017 has been presented in order to collate advice from the group regarding the approach as such, and the statistical methods applied.

In this conceptual contribution, 'stratification' was considered as a possible perspective, in which the design of future surveys would be changed and lead to a modified allocation of stations within the survey area. In contrast, 'post-stratification' would mean that the design of the survey itself remains as it has traditionally been performed, while only the analysis of the data would be changed based on spatial subunits of the survey area. The WG considered that the quantitative evaluation of the gains in precision achieved by post-stratification could be used to infer potential gains in survey efficiency by similar stratification in the development of potential future design options.

There was some debate regarding the appropriate metric to measure improvements through stratification by with several being considered. There was insufficient time to reach a conclusion based on theoretical principles and no time to do the appropriate calculations. This will be taken up in the next three-year WG period.

Proposal for further analyses regarding IBTS survey design and data analysis

WGISDAA identified several aspects in the design of the IBTS, which require in-depth analyses in order to evaluate the potential for increasing the survey's efficiency with respect to current needs for its products. Research questions include in particular:

1. How can stratification or post-stratification be applied to increase the survey's precision and efficiency with regard to prioritized products?
2. How can efficiency in the collection of otoliths or other biological samples be gained to best support new assessment methods, and to address any important existing spatial structure in the data (e.g., regionally different growth rates for target species).
3. What guidance can be given for switching to a new gear for the IBTS? (The current gear is too fragile for many areas in the North Sea, limiting the options for random selection of stations).
4. What guidance can be given for commissioning new research vessels? (Several IBTS partners plan to switch to new survey vessels during the upcoming years.

What recommendations should be given for planning of the transition phase; is inter-calibration between old and new vessels necessary?).

In order to create sufficient momentum and to answer these questions before foreseeable changes in the survey come into effect, the group decided to pursue options for funding in the respective national labs, via EMFF / DCF.

Results from these investigations could then be used for informed decisions during a proposed '**Workshop on Impacts of planned changes in the North Sea IBTS'** (WKNSIMP).

5.6.4 Boot strapping index calculation to inform on precision and effectiveness of surveys.

Bootstrapping is a powerful and versatile, yet relatively simple technique for estimating the accuracy (including bias, variance, confidence intervals, and prediction error) of some estimate based on samples. It relies on the empirical distribution of the samples, and hence does not require assumptions about the particular distributions (e.g. normal distribution) of the data or the residuals. Since data from surveys are typically not normally distributed, and typical output such as total numbers-at-age are complicated functions for which variances are difficult to derive, bootstrapping is a very useful technique to consider in this respect. It is based on resampling data (or residuals) with replacement. When the measurement error is correlated in blocks (e.g. hauls) the resampling should be done on block level in order to keep the inherent correlation in the data. This is sometimes known as block or cluster bootstrapping. The DATRAS R-package (https://www.rforge.net/DATRAS/Tutorial.html) is designed such that it is easy to perform block bootstrapping on haul level. The working group have produced some example R scripts that demonstrates this (numbers- and weight-at-age for North Sea Cod IBTS data). The examples can be found in "Software" folder on the WGISDAA 2017 SharePoint. These examples are intended merely as illustrations and starting points for further development, and the working group intends to add more examples over time. It should be emphasized that for small sample sizes bootstrap methods tend to under-estimate the variance, and if the correct blocking is not chosen, such that samples are not independent, the results may be seriously biased.

6 Cooperation

The WG had no formal cooperation with other working groups or advisory structures during its three-year cycle, because this has not proven very successful in meeting the objectives of the group. Instead the WG has largely relied on the ‘individual-based’ approach, where shared members with other WGs have informed WGISDAA of their needs and those same individuals have then taken this results / recommendations back to those WGs. The same model has been much less successful for the assessment WGs.

7 Summary of Working Group self-evaluation and conclusions

The active participants in this WG have been variable which because of its purpose was always expected. However, the numbers have also shrunk considerable since its inception 6 year ago. Never -the-less a group of stalwart participants has developed over the last 3 years with very good connections to a number of survey WGs. Because of this and because the ICES need for better use of survey information in science and advice the group decided it was worth continuing despite the fact that it had been hard to make progress on cooperating with assessment working groups which had been one of the major objectives of the group from the outset.

