Resolution of biodiversity and assemblage structure in demersal fisheries surveys: the role of tow duration

Corresponding Author: Meadhbh Moriarty[1] m.moriarty@marlab.ac.uk

M. Moriarty[1/2], A.F. Sell[3], V.M. Trenkel[4], C.P. Lynam[5], F. Burns[1], E.D. Clarke[1], S.P.R. Greenstreet[1], C. McGonigle[2]

1. Marine Scotland Science, 375 Victoria Rd, Aberdeen AB11 9DB

2. Ulster University, Cromore Road, Coleraine BT52 1SA

3. Thünen Institute of Sea Fisheries, Palmaille 9, 22767 Hamburg, Germany

4. Ifremer, Rue de I'lle d'Yeu, BP 21105, 44311 Nantes Cedex 03, France

5. CEFAS, Pakefield Road, Lowestoft, Suffolk NR33 0HT

Abstract

An experiment during a fisheries independent survey in the North Sea was conducted to test whether sampling effort could be reduced without a significant loss in data precision. To examine potential effects of reducing tow duration from the standard 30 minutes to a proposed 15 minutes estimates of species encounter rates, species richness, and estimates of abundance, biomass and body size were analysed. Results show species richness estimates are lower in the short tow category. While biomass and abundance at length and body size are significantly affected by the change in tow duration, estimates of Large Fish Indicator, the Typical length and Mean-max length are not significantly affected by the regime change. The results presented here suggest that a reduction of tow duration did not optimise the resolution of biodiversity, and it may affect other survey objectives, such as, providing estimates of abundance or biomass for assessment of commercial species.

Keywords: North Sea, species diversity, biomass, abundance,body size, IBTS, International Bottom Trawl Survey, fisheries independent survey, linear mixed model

Introduction

Maximising survey resources in pressing economic conditions is a major concern in fisheries science. Since 1998 the standard tow duration for the International Bottom Trawl Survey (IBTS) in the North Sea has been 30 minutes (ICES, 2012; 2015a). Recently tow duration has come under scrutiny within the International Council for the Exploration of the Sea (ICES) community (IBTS Working Group; IBTSWG). An experiment to test the effect of moving to a tow duration of 15 minutes was initiated on the basis that this would:

1. Reduce the risk of gear damage during any single tow, thereby reducing the number of tows classified as invalid due to gear damage;
2. Potentially allow more tows to be carried out, which could improve the precision of species abundance indices, and
3. Potentially reduce overall survey time, with consequent savings in resources (ICES, 2015b).

Fisheries independent bottom trawl surveys have historically been undertaken to meet fisheries management requirements under the European Common Fisheries Policy (CFP) and Data Collection Framework (DCF). However, in 2014 the member states involved in the IBTS nominated their own surveys to fulfil monitoring obligations under the Marine Strategy Framework Directive (MSFD) (EC, 2008; 2010). Therefore, the IBTS must supply the data required to derive the ecological indicators necessary to assess the status of the whole fish community. Changes in survey design must now take account of the needs of stock assessments used for fisheries management purposes and should also consider the data requirements for the indicators used for broader ecological assessments (Jennings, 2005).

Several different types of indicators have been used to assess variation in the state of fish communities (Trenkel and Rochet, 2003; Shin *et al.,* 2010; Shannon *et al.,* 2010; Greenstreet *et al.,* 2012a). Some focus on community size composition such as the large fish indicator (LFI) (Greenstreet *et al.,* 2011; Shephard *et al.,* 2011; Modica *et al.,* 2014), mean fish weight (Greenstreet and Rogers, 2006) and the size spectra slope coefficient (Gislason and Rice, 1998). Others focus on species composition, capturing aspects of the evenness of species abundance across all species sampled (Bianchi *et al.,* 2000; Greenstreet and Hall, 1996; Greenstreet *et al.,* 1999; Heath *et al.,* 2011). Outside of this established framework, additional studies address changes in the abundance of specified suites of species (Greenstreet *et al.,* 2012b). Any change in survey design that alters the apparent relative abundance of scarce components (rare species, large size classes) compared with the more abundant components (common species, small size classes) will have an impact on these indicator values.

Scientific surveys provide fisheries independent indices of species abundance. Fisheries managers are concerned about commercial fish, which are generally the more abundant species. However, ecosystem assessments are often concerned with some of the rarer species (Dulvy *et al.,* 2003; Greenstreet *et al.,* 2012b). Metrics of species richness, for example, are confounded by survey techniques that inadequately sample rare species (Greenstreet and Piet, 2008). Previous work has examined the effects of tow duration (eg. Ehrich and Stransky, 2001), where a reduction from 60 to 30 minutes led to a slight reduction in the number of observed species. Changes in survey design that may have little impact on indices of abundance of more common species could potentially have considerable and adverse consequences for abundance metrics in rarer species (Magurran, 2014).

Two commonly accepted concepts in fisheries science are: 1) a longer tow provides a more reliable measure of species richness occurring in the habitat being sampled as they cover a larger swept area presenting a greater opportunity to resolve rarer species, and 2) large fish that are stronger swimmers are more efficiently captured (Wardle, 1986). Conversely the “catch-by-surprise” hypothesis held by Godø *et al.* (1990), suggests that the catch per unit effort of herding fish species may decrease with increased tow duration. Nevertheless, the number of individuals caught before and after the official duration (end effect) increases in shorter tows, which can primarily affect abundant species with a higher degree of mobility (Battaglia *et al.,* 2006).

Fisheries survey data are highly variable, effects on catch rates may be associated with fish reaction to the survey gear. Reactions to gear are partially determined by their distribution in the water column, size/shape, behaviour, or the degree of association to the seabed (Engås and Godø, 1986; Aglen, 1996; Godø, 1990; Fréon *et al*., 1993; Adlerstein and Ehrich, 2002). The catchability of different species depends on many factors, including fish behaviour in relation to the gear type (otter trawl or beam trawl), herding efficiency, and the probability of escape at the entrance to the net (Wardle, 1993; Engås, 1994). For some species, catch rates may vary because their behaviour changes throughout the day (Trenkel *et al*., 2008; Doray *et al*., 2010); while for other species catch rates may also vary over the duration of the tow due to spatial heterogeneity (Kingsley *et al.,* 2002) .

No significant difference in the mean length of fish caught or the catch per unit effort between 15 min and 30 min survey tows (Godø *et al.,* 1990; Walsh, 1991) was identified for cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*) and long rough dab (*Hipploglossoides platessoides*), yellowtail flounder (*Limanda ferruginea*) and thorny skate (*Amblyraja radiata*). Similarly, no significant effect was found on catch per unit effort, size composition or notably maximum length for Northern shrimp (*Pandalus borealis*) and Greenland halibut (*Reinhardtius hippoglossoides*) off West Greenland (Wieland and Storr-Paulsen, 2006). However, Somerton *et al.,* (2002) noted that the catch per swept-area increased significantly for two commercial species of crab when tow duration was decreased from 30 min to 15 min.

The overall aim of this current research is to demonstrate the impact of varying tow duration on catch composition for groundfish surveys in the North Sea, at scales relevant to fisheries management. Here we look at the fish community sampled by the gear using the longer tows (28-32 minute) and ascertain if the shorter tows (14-16 minute) are significantly different in terms of the features that they resolve. We examine the variance in abundance and biomass estimates in both long and short tow categories using linear mixed models. We also examine evenness and richness across the sample area and test the MSFD community and population level indicators.

