An analysis of the North Sea International Bottom Trawl Survey Data

Abstract

In this research we present non-parametric estimation procedures for calculating abundance at age indices, and investigate the sensitivity of these estimates with respect to the number of otholits collected at sea. The procedures presented are applied to the North Sea International Bottom Trawls Survey data for cod (*Gadus morhua*) and saithe (*Pollachius virens*). We demonstrate how much information would be lost if the survey design was defined such that fewer otholits were collected. Age length keys (ALKs) are used to map lengths to age, and we use ALKs with and without the assumption of constant age length structures over relatively large areas. All abundance at age indices are presented with variance estimates.

1 Introduction

Fish stock assessments are used by fishery managers for making management decisions regarding catch quotas. The assessments provide fundamental information about the status of the stock, for instance, whether the stock is increasing and support for increased levels of harvest should be given, or whether the stock is decreasing and stricter control on harvest should be implemented. Associated with the parameters used in fish stock assessment is their uncertainty, which should not be ignored when formulating management policies (Walters and Ludwig, 1981; Ludwig and Walters, 1981; Berg et al., 2014). This uncertainty can arise from many sources including natural variability, estimation procedures and lack of knowledge regarding the parameter (Ehrhardt and Legault, 1997). The North Sea International Bottom Trawl Survey (IBTS) data, coordinated by the International Council for the Exploration of the Sea (ICES), provides information on seasonal distribution of stocks and estimates of abundance indices and catch in numbers of fish per age-class without an assessment of the accuracy of these estimates. As stated by Ludwig and Walters (1981) it is relevant for managers to take into the uncertainty related to stock size when making management polices. The indices of abundance at age from IBTS are based on data obtained from a stratified semi-random sampling approach of trawl stations, and it is essential to account for the sampling approach so as to produce reliable variance estimates (Lehtonen and Pahkinen, 2004). If the sampling approach is ignored, the effect on the variance of the parameters could be substantial. In particular, the variance could be greatly inflated due to the clustering effect, which involves intra-cluster correlation of the variables (Aanes and Vølstad, 2015; Lehtonen and Pahkinen, 2004).

There are two separate stages of the North Sea International Bottom Trawl Survey (IBTS) for generating abundance indices per age. The first consist of calculating indices per *length* class, which are obtained by trawling in a stratified manner and counting the number of fish caught. Then that knowledge is transformed to indices with respect to age. The latter part is achieved with an age-length key (ALK), which is constructed by sampling otoliths in a stratified procedure from each haul and/or sub-area. To our best knowledge, there has been no research on how much the uncertainty of the abundance indices is related to these two distinct parts. The main contribution of this article is to shed light on how the indices estimates and their associated uncertainty estimates change if less effort was spent on collection of otoliths. We achieve the reduction of otoliths by mimicking a defined sampling procedure with less effort. We also focus on the spatial distribution of the ALK, and such spatial structures in the ALK has also been investigated in Berg and Kristensen (2012); Hirst et al. (2012).

Currently, abundance indices from IBTS are reported in DATRAS (ICES, 2018c) using an age-length key (ALK) (Fridriksson, 1934) which is assumed to be constant over relatively large areas. In this research we propose two ALKs which accounts for spatial variation: i) a nonparametric haul based ALK, and ii) a spatial model-based ALK. These ALKs are described in section 2, and the results from the model based ALK gives a strong case for assuming variation in the ALK within RFAs. A spatial model based ALK (Berg and Kristensen, 2012; Berg et al., 2014) known as the NS-IBTS Delta-GAM index (ICES, 2016b) is currently being used to calculate standardized age-based survey indices used in assessment for the North Sea stock. And as far as we are aware the variance estimates of parameters estimated from NS-IBTS Delta-GAM index are *only* utilized for assessment of Herring (*Clupea harengus*) in the North Sea.