8 References

Berg, C. W., & Kristensen, K. (2012). Spatial age-length key modelling using continuation ratio logits. Fisheries Research, 129-130(2012), 119–126. doi:10.1016/j.fishres.2012.06.016

Berg, C. W., Nielsen, A., & Kristensen, K. (2014). Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. Fisheries Research, 151, 91–99. doi:10.1016/j.fishres.2013.10.005

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). Time series analysis forecasting and control. Wiley.

Cressie, N., & Wikle, C. K. (2011). Statistics for spatio-temporal data. Wiley.

deYoung, B., Barange, M., Beaugrand, G., Harris, R., Perry, R. I., Scheffer, M., & Werner, F. (2008). Regime shifts in marine ecosystems: detection, prediction and management. Trends in Ecology and Evolution, 23(7), 402–409. doi:10.1016/j.tree.2008.03.008

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography, 16(2), 129–138. doi:10.1111/j.1466-8238.2006.00279.x

Fissel, B., Dalton, M., Felthoven, R., Garber-yonts, B., Haynie, A., Kasperski, S. Seung, C. (2015). Stock assessment and fishery evaluation report for the groundfish fisheries of the Gulf of Alaska and Bering Sea / Aleutian Islands area : economic status of the groundfish fisheries off Alaska, 2014. NPFMC Economic SAFE.

Hoeting, J. A. (2009). The importance of accounting for spatial and temporal correlation in analyses of ecological data. Ecological Applications, 19(3), 574–577.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? Ecology, 74(6), 1659–1673. Retrieved from http://www.esajournals.org/doi/abs/10.2307/1939924

Maunder, M. N., & Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. Fisheries Research, 70(2-3), 141–159. doi:10.1016/j.fishres.2004.08.002

Ono, K., Punt, A. E., & Hilborn, R. (2015). Think outside the grids: An objective approach to define spatial strata for catch and effort analysis. Fisheries Research, 170, 89–101. doi:10.1016/j.fishres.2015.05.021

Perretti, C. T., Munch, S. B., & Sugihara, G. (2013). Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. Proceedings of the National Academy of Sciences of the United States of America, 110(13), 5253–5257. doi:10.1073/pnas.1216076110

Stewart, I. J., & Monnahan, C. C. (2016). Overview of data sources for the Pacific halibut stock assessment and related analyses. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2015 (pp. 99–187).

Stewart, I. J., Monnahan, C. C., & Martell, S. J. D. (2016). Assessment of the Pacific halibut stock at the end of 2015. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2015 (pp. 188–219).

Stewart, I. J., & Martell, S. J. D. (2016). Appendix: Development of the 2015 stock assessment. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2015 (pp. 623–768).

Thorson, J. T., Shelton, A. O., Ward, E. J., & Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science. doi:10.1093/icesjms/fsu243

Wilberg, M. J., Thorson, J. T., Linton, B. C., & Berkson, J. (2009). Incorporating time-varying catchability into population dynamic stock assessment models. Reviews in Fisheries Science, 18(1), 7–24. doi:10.1080/10641260903294647

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 International Bottom Trawl Survey Working Group (IBTSWG), 23-27 March 2015, Bergen, Norway. ICES CM 2015/SSGIEOM:24. 278 pp.

ICES 2016a. First Interim Report of the International Bottom Trawl Survey Working Group (IBTSWG), 4-8 April 2016, Sète, France. ICES CM 2016/SSGIEOM:24. 292 pp.

ICES 2016b. Interim Report of the working Group on Improving use of Survey Data for Assessment and Advice (WGISDAA), Hamburg, Germany, 12-14 July 2016. ICES CM 2016/SSGIEOM:28. 31 pp.

ICES 2016c: Report of the Workshop on North Sea herring larvae surveys, data needs and execution (WKHERLARS2). ICES CM 2016/SSGIEOM:33

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

ICES 2017b. Interim Report of the International Bottom Trawl Survey Working Group. IBTSWG Report 2017 27-31 March 2017. ICES CM 2017/SSGIEOM:01. 337 pp.

Annex 1: List of participants

Name	Institute	Country (of institute)	Email
Sven Kupschus	Cefas	UK	sven.kupschus@cefas.co.uk
Matthias Kloppmann	Thuenen-Institute of Sea Fisheries	D	matthias.kloppmann@thuenen.de
Kai Wieland	DTU Aqua	DK	kw@aqua.dtu.dk
Casper Berg	DTU Aqua	DK	cbe@aqua.dtu.dk
Stan Kotwicki (by correspondence)	NOAA	US	stan.kotwicki@noaa.gov
David Stokes	Marine Insitute	IR	david.stokes@marine.ie
Anne Sell	Thuenen-Institute of Sea Fisheries	D	anne.sell@thuenen.de

Annex 2: Recommendations

No recommendations from WGISDAA.