Data and Analysis

Survey design

The Greater North Sea International Otter Trawl Survey is a coordinated survey involving England, Scotland, Norway, Germany, Sweden and Denmark in the annual sampling during quarter 3. The survey follows a systematic unaligned sampling design (Cochran, 1977), where each ICES statisitcal square is sampled at least twice. The general protocol is that each vessel fishes for 30 minutes at a speed of 4 knots using a standardised *Grande Overture* Vertical *trawl. Fish species, numbers at length and weight are some of the parameters that are recorded*. In 2015 and 2016, the IBTSWG elected to do experimental tows, where at least one of the tows in each rectangle would be 15 minutes and at least one would be 30 minutes. This resulted in an almost 50: 50 spilt in 2015, whereas in 2016 there was an increased effort to produce 15 minute tows (Figure 1).

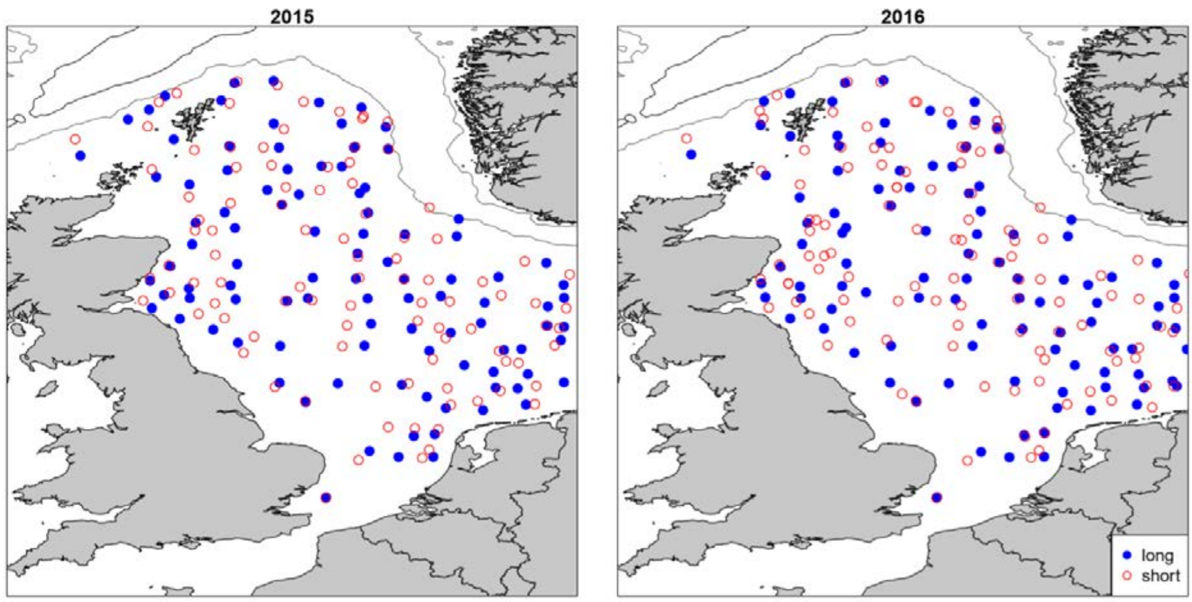

Figure 1: Stations sampled in 2015 and 2016 North Sea Q3 survey which were seleced as part of this study.

Data source

The quality assured monitoring and assessment data set for the Greater North Sea International Quarter 3 Otter Trawl Survey (Moriarty and Greenstreet, 2017), dervied from the NS-IBTS Q3 ICES DATRAS dataset, was used for this study. This is a publicly available data source with supporting technical documentation describing techniques used to quality assure the DATRAS data downloaded on 30-03-2017 (Moriarty *et al.*, 2017; Greenstreet and Moriarty, 2017a/b). The catch data for all species is expressed as recorded numbers at length, numbers per km^2 at length or biomass per km^2 at length. Table S1 highlights taxonomic corrections made to the quality assured monitoring and assessment data set, which were made in consultation with expert advice (e.g. Heessen *et al.,* 2015). The tow data includes geographical position (longitude, latitude), depth (m), tow number, vessel, statistical rectangle, tow duration (minutes), swept area (in km^2) and year.

Data selection

A subset of data was selected from the 2015 and 2016 survey data based on two criteria. The first criteria was that an ICES rectangle must have been sampled at least twice one long tow (30 ±2 minute) and one short tow (15 ±1 minute). The second criteria was that the experimental short tow must be within 30% of the depth range of the standard long tow. This was chosen arbitrarily to reduce variation caused by samples with extreme depth ranges. The depth difference in most of the tows are small, but in 16% of the tows there is a substantial difference in the depth (e.g. a tow at 17m and a tow at 73m giving a difference of 56m; Figure S1, supplementary material). The largest difference in the depth of paired tows within the same rectangle was 102m (Figure S1 in supplementary material). These two criteria produced a suite of tows covering 97 ICES rectangles (0.5 x 1°), consisting of 99 long and 99 short tows in 2015 and 103 long and 110 short tows in 2016 (Figure 1). To assess the individual tow variablity per rectangle between short and long tows the range of depth, time and the differences in speed over ground were examined (Figure S2, supplementary material).

Analyses of biodiversity

The primary aim was to demonstrate the effect of varying tow duration on species diversity, richness and evenness in the North Sea survey in Q3. The mean species richness for the long and short tows for each year was calculated to ascertain if there was a difference between the two categories.

Linear mixed effect models were used to determine the relationships between richness and tow duration, ship, speed over ground, time of tow (diel fluctuations), year, month/day, and depth. The interactions between tow duration and ship, and tow duration and year were also tested. ICES statistical rectangle was added as a random effect in the model to account for spatial auto correlation. Models were implemented using the package "lme4" (Bates *et al*., 2015) in R (R Core Team, 2017). The global linear mixed effect model had the form

Equation 1: $$\gamma = X\beta + Zu + \varepsilon$$

Where $N = 411$, and γ is a $N \times 1$ column vector of the outcome variable (e.g. richness of fish in a tow). X is a $N \times p$ matrix of the $p = 9$ fixed effects predictor variables; tow duration (long/short), ship (5 ships), speed over ground (km/hr), time of day (diel fluctuations), year, month/day (Julian days), and depth (m), the interactions between tow duration and ship, and the interactions between tow duration and year. β is $p \times 1$ column vector of fixed effect regression coefficients. Z is a $N \times q$ matrix with 1 for the corresponding random effect of ICES statistical square and 0 otherwise, $q = 97$, as we suspect that samples in the same statistical square are correlated. u is a $q \times 1$ vector of the random effects; and ε is a $N \times 1$ column vector of the residuals not explained by $X\beta + Zu$. A gaussian identity link distribution was used. Parameters were estimated by maximum likelihood. The best fitting model was determined based on Akaike's information criterion (AIC) scores. All pairwise interactions of explanatory variables were tested.