The spatial ALK model introduced in Berg and Kristensen (2012) is similar to the model used in this paper; the main difference is that we include the spatial structure through a spatial random field (Lindgren et al., 2011) and not through two dimensional splines (Wood, 2017).An overview of the North Sea International Bottom Trawl Survey is given in Section 1.1. The current estimators for ALK and catch per unit effort (CPUE) used by ICES in their database for trawl surveys (DATRAS) and our proposed ALK estimators are given in Section 2. Two case studies, in which the methods described in Section 2 are applied to, are given in Section 3, and a discussion is given in Section 4.

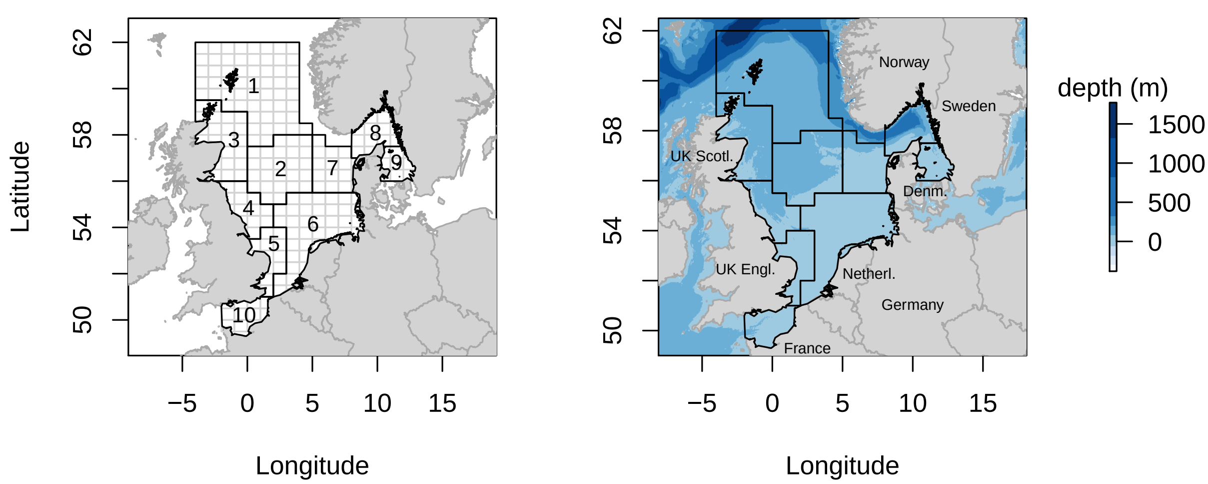
1.1 Overview of the North Sea International Bottom Trawl Survey

The North Sea International Bottom Trawl Survey was formed in 1991, which is a combination of the International Young Herring Survey (IYHS) and eight national surveys in the North Sea, Skagerrak and Kattegat areas. These surveys began in the 1960's, and the 1970's and 1980's, respectively. The IYHS was developed with the aim of obtaining annual recruitment indices for the combined North Sea herring *Clupea harengus* stock (ICES, 2012), but yielded valuable information on other fish species such as cod *Gadus morhua* and haddock *Melanogrammus aeglefinus*.

The North Sea IBTS began with quarterly surveys providing information on seasonal distribution of stocks sampled, hydrography and the environment, which allows changes in fish stock to be monitored and abundance of all fish species to be determined. These quarterly surveys, however became difficult to sustain as countries experienced budget cuts making it impossible to maintain high levels of research vessel effort. As such, in 1997 countries carried out a survey only twice a year; a first quarter survey (January-February) and a third quarter survey (August-September). The target species of IBTS fished from 1991-2018 includes standard pelagic species: Herring (Clupea harengus), Sprat (Sprattus sprattus) and Mackerel (Scomber scombrus); and standard roundfish species: Cod (Gadus morhua), Haddock (Melanogrammus aeglefinus), Saithe (Pollachius virens), Norway Pout (Trisopterus esmarkii) and Whiting (Merlangius merlangus).