Annex 3: WGISDAA terms of reference

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	ICES HQ, Copenhagen, Denmark	Interim report by 20 August to ACOM / SCICOM	
Year 2019	TBD July	Copenhagen, Denmark	Interim report by Date Month to ACOM / SCICOM	
Year 2020	TBD July	Copenhagen, Denmark	Final report by Date Month to ACOM / SCICOM	

ToR descriptors[1]

ToR	Description	Background	Science Plan topics addressed	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	26, 31		

[1] Avoid generic terms such as "Discuss" or "Consider". Aim at drafting specific and clear ToR, the delivery of which can be assessed

		in other assessments.			
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” options for survey working groups.	31		
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		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	WGISDAA has had difficulty in attaining wider participation in its work			

	learned from the specifc cases dealt with in TOR a and b

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

Annex 4: Copy of Working Group self-evaluation

1. Working Group name: Working group on improving survey data for analysis and advice
2. Year of appointment: 2015
3. Current Chairs: Sven Kupschus
4. Venues, dates and number of participants per meeting. ICES HQ, Copenhagen, Feb 2015 (6); Thune Inst, Hamburg, Jul 2016 (7); ICES HQ, Copenhagen, July 2017 (6).

WG Evaluation

5. If applicable, please indicate the research priorities (and sub priorities) of the Science Plan to which the WG make a significant contribution.
6. The WG has:

   a) Investigated various methods of model based index calculations to overcome changes in survey design or implementations to provide consistent time series of catch at age information.

   b) Provided options and advice for improving efficiency and consistency of SSB based indices in the face of environmentally induced changes in larval / spawning distributions.

   c) Examined the effects of reduced tow durations in the NS-IBTS survey on both the mean and variance aspects of different advisory output metrics including biodiversity.

   d) Explored some methods of delivering increased survey precision through improved stratification methods.

7. The WG is not intended to directly provide advisory products, but it has contributed to advice indirectly through contributions to the upcoming NS herring benchmark assessment and the mackerel assessment. The contributions were in respect to survey design and evaluation.
8. The WG has not yet conducted any outreach activity outside the ICES network.
9. The main difficulty in achieving the full potential of the WG has been the inability to actively engage more stock assessment scientists with specific survey index related issues in the WG. The WG has made good progress in involving survey groups and providing a forum to discuss survey design and statistical questions as well as explaining the statistical assumptions underlying monitoring more generally.

Future plans

10. Does the group think that a continuation of the WG beyond its current term is required? (If yes, please list the reasons). Significant benefits to ICES advice and the wider ICES community remain while communication between assessment and survey WG remains impeded by limited overlap in interest and understanding of the monitoring and modelling expertise. Although the progress in engaging the assessment WG has been poor, significant improvements have been made with several survey WGs and the group plans to tackle the former through greater engagement in the benchmark system.
11. If you are not requesting an extension, does the group consider that a new WG is required to further develop the science previously addressed by the existing WG.

    (*If you answered YES to question 10 or 11, it is expected that a new Category 2 draft resolution will be submitted through the relevant SSG Chair or Secretariat.*)

12. What additional expertise would improve the ability of the new (or in case of renewal, existing) WG to fulfil its ToR? At present the WG relies on a very small pool of assessment / statistical modelling experts to provide answers and explanations to problems. This means a heavy workload as well as a constrained breadth of opinion and experiences being brought to bear on specific issues. Historically there was interest from acoustic survey experts in the WG, but this interest has waned in recent years and currently the group has no active acoustic survey specialists. More expertise in this area would improve the ability of the WG to provide advice in these matters.
13. Which conclusions/or knowledge acquired of the WG do you think should be used in the Advisory process, if not already used? (please be specific). The WG generally and the members individually have contributed to an improved survey group understanding of the data requirements for assessment purposes. Based on this the WG concludes that significant benefits remain by involving the assessment groups in this process. To accelerate the up-take and communication between assessment and survey groups the group aims to take a focussed workshop approach.