Species richness curves with bootstrapped confidence intervals were plotted against number of tows for both long and short tows categories using a randomised method. Pielou's evenness index, derived from the Shannon diversity index, was calculated for both the long and the short tows. Exploratory analysis demonstrated violations of assumptions for parametric testing, therefore, a Scheirer–Ray-Hare test (Dytham, 1999), was performed to test the hypothesis there is no difference in mean species evenness in long and short tows in each year, and no significant interaction between haul duration and year. In addition to using non-parametric test, we log transformed the data and assessed the interactions using linear models. Both approaches gave the same results, so have elected to report only the Scheirer–Ray-Hare test.

Differences in abundance and biomass

The second aim was to determine if varying tow duration affected the biomass and abundance estimates calculated from the survey data. Again, a linear mixed effects model was employed, the global model followed Equation 1, where γ was the log transformed abundance (n/km^2) and the log transformed biomass (kg/km^2) respectively. The same model parameters and model selection criteria were used.

Differences in body size

To assess differences in average body size, fish were grouped in 10 cm length classes and the log transformed mean biomass/abundance at grouped length classes by tow in each tow category and year. To assess if there were significant differences in actual body size, estimates of mean size and standard deviation per tow category per year were calculated and a Pearson's Chi-squared test with simulated p-value (based on 2000 replicates sampled with replacement from all tows) was undertaken.

To test if there was a significant difference between the short and long tows in estimates of MSFD indicators being derived from this survey the Large Fish Indicator (LFI), the Typical length (TyL) and Mean-max length (MML) were calculated for the appropriate suite of species in the samples (OSPAR, 2017). The LFI is the ratio of the average biomass (kg/km^2) of large demersal fish ($\geq 50\ cm$) per ICES statistical rectangle over the average biomass (kg/km^2) of all demersal fish sizes per ICES statistical rectangle. The LFI were calculated in both 2015 and 2016 to test for a significant difference in the estimated LFI for long and for short tows. TyL is the geometric mean length where length is weighted by biomass; MML is the arithmetic average of the maximum length obtained by species in the survey weighted by biomass; the species were split into two groups, "pelagic" and "demersal" species for the two indices. It should be noted, however, that all of these results are based on comparisons of relative abundance, relative biomass and relative mean length; therefore, accuracy cannot be evaluated as true values are unknown.

Results

Biodiversity

The best fitting linear mixed model for estimation of the mean species richness in the tow suggested that tow duration was a significant factor in describing the variability seen in the data. Other fixed effect variables that were important in describing the mean species richness were time of day, the effect of ship, the depth (m) and the year (Table S3, Figure S3). There was no significant interaction between any fixed effects, and they were therefore not included in the final model (Table 1).

Table 1: Summary of explanatory variables included (✓) or excluded (×) *in the best fitting linear mixed models for estimating the factors that explain the variance in species richness, abundance* (n/km^2) *and biomass* (kg/km^2) *in tows.*

Fixed Effects	Richness (no spp)	Abundance (n/km^2)	Biomass (kg/km^2)
Tow Duration	✓	✓	✓
Ship	✓	✓	✓
Tow Duration : Ship	×	×	×
Year	✓	×	✓
Tow Duration : Year	×	×	×
Month/Day (Julian days)	✓	✓	×
Time of Day (diel fluctuation)	✓	✓	✓
Depth (m)	✓	×	×
Speed Over Ground (km/hr)	×	×	×
Random Effect			
1 | ICES Statistical Square	✓	✓	✓

Sixteen species were uniquely present in long tows, but not short tows. These include some pelagic fish like *Belone belone* (Garfish) *Sarda sarda* (Atlantic bonito) and *Scomber colias* (chub mackerel), flatfish like *Phrynorhombus norvegicus* (Norwegian topknot) and *Zeugopterus punctatus* (topknot) and rays such as *Leucoraja fullonica* (Shagreen ray). Conversely, seven species were collected in short tows but not in long tows these included elasmobranchs such as *Etmopterus spinax* (lantern shark), *Mustelus spp.* (smooth-hound) and *Raja brachyura* (blond ray). Sharks and rays such as *L. fullonica* (Shagreen ray) and *R. brachyura* (Blond ray), *E. spinax* (lanternshark), *Mustelus spp* (smooth-hounds) were not consistently sampled (see Table S1 for full list). In the area selected for analysis the mean number of species collected in the five years prior to the start of the experiment (2010-2014) was 78 species. While the long tows are consistent with previous years with 77 and 78 species encountered, as expected when looking at a similar total number of hauls, the short tows fell short of this with 71 and 73 species encountered respectively (Table 2). The increased effort to sample more diverse habitats meant the total number of species reported in 2015 and 2016 was above average in the area sampled (83 and 87 species, respectively), the increased species were predomently reported by England, who exclusively fished for 30 min and fished similar stations in both years. In 2016, within our study area, England was the only country to report *Belone belone* (Garfish), *C. maximus* (basking shark), *L. liparis liparis* (common seasnail), *L. vahlii* (Vahl's eelpout)*, P. marinus* (sea lamprey) *and S. trutta trutta* (sea trout).

All species encountered and the number of times a species occurred is listed in Table S1. In some cases, species only occurred once within the study area, 50 species occurred in less than 5% of samples. To ascertain the effect of these species which were not well sampled testing for differences in abundance and biomass was performed on the full data set and a reduced dataset that excluded the poorly sampled species. The results were not significantly different between the two data sets; therefore, the following analyses included all the species listed in Table S1. There was no significant interaction between year and tow duration, nor were any significant differences found for the year or tow duration in the Pielou's evenness.

Table 2: Summary of the mean number of species encountered per ICES rectangle in each category and year. In 2015 a total of 83 fish species were encountered while in 2016 87 species where identified. The average number of species in sampled in the period from 2010-2014 was 78.

Year	Category	Mean number of species per rectangle	Standard deviation	Total number of species encountered
2010	long	17.54	3.98	75
2011	long	18.14	4.31	78
2012	long	18.61	3.94	78
2013	long	17.68	4.89	78
2014	long	18.62	4.39	82
2015	long	16.29	3.96	77
2015	short	13.94	3.64	71
2016	long	16.87	4.06	78
2016	short	14.77	4.35	73

The difference in potential species richness within the two tow categories, showing the difference in ability to reach a species richness of 50 species is highlighted in figure 2. The long tow category was 33% more effective at sampling species richness, this suggests that a 33% increase in the number of short tows would provide a similar species richness estimate. When increased to a species richness of 75 species this gap widens and a 67% increase in the number of short tows to long tows would be needed.