Research vessels from seven nations in the first quarter (Q1) and six nations in the third quarter (Q3) are used for conducting surveys on all finfish species in the North Sea during January-February and July-August, respectively, between 1997-2018 (Table 2 in Web appendix A gives details of the different research vessels). The sampling frame is defined by the ICES index or roundfish areas (RFA) as shown in Figure 1 numbered 1 to 10, and which we refer to as superstrata (Nottestad et al., 2015; Fuller, 2011). These roundfish areas were substratified into small strata defined by non-overlapping statistical rectangles of roughly 30×30 nautical miles (1^o Longitude $\times$ 0.5^o Latitude), and were convenient to use for North Sea IBTS as they were already being used for fisheries management purposes. Most statistical rectangles contain a number of possible tows that are deemed free of obstructions, and vessels are free to choose any position in the rectangles as long as the hauls are separated by at least 10 nautical miles within and between rectangles. However, all countries select tows based on a semi-random approach from datababes of national safe tows or DATRAS or commercial fishing data, except Sweden who uses fixed stations and in some cases depth-stratified semi-random sampling design (ICES, 2018b), and England who also uses fixed stations and only conduct surveys during the third quarter. In some rectangles, sampling may be further stratified due to significant changes in seabed depth which may, in turn, cause variations in the fish population. In particular, the North Sea IBTS herring, saithe and sprat data are weighted by depth strata in the statistical rectangle (see Table 4 in appendix C). It is also a requirement that countries avoid clustering their stations between adjacent rectangles in order to reduce positive serial correlation, and thereby maximize survey precision. The latest major reallocation of rectangles occurred in 1991, but since then the survey has tried to keep at least one vessel in every subarea in which it had fished in the most recent years. Minor reallocation of rectangles between Norway, Scotland and Germany was done in 2013. Each rectangle was typically sampled twice by two different countries before 1997, but after that target coverage of two trawl hauls per rectangle per survey was introduced because of national financial constraints (ICES, 2015). But in some rectangles in the Eastern English Channel, Southern North Sea and Central North Sea intensified sampling is carried out.

The recommended standard trawling gear of the North Sea IBTS is the mulitpurpose chalut à Grande Ouverture Verticale (GOV) trawl (ICES, 2012), which has been used on all participating vessels since 1992, while different pelagic and bottom trawls suitable for fishing finfish species were used before 1992. Standardized trawling protocols were adopted with a towing speed of 4 knots but depending on vessel performance, tide and weather conditions the average towing speed can be at minimum 3.5 and maximum 4.5 knots. From 2000-2018 trawling is done during the daylight hours, which are considered 15 minutes before sunrise to 15 minutes after sunset (ICES, 2012). After each trawl the total catch of the different species is weighed on board and biological parameters such as length for all fish species caught (to 0.1cm below for shellfish, to 0.5cm below for herring and sprat and to 1cm below for all other species) are collected. Where the numbers of individuals are too large for all of them to be measured to obtain the length distribution, a representative subsample of 100 fish is selected. Otoliths are collected on board from a small fraction of all the target species from all RFAs (Figure 1) to retrieve age reading. Table 3 in Web appendix B gives the minimum sampling levels of otoliths for the target species.



(a)

Figure 1: Standard roundfish areas used for roundfish since 1980, for all standard species since 1991 (left panel). Additional RFA 10 added in 2009. For example, the number 1 indicates ICES Index Area 1, and an ICES Statitical rectangle (ST) in IA 1 is 43F1. The map on the right panel shows norwegian trench and shelf edge (depths 1000-1500).

2 METHODS

This section gives the estimators of abundance indices. The estimators are haul time-based and utilizes an ALK approach. We consider the ALK approach used in DATRAS and we propose two ALK estimators.

The ALK used in DATRAS for computing abundance indices does not account explicitly for the spatial distribution in the age-length composition, which may be different and would result in a biased ALK. This difference may be caused either by variation in length-at-age distributions or by variations in the relative abundance of age classes, that is age-at-length distributions (Gerritsen et al., 2006). To account for the spatial distribution we propose a design-based ALK estimator that is haul dependent (Section 2.2.2) and a model-based ALK estimator (2.2.3).

2.1 Catch per unit effort

In this research, the catch per unit effort (CPUE) is defined as the number of fish of a certain species and age or length which are caught per hour trawl. In this section we define the CPUE mathematically, which explains how the index is calculated.