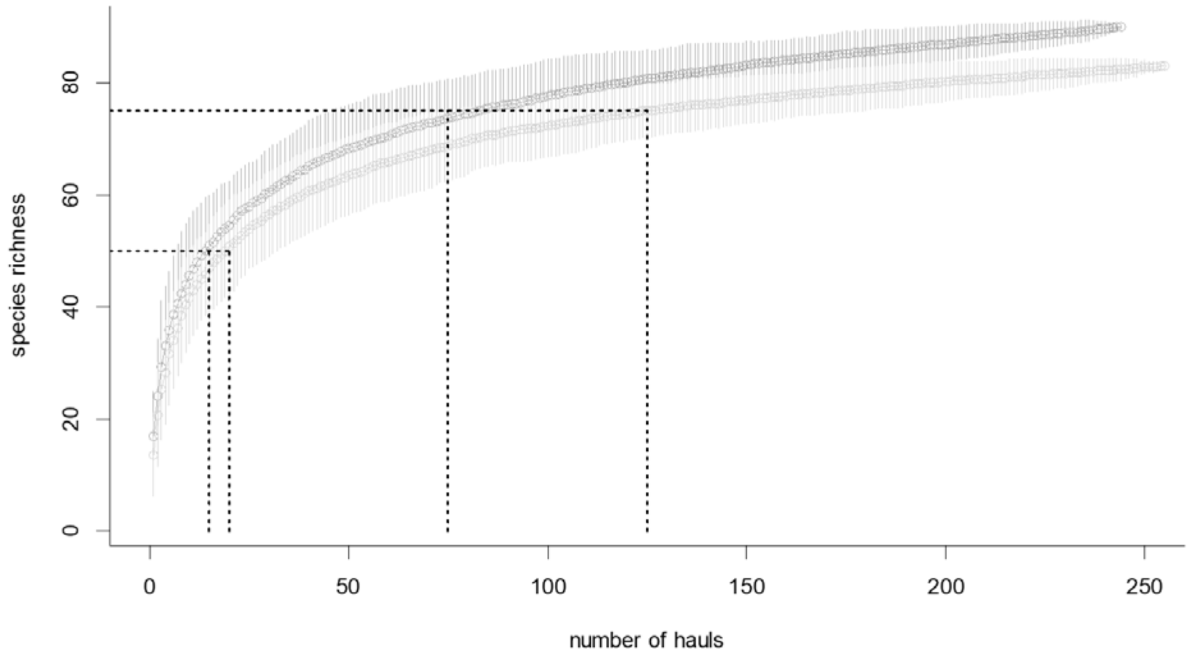

Figure 2: Cumulative species richness curves for long tows in dark grey and short tows in light grey. The black dotted lines show that to reach a species richness of 50 species, approximately 15 tows in the "long" or 20 tows in the "short" tow category are needed (33% increase in effort). Whereas to reach a species richness of 75 species, using the "long" tow category approximately 75 tows are needed and using the "short" tow category approximately 125 tows are needed (67% increase in effort). Vertical lines provide the standard deviation from random permutations of the data.

Abundance and Biomass

The best fitting linear mixed model for estimation of the mean abundance (n/km^2) in the tow suggested that tow duration is a significant factor in describing the variance seen in the data. Other fixed effect variables that were important in describing the mean abundance (n/km^2) according to the best fit model were time of day, the effect of ship, and the day of the month (Table S4, Figure S4). There was no significant pairwise interaction between any fixed effects (Table 1). Similarly tow duration was a significant factor in describing the variance seen in the mean biomass (kg/km^2), other fixed effect variables that are important in describing the mean biomass (kg/km^2) are time of day, the effect of ship, and the year (Table S5, Figure S5).

Body Size

When samples were grouped into length classes and the average biomass/abundance at size was compared, evidence for differences between the short and long tows was found for abundance. The results showed that there was no significant interaction between year and tow duration, but there were significant differences ($p < 0.05$) in the tow categories, and year was not found to be a significant factor. The log-transformed mean abundance and mean biomass calculated by summing the the log transformed mean biomass/abundance at grouped length classes are outlined in table 3. Generally short tows had a higher mean biomass and abundance at size, in particular for larger sizes (> 40 cm) than long tows when the data was standardised for swept area (km^2) (Figure 3). The > 99cm class shows there is a higher mean abundance at size for short tows, while the biomass reflects virtually no difference, this is due to several larger fish in the long tows that balanced out the more numerous smaller fish in the short tows.

The 30-39 cm class is dominated by three pelagic species Atlantic horse mackerel (*Trachurus trachurus),* herring (*Clupea harengus),* and mackerel (*Scomber scombrus*) which accounted for about 65% of the abundance in this length class in long tows and 40% of the catch in short tows. The other dominant species in this class are haddock (*M. aeglefinus*) and whiting *(M. merlangus*) which accounted for about 25% of the abundance in this length class in long tows and 37% of the catch in short tows.

Table 3: Summary of the mean across length classes of the log-transformed abundance and biomass at length in each tow duration category and year.

Year	Category	Mean log-abundance *(numbers/km^2)*	Standard deviation (Log-abundance)	Mean log-biomass *(kg/km^2)*	Standard deviation (Log- biomass)
2015	long	4.24	1.99	4.16	1.22
2015	short	4.77	1.94	4.50	1.11
2016	long	4.10	1.89	4.02	1.22
2016	short	4.58	1.69	4.47	1.12

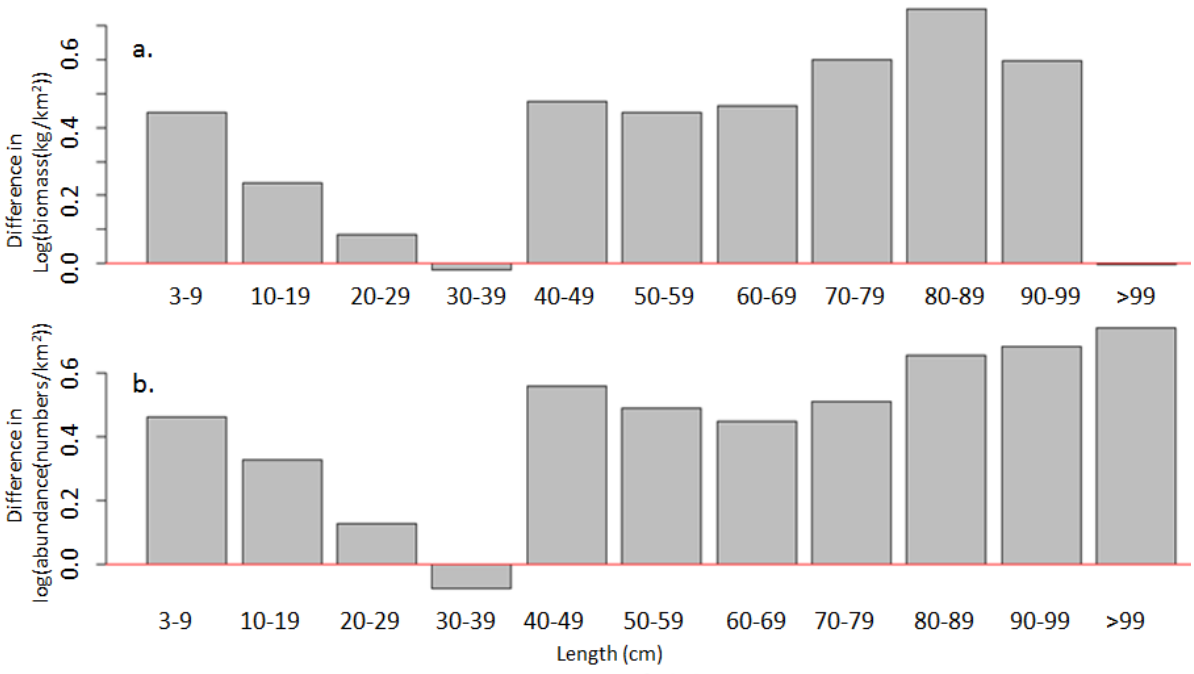

Figure 3 (a.) Bar charts showing the difference (short tow – long tow) between log-transformed mean biomass (kg/km^2) in short and long tow categories for groups of length of all fish over the two years. (b.) Showing the difference between log-transformed mean abundance (numbers/km^2) in short and long tow categories for groups of length of all fish over the two years.

Given the apparent differences in abundance at size between short and long tow categories a Pearson's chi-squared (χ^2) test was carried out to examine differences in size composition of fish in each category. A χ^2 value of 2963600 ($p < 0.001$) for 2000 bootstrapped resamples was calculated. This suggests a significant relationship between the tow duration categories for the length of fish caught. Table 4 highlights the mean size of fish caught in each category and year. The long tow category had a higher mean size than the short tow category.

Table 4: Summary of the mean size and standard deviation of fish caught in each category and year, based on number/km^2 of individuals caught in the haul.

Year	Category	Mean size of fish (cm)	Standard deviation
2015	Long	13.28	6.13

2015	Short	12.2	5.58
2016	Long	13.19	7.11
2016	Short	11.89	6.80

The LFI with a 50 cm threshold was dominated by common skate (*Dipturus batis)*, followed by cod (*Gadus morhua*), monkfish (*Lophius piscatorius)*, pollack (*Pollachius virens*), and hake (*Merluccius merluccius*). Other species which made up the community of larger fish included rays and sharks (*Squalus acanthias, Raja clavata, R. montagui, Mustelus spp.*), and commercially important species such as haddock (*M. aeglefinus*). There was no significant difference (p = 0.05) in the MSFD indicator (LFI, MML and TyL) results in the long and short tows (Figure 4).

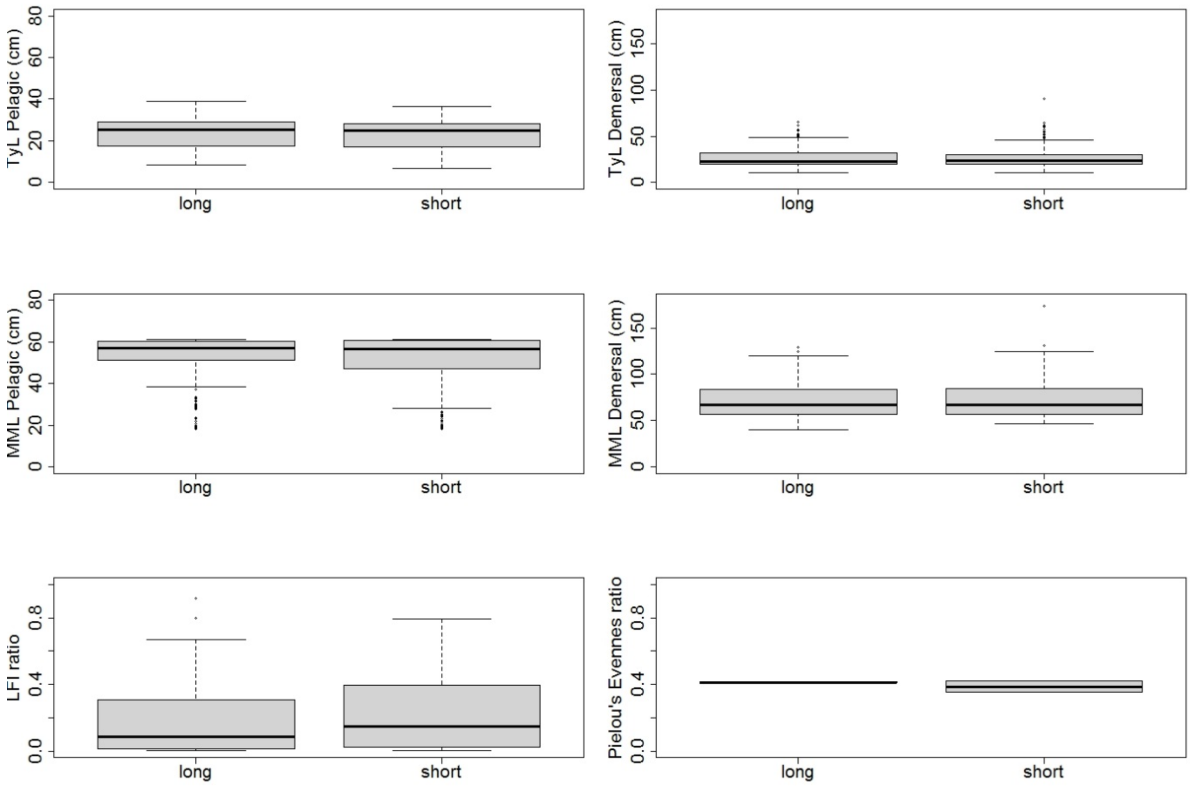

Figure 4. Box plots show the mean and the variance for TyL, MML LFI and Pielou's evenness metrics in both long and short tow categories. There was no significant difference between long and short tow categories in any of these metrics.

Discussion

This study demonstrates that tow duration , depth (m), diel fluctuation, ship, and year, are significant variables in predicting number of species in a haul (Table 1; Table S3; Figure S3). The long tow category reached a higher species diversity than the shorter tows (Table 2). The individual species recorded did differ in each category in each year, with rarer species not consistently present across years. The culumative frequency curves in figure 2 highlight the disparity between the long and short tows. Subsampling the 30 min tows has been found to limit the species richness in a similar manner to the shortened tow duration (Ehrich and Stransky, 2001), so understanding to what degree are stations subsampled in the current 30 min tow regime is important. If one country consistently subsamples stations it may lead to a bias in species richness over time in that area. On the balance of evidence presented, in terms of the species richness estimates staying with the longer 30 minute tows is capturing additional information, and will give better estimates of species richness (this is a function of the larger net swept area in the longer tows). However, this experiment only addresses the consequences of a reduction in fishing effort, as the data available generally has one short tow for each long tow in each rectangle sampled. Therefore, in order to see how this applies more broadly we would need to examine paired tows that account other factors that we cannot adequately test here, such as vessel and crew effect.

With a longer tow and a greater net swept area the probability of encountering a rare species increases, the argument stands that using a shorter tow duration and at the same time an increased number of stations might allow additional habitats to be sampled and could thereby increase the probability of encountering rare species (Pennington and Vølstad, 1991). However, this only holds if the time gained permits to carrying out additional tows (including steaming time) in a given day, and the survey design moves from rectangle to habitat stratification. The increase in species richness through sampling a variety of habitats in the North Sea is demonstrated in Wieland (2017a). Species richness varied at different depths, being significantly higher at deeper stations, due to an increase in numbers of sharks and rays (Wieland, 2017a). This is reflected in our linear mixed model investigating species richness, depth was a significant variable in predicting number of species in a haul (Table S3). A key aspect of reducing tow duration is to adapt the current sampling regime and sample more habitats within the North Sea. In 2016 a greater variety of habitats were sampled, Wieland, (2017a) found that there was an increase in biomass at depth. Depth of the tow will affect the community composition and the performance of the gear, so rectangles with paired tows that have a large difference in depth may not be directly comparable; this has been addressed by limiting the depth band of tows within rectangles for the standard survey area. Thus, by reducing the tow time to 15 min and freeing up time to sample more habitats at different depths the survey may in fact become impaired in its primary goal of detecting trends in abundance and biomass in the fish community.