For a given species of interest, let $n_{h,l}$ be the number of fish with length l caught by trawl haul h. The CPUE for a given length l by trawl haul h is defined as

$$\text{CPUE}_{h,l} = \frac{n_{h,l}}{d_h}, \tag{2.1}$$

were d_h is the duration of the trawl in hours. The CPUE per age class is further defined as

$$\text{CPUE}_{h,a} = \sum_{l \in \mathbf{L}} \text{CPUE}_{h,l} \times ALK_{a,l,h}, \tag{2.2}$$

where $\mathbf{L}$ is the set of all length classes and $ALK_{a,l,h}$ is the age length key, which represents the estimated proportion of fish with age a in lth length class in haul h. For a given number of trawl hauls in a statistical rectangle, the mean CPUE defined as mCPUE in a statistical rectangle can be expressed as the average CPUE of the trawl hauls in the statistical rectangle:

$$\text{mCPUE}_{s,a} = \sum_{h \in H_s} \frac{\text{CPUE}_{h,a}}{|H_s|}. \tag{2.3}$$

Here H_s represents the set of trawl hauls taken in statistical rectangle s, and $|H_s|$ is the number of hauls taken in the rectangle. The mCPUE in pth RFA is further defined as

$$\text{mCPUE}_{p,a} = \sum_{s \in S_p} \frac{\text{mCPUE}_{s,a}}{|S_p|}\omega_s, \tag{2.4}$$

where S_p is the set of all statistical rectangles in RFA p, $|S_p|$ is the number of statistical rectangles in RFA p, and ω_s is a weight variable for each statistical rectangle. The weight variable ω_s varies between species. For some species ω equals 1 (e.g. Gadus morhua) for all s, and for other species it is the proportion of the statistical rectangle which has depth between 10 to 200 meters, for example Pollachius virens (see Table 4 in Web appendix C for weightings of statistical rectangles). The index for abundance at age in the whole study area, λ_a, is further defined by

$$\lambda_a = \frac{\sum_{p \in \mathbf{P}} A_p \mathrm{mCPUE}_{p,a}}{A_{\mathrm{total}}}. \tag{2.5}$$

Here $\mathbf{P}$ is the set of RFAs, A_p is the area of RFA p, and $A_{\mathrm{total}} = \sum_{p \in \mathbf{P}} A_p$.

2.2 ALK estimators

The definition of the CPUE of age includes an ALK, see (2.2), which we described in this section. Three ALK estimators are included in this research, which are named as follows: *i*) DATRAS ALK, *ii*) haul based ALK and *iii*) model based ALK.

2.2.1 DATRAS ALK

Let $\mathrm{ALK}^{\mathrm{D}}$ denote the DATRAS ALK. The $\mathrm{ALK}^{\mathrm{D}}$ is defined as constant within each RFA, and is calculated for each RFA by aggregating the age observation from each RFA. $ALK^{\mathrm{D}}_{a,l,h}$ used in equation (2.2) is defined as the proportion of observed fish with age a in length class l in the RFA h. If there are no observed fish in length class l in the RFA, ages from length classes close to l is used. The details of the procedure for borrowing strength from neighbouring length classes are given in Web appendix D.1. The underlying assumption of this ALK is that age-length compositions are homogeneous within the RFAs. This is a rather strong assumption, and any violation would have an unknown impact on the estimates of abundance indices. Aanes and Vølstad (2015) illustrated that violation of the assumption of constant ALK leads to biased estimates of CPUEs.

2.2.2 Haul based ALK

We define a haul dependent ALK by ALK^H. The $ALK^H_{a,l,h}$ is defined as the average proportion of observed fish with age a in length class l in haul h. If there are no observed ages of fish in a length class l in the haul, ages from the same length class in the haul close by is used (Web appendix D.2 describes the procedure in detail).