The general picture in the abundance in the short tow category is higher than in the long tow, with the exception of fish in the 30-39 cm class and the >100cm class (Figure 3). The same picture is seen in the biomass estimates. The short tow category had a significantly higher logged mean abundance at length than the long tow category. The investigation of arithmetic mean body size class by abundance/biomass suggested a significant difference in the mean body size caught in long and short tows, with long tows catching slightly larger fish on average than the short tows. However, the geometric mean length weighted by biomass (TyL) was robust to this influence. Similarly, the species composition metric (MML) was also robust to the change in tow duration.

The effective sample sizes for estimating population characteristics (e.g. age) are typically low for the IBTS surveys, around one fish at length, on average, per tow, this implies that there may be little to gain by increasing tow duration beyond 15 min for estimating population characteristics. Devine and Pennington (2017) suggest that for the IBTS survey, 15 min tows are more efficient for estimating catch per unit effort series than 30 min tows. In addition, other studies on the North Sea Q3 experimental tow data have examined the effect on catch rates by ages for individual species such as cod (*G. morhua*), and whiting (*Merlangius merlangus)* (Wieland, 2017b), haddock (*M. aeglefinus*) and

Norway pout (*Trisopterus esmarkii*) (Jaworski *et al.,* 2017). There was no clear indication that the experimental 15 min tows were any less representative than the standard 30 min for catch rates at age of these four species (Jaworski *et al.,* 2017; Wieland, 2017b).

Attributing the variation in species richness, abundance, biomass and body size in a tow to just one factor, duration of a tow, is not always possible, as the survey data is highly variable the community structure varies in space and time, and the North Sea environment is heterogeneous. Efforts have been made to standardize protocols in the North Sea surveys, by fixing tow duration, vessel speed and standardizing the gear. However, in practice, tow duration varies, for example if a very large pelagic shoal is detected on the sonar then a chief scientist may decide to tow early to protect the nets.

Vessel speed, also known as speed over ground, is difficult to regulate as this is only one measure of speed, without a clear measure of speed in water it is difficult to ascertain how the variation of vessel speed will affect the catch composition. Figure S2 shows how each vessel performed at 30 min tows and 15 min tows respectively. In some cases, the vessels deviated from the expected 4 knots (speed over ground). The protocol set out for this survey, to maintain a constant speed of 4 knots through water and over ground is impracticable. A departure from target speed has been found to affect catch rates of target species in previous studies (Adlerstein and Ehrich, 2002; Koeller, 1991; Main and Sangster, 1981; Neproshin, 1979; Ona and Chruickshank, 1986; Olsen *et al*., 1982; Olsen, 1990; Ona and Godø, 1990). In our linear mixed models speed was not a significant factor in describing variance in richness, abundance or biomass estimates (Table 1), which is not surprising, since vessels operate around a given target speed (Figure S2). It is noted that the "standard" gear as described in the survey manual is not used by any participating nation (ICES, 2015b).

Time of day plays a part in variation of catch rates for some species, this is reflected in our linear mixed models, where time of day is a significant factor in all three models. In this study the time of day varies for paired tows in a given rectangle, in some cases the tows occur as close as 2 mins apart, in other cases the range is much higher, for example 2.45 am for one tow and 6.55 pm for the second. Catches of several species are known to fluctuate with time of day, (Adlerstein and Trumble, 1993; Adlerstein and Ehrich, 2002; Ehrich and Gröger, 1989; Pitt *et al.*, 1981; Wieland *et al.*, 1998), so paired tows should be performed as close together as possible to limit bias. Depth of the tow will affect the community composition and the performance of the gear, so rectangles with paired tows that have a large difference in depth may not be directly comparable. Estimates of wing swept area are also imprecise. These mechanical parameters alongside fish behaviours lead to uncertainty in estimates of fish abundance (numbers per km^2) and biomass (kg/km^2).

These considerations may compromise the ability to assess the differences in one factor, as best practice would be to control all other variables. Given the time and financial constraints on participating nations in the current economic climate, it would not be practical to perform such an experiment on this scale. As these experimental tows are not truly paired tows, i.e. two vessels towing side by side, at the same speed and at the same depth, there is a high amount of additional variation. However, when paired experiments have been carried out, the results still showed a large variability between tows carried out in close proximity at the same time (Doray *et al.,* 2010). This makes it very difficult to draw any significant results from any tests performed. As a result of this variation we have elected to look at the average changes over the whole study area, to ascertain if a signal is present that suggests a consistent bias based on tow duration.

Optimisation of survey resources while managing the needs and expectations of the end users is an issue that affects many nations. In this case the dicussion that has been initiated on optimising the survey design will require big picture thinking. This experiment, addressing one factor, tow duration, must be set in the context of the wider discussion which considers all the potential future changes, such as a new fishing gear, that will be required to maintain this survey, and other similar surveys into the future. Fisheries survey data are highly variable and disentangling within survey variation and understanding how this affects individual samples is a difficult task. By changing a key factor in the survey design there is a risk of undermining the primary goal of the survey. Such a change must be decided on balance of the potential gains for example, reducing tow duration may increase precision of a survey by allowing time to collect more samples. The average number of stations sampled by the full survey from 2011-2014 was 323 (Table S2 in supplementary material), assuming reducing tow duration to 15 mins would allow one additional tow per day for each vessel, this increases the total number of stations sampled to 424 stations, representing a 31% increase in the number of stations sampled. If each nation could carry out 1.5 extra tows per day then there would be a 47% increase in stations sampled, however this is unlikely given the distance between stations. Based on projections using a semi-Gleason fit on the species accumulation curves, a 31% increase in short hauls may provide a similar amount of species richness information as the current survey design.

A major concern when looking at historic surveys with longer time series is disrupting the time series and therefore losing long term information. In this particular case, the is another survey conducted in Q1 which largely samples the same community (with the exception of a few migratory species) and over a much longer time period therefore the histoical information for this community may still be maintained despite change to the Q3 survey. There are many practical benefits to implementation of a reduction of tow duration within the North Sea Q3 survey such as less wear and tear on gear; increased coverage of habitats; a reduction in subsampling of large tows; and a potential reduction in animal mortality. Reducing the impact of marine surveying is important and a reduction in tow duration may be part of the solution. However, if there is a substantive increase in number of tows carried out, the displacement in effort may impact on more habitats. The results presented illustrate the potential losses involved as it supports the assertion that a reduction in tow duration, given the current survey design, would have a negative impact on the capacity to resolve species richness, and may also affect the main survey objectives to supply data to the assessment working groups to fine-tune North Sea regional calculations of estimates of species abundance and biomass in support of the first quarter assessments. Before any longterm changes are made to a surveys design it is imperative that a broader strategy on survey modernisation and impact reduction is discussed and agreed upon.

Acknowledgements

This contribution was initiated by the work carried out by the ICES IBTSWG. All who took part in the working group are thanked for their contributions.

Funding

S.P.R.G and M.M were supported under the Scottish Government's schedual of service ST02h/20451. F.B and E.D.C were supported under the Scottish Government's schedual of service SU02c0/20241. C.P.L acknowledges support from Defra funding for Stock Monitoring Charter and RV Surveys (MD003).

References

Adlerstein, S. A., and Ehrich, S. 2002. Effect of deviations from target speed and of time of day on catch rates of some abundant species under North Sea International Bottom Trawl Survey protocol conditions. – ICES Journal of Marine Science, 59: 594–603.

Adlerstein, S. A., and Trumble, R. J. 1993. Management implications of changes in by-catch rates of Pacific halibut and crab species caused by diel behaviour of groundfish in the Bering Sea. ICES Marine Science Symposia, 196: 211– 215.

Aglen, A. 1996. Impact of fish distribution and species composition on the relationship between acoustic and swept area estimates of fish density. ICES Journal of Marine Science, 53: 501–505.

Bates, B., Maechler, M., Bolker, B., Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Battaglia, A., Trenkel, V. M., and Rochet, M.-J. 2006. Estimating end effects in trawl catches. ICES Journal of Marine Science, 63: 956-959.

Bianchi, G., Gislason, H., Graham, K., Hill, L., Jin, X., Koranteng, K., Manickchand-Heileman, S., Payá, I., Sainsbury, K., Sanchez, F., and Zwanenburg K. 2000. Impact of fishing on size composition and diversity of demersal fish communities, *ICES Journal of Marine Science*, Volume 57:3, pp 558–571, https://doi.org/10.1006/jmsc.2000.0727

Cochran, W.G. (1977) Sampling Techniques. John Wiley & Sons, New York.

Devine, J., and Pennington M., 2017. The 2016 North Sea IBTS Q3 tow duration experiment. Working Document in ICES. 2017. Interim Report of the International Bottom Trawl Survey Working Group. IBTSWG Report 2017 27-31 March 2017. ICES CM 2017/SSGIEOM:01. 337 pp 285.

Doray, M., Mahévas, S., and Trenkel, V. M. 2010. Estimation of gear efficiency in a combined acoustic-trawl survey, with reference to demersal fish spatial distribution. ICES Journal of Marine Science, 67: 668-676.

Dulvy, N.K., Sadovy, Y. and Reynolds, J.D. 2003. Extinction vulnerability in marine populations. Fish and Fisheries 4, 25–64.

Dytham, C. 1999. Choosing and using statistics: a biologist's guide. Blackwell Science, Oxford, United Kingdom.

EC. 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008; establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive), Official Journal of the European Union, L 164/19

EC. 2010. Commission Decision of 1 September 2010 on criteria and methodological standards on good environmental status of marine waters (notified under document C(2010) 5956)(2010/477/EU)

Ehrich, S., and Gröger, J. 1989. Diurnal variation in catchability of several fish species in the North Sea. ICES CM 1989/B: 35.

Ehrich, S., and Stransky, C. 2001. The influence of towing time, catch size and catch treatment on species diversity estimates from groundfish surveys. Archive of fishery and marine research, 49:37-44

Engås, A. 1994. The effects of trawl performance and fish behaviour on the catching efficiency of demersal sampling trawls. In Marine fish behaviour in capture and abundance estimation, pp. 45-68. Ed. by A. Fernö, and S. Olsen. Fishing News Books, Oxford.

Engås, A., and Godø, O. R. 1986. Influence of trawl geometry and vertical distribution of fish on sampling with bottom trawl. Journal of Northwest Atlantic Fisheries Science, 7: 35–42.

Fréon, P., Gerlotto, F., and Misund, A. 1993. Consequences of fish behaviour for stock assessment. ICES Marine Science Symposia, 196: 190–195.

Gislason, H., and Rice, J. 1998. Modelling the response of size and diversity spectra of fish assemblages to changes in exploitation. ICES Journal of Marine Science. 55.362-370

Godø, O. R. 1990. Factors affecting accuracy and precision in abundance indices estimates of gadoids from scientific surveys. Thesis, University of Bergen, Norway. 169 pp.

Godø, O. R., Pennington, M., and Vølstad, J. H. 1990. Effect of tow duration on length composition of trawl catches. Fisheries Research, 9: 165–179.

Greenstreet, S.P.R., and Hall S.J., 1996. Fishing and the groundfish assemblage structure in the north-western North Sea: an analysis of long-term and spatial trends. Journal of Animal Ecology Vol 65:577-593

Greenstreet, S. P. R., Fraser, H. M., Rogers, S. I., Trenkel, V. M., Simpson, S. D., and Pinnegar, J. K. 2012b. Redundancy in metrics describing the composition, structure, and functioning of the North Sea demersal fish community. ICES Journal of Marine Science, 69: 8–22.

Greenstreet, S. P. R., Rogers, S. I., Rice, J. C., Piet, G. J., Guirey, E. J., Fraser, H. M., and Fryer, R. J. 2012a. A reassessment of trends in the North Sea Large Fish Indicator and a re-evaluation of earlier conclusions. ICES Journal of Marine Science, 69: 343–345.

Greenstreet, S. P. R., Rogers, S. I., Rice, J. C., Piet, G. J., Guirey, E. J., Fraser, H. M., and Fryer, R. J. 2011. Development of the EcoQO for the North Sea fish community. – ICES Journal of Marine Science, 68: 1–11.

Greenstreet, S.P.R and Piet, G.J., 2008. Assessing the sampling effort required to estimate α species diversity in the groundfish assemblages of the North Sea. Marine Ecology Progress Series. Vol. 364: 181-197, doi:10.3354/meps07499.

Greenstreet, S.P.R., and Moriarty, M. 2017a. OSPAR Interim Assessment 2107 Fish Indicator Data Manual (Relating to Version 2 of the Groundfish Survey Monitoring and Assessment Data Product). Scottish Marine and Freshwater Science Vol 8 No 17, 83pp. DOI: 10.7489/1985-1

Greenstreet, S.P.R., and Moriarty, M. 2017b. Manual for Version 3 of the Groundfish Survey Monitoring and Assessment Data Product. Scottish Marine and Freshwater Science Vol 8 No 18, 77pp. DOI: 10.7489/1986-1

Greenstreet, S.P.R., and Rogers, S.I., 2006. Indicators of the health of the fish community of the North Sea: identifying reference levels for an ecosystem approach to management. ICES Journal of marine Science Vol 63: 573-593

Greenstreet, S.P.R., Spence F.E., and McMillan, J.A,. 1999. Fishing effects in northeast Atlantic shelf seas: patterns in fishing effort, diversity and community structure. V. Changes in structure of the North Sea groundfish assemblage between 1925 and 1996. Fisheries Research Vol 40: 153-183.

Heath, M. R., and Speirs, D. C. 2011. Changes in species diversity and size composition in the Firth of Clyde demersal fish community (1927–2009). Proceedings of the Royal Society of London B: Biological Sciences, rspb20111015.

Heesen, H.J.L, Dann, N., Ellis, J.R. (2015) Fish Atlas of the Celtic Sea, North Sea and Baltic Sea: Based on International Research Vessel Data

ICES, 2012. *Manual* for the International Bottom Trawl *Surveys*. Series of ICES. *Survey* Protocols. SISP 1-*IBTS* VIII.

ICES. 2015a Manual for the International Bottom Trawl Surveys. Series of ICES Survey 479 Protocols SISP 10 - IBTS IX. 86 pp. 480

ICES, 2015b. Report of the International Bottom Trawl Survey Working Group (IBTSWG), 23-27 481 March 2015, Bergen, Norway. ICES CM 2015/SSGIEOM:24. 278 pp.

Jaworski, A., Clarke, L., and Burns, F., 2017. Data Analysis For Haddock and Norway Pout From Tow Duration Experiment NS-IBTS 3rd Quarter 2015–2016.Working Document in ICES. 2017. Interim Report of the International Bottom Trawl Survey Working Group. IBTSWG Report 2017 27-31 March 2017. ICES CM 2017/SSGIEOM:01. 337 pp

Jennings, S. 2005. Indicators to support an ecosystem approach to fisheries. Fish and Fisheries, 6: 212–232. doi:10.1111/j.1467-2979.2005.00189.x.

Kingsley, M. C. S., Carlsson, D. M., Kanneworff, P., and Pennington, M. 2002. Spatial structure of the resource of Pandalus borealis and some implications for trawl survey. Fisheries Research, 58: 171–183.

Koeller, P. A. 1991. Approaches to improving groundfish survey abundance estimates by controlling the variability of survey gear geometry and performance. Journal of Northwest Atlantic Fisheries Science, 11: 51–58.

Magurran, A.E., 2014. Measuring Biological Diversity, Hoboken, NJ: Wiley-Blackwell.

Main, J., and Sangster, G. I. 1981. A study of fish capture process in a bottom trawl by direct observations from a towed underwater vehicle. Scotland Fisheries Research Report, 23: 1–23.

Modica, L., Velasco, F., Preciado, I., Soto, M., and Greenstreet, S. P. R. 2014. Development of the large fish indicator and associated target for a Northeast Atlantic fish community. – ICES Journal of Marine Science, 71: 2403–2415.

Moriarty, M., and Greenstreet, S.P.R. 2017. Greater North Sea International Otter Trawl Quarter 3 Groundfish Survey Monitoring and Assessment Data Products. doi: 10.7489/1923-1

Moriarty, M., Greenstreet, S.P.R. and Rasmussen, J. 2017. Derivation of Groundfish Survey Monitoring and Assessment Data Product for the Northeast Atlantic Area. Scottish Marine and Freshwater Science Vol 8 no 16, 240pp. DOI: 10.7489/1984-1

Neproshin, A. Y. 1979. Behaviour of the Pacific mackerel, Pneumatophorus japonicus, when affected by vessel noise. Journal of Ichthyology, 18(4): 695–699.

Olsen, K. 1990. Fish behaviour and acoustic sampling. Rapports et Procès-verbaux des Réunions du Conseil International pour l'Exploration de la Mer, 189: 159–166.

Olsen, K., Angell, J., and Pettersen, F. 1982. Observed fish reaction to a surveying vessel with special reference to herring, cod, capelin and polar cod. FAO Fisheries Report, 300: 139–149.

Ona, E., and Chruickshank, O. 1986. Haddock avoidance reactions during trawling. ICES CM 1986/B: 36, 13pp.

Ona, E., and Godø, O. R. 1990. Fish reactions to trawling noise: the significance for trawl sampling. Rapports et Procès-verbaux des Réunions du Conseil International pour l'exploration de la Mer, 189: 159–166.

OSPAR., 2017., OSPAR Assessment Portal <https://oap.ospar.org/en/ospar-assessments/intermediate-assessment-2017/> accessed 21/12/2017

Pennington, M., and Vølstad, J. H. 1991. Optimum size of sampling unit for estimating the density of marine populations. Biometrics, 47: 717–723.

Pitt, T. K., Wells, R., and McKone, W. D. 1981. A critique of research vessel otter trawl surveys by the St. Johns research and resource services. In Bottom trawl surveys, pp. 42–81. Ed. by W. G. Doubleday, and D. Rivard. Special Publication of Canadian Journal of Fisheries and Aquatic Science, 58.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shannon, L. J., Coll, M., Yemane, D., Jouffre, D., Neira, S., Bertrand, A., and Diaz, E. 2010. Comparing data-based indicators across upwelling and comparable systems for communicating ecosystem states and trends. ICES Journal of Marine Science, 67: 807–832.

Shephard, S., Reid, D. G., and Greenstreet, S. P. R., 2011. Interpreting the large fish indicator for the Celtic Sea. – ICES Journal of Marine Science, doi:10.1093/icesjms/fsr144.

Shin, Y-J., Shannon, L. J., Bundy, A., Coll, M., Aydin, K., Bez, N., Blanchard, J. L., Borges, M. F., Diallo, I., Diaz, E., Heymans, J. J., Hill, L., Johannesen, E., Jouffre, D., Kifani, S., Labrosse, P., Link, J. S., Mackinson, S., Masski, H., Mo¨llmann, C., Neira, S., Ojaveer, H., ould Mohammed Abdallahi, K., Perry, I., Thiao, D., Yemane, D., and Cury, P. M. 2010. Using indicators for evaluating, comparing, and communicating the ecological status of exploited marine ecosystems. 2. Setting the scene. – ICES Journal of Marine Science, 67: 692–716.

Somerton, D. A., Otto, R. S., and Syrjala, S. E. 2002. Can changes in tow duration on bottom trawl surveys lead to changes in CPUE and mean size? Fisheries Research, 55: 63–70.

Trenkel, V. M., and Rochet, M.-J. 2003. Performance of indicators derived from abundance estimates for detecting the impact of fishing on a fish community. Canadian Journal of Fisheries and Aquatic Sciences, 60: 67-85.

Trenkel, V. M., Rochet, M.-J., and Mahévas, S. 2008. Interactions between fishing strategies of Nephrops trawlers in the Bay of Biscay and Norway lobster diel activity patterns. Fisheries Management and Ecology, 15: 11-18.

Walsh, S. J. 1991. Effect of tow duration on gear selectivity. NAFO SCR Doc., 91/84: 9 pp.

Wardle, C. S. 1993. Fish behaviour and fishing gear. In Behaviour of teleost fishes, 2nd edn, pp. 609-643. Ed. by T. J. Pitcher. Chapman & Hall.

Wardle, C.S. 1986. Fish behaviour and fishing gear. In: The Behaviour of Teleost Fishes. Ed. By T. J. Pitcheer. Croom Helm, London, pp. 463–495.

Wieland, K. 2017a. Within-rectangle variation of species composition and catch rates – Implications for survey design, station allocation and tow duration. Working Document in ICES. 2017. Interim Report of the International Bottom Trawl Survey Working Group. IBTSWG Report 2017 27-31 March 2017. ICES CM 2017/SSGIEOM:01. 337 pp

Wieland, K. 2017b Effect of tow duration on catch rates of whiting and cod in the NS-IBTS 3Q 2015 and 2016. Working Document in ICES. 2017. Interim Report of the International Bottom Trawl Survey Working Group. IBTSWG Report 2017 27-31 March 2017. ICES CM 2017/SSGIEOM:01. 337 pp

Wieland, K., and Storr-Paulsen, M. 2006. Effect of tow duration on catch rate and size composition of Northern shrimp (Pandalus borealis) and Greenland halibut (Reinhardtius hippoglossoides) in the West Greenland Bottom Trawl Survey. Fisheries Research, 78: 276–285.

Wieland, K., Fosdager, L., Holst, R., and Jarre-Teichmann, A. 1998. Spatial distribution of estimates of juvenile (age 1 and 2) whiting and cod in the North Sea. ICES CM 1998/J: 7.