2.2.3 Spatial model-based ALK estimator

In this section we introduce a spatial model based ALK. Using such a model enables us to obtain smooth structures in the distribution of age given length. It further enables us to utilize spatial latent effects. Spatial model-based approach of age-lengths are widely used (Berg and Kristensen, 2012; Hirst et al., 2012; Rindorf and Lewy, 2001), and are used for stock assessment in the North Sea (Berg et al., 2014).

Let the response variable of the age group of a fish be $a = M, ..., A$ where M is the youngest age, and A is the oldest age which is typically defined as a "plus group". Suppose $y(l, \mathbf{s}, h)$ is the age of a fish with length l caught at location $\mathbf{s}$. As in Berg and Kristensen (2012) we use a continuous ratio model for the spatial age given length model. Define

$$\pi_a[y(l,\mathbf{s})] = P(y = a|y \geq a, l, \mathbf{s}) = \frac{p_a(l,\mathbf{s})}{p_a(l,\mathbf{s}) + \cdots p_{A-1}(l,\mathbf{s})} \quad \text{for } a = M, ..., A-1, \tag{2.6}$$

where $p_a(l, s)$ is the probability of a fish with length l at location $\mathbf{s}$ to be of age a. Note that $\pi_a(y(l, \mathbf{s}))$ is the probability of age a given that it has age greater than or equal to its age with length l at location s. Further is it assumed a logit link

$$\log\left[\frac{\pi_a[y(l,\mathbf{s})]}{1-\pi_a[y(l,\mathbf{s})]}\right] = \beta_a + f_a(l) + \gamma_a(\mathbf{s}). \tag{2.7}$$

Here β_a is an intercept, $f_a(l)$ is a continuous function of length and $\boldsymbol{\gamma}$ is a mean zero Gaussian spatial random field with Matérn covariance function. The spatial random field is intended to capture any spatial variation in the ALK.

The continuous function $f_a(l)$ in (2.7) is modelled with usage of P-splines (Wood, 2017), and these spline regression coefficients are included as a Gaussian random effect. The precision matrix for the spline regression coefficients is constructed such that the variability (or wryggliness) in the spline is penalized, see

Wood (2017, page 239) for details. The R package mgcv (Wood, 2015) is used for extracting the precision matrix needed for the spline regression coefficients.

We assume that the spatially Gaussian random field in (2.7), $\boldsymbol{\gamma}$, follows a stationary Matérn covariance structure:

$$\text{Cov}(\gamma(\mathbf{s}_1), \gamma(\mathbf{s}_2)) = \frac{\sigma_\gamma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa_\gamma||\mathbf{s}_1 - \mathbf{s}_2||)^\nu K_\nu(\kappa_\gamma||\mathbf{s}_1 - \mathbf{s}_2||), \tag{2.8}$$

where σ_γ^2 is the marginal variance, $||\cdot||$ is the Euclidean distance measure in kilometres, ν is a smoothing parameter, κ_γ is a spatial scale parameter and $K_\nu(\cdot)$ is the modified Bessel function of the second kind with $\nu = 1$. The spatial field is estimated with the stochastic partial differential equation (SPDE) procedure described in Lindgren et al. (2011). The main concept behind the SPDE procedure is that the precision matrix of a spatial field with Matérn covariance function can be approximated by a sparse matrix on a grid covering the area of interest. Such a grid and sparse precision matrix are constructed with use of the R-INLA package (Rue et al., 2009).

The model based ALK estimate is obtained by maximizing the likelihood. We maximize the likelihood with use of an R-Package called Template Model Building TMB (Kristensen et al., 2015), combined with the optimizing function nlminb in R. In this application TMB is advantageous as it uses Laplace approximation for the latent fields gaining computational efficiency, it also utilizes sparse structures in the latent fields, and uses automatic derivation.

2.3 Uncertainty estimation

In this section we describe how the uncertainty of the CPUE estimates are calculated. We use nonparametric bootstrapping to quantify the uncertainty of the CPUEs. In nonparametric bootstrapping independent samples of lengths and age are drawn with replacement from the original data and approximate 95% confidence intervals are obtained using bias-corrected percentile method (Carpenter and Bithell, 2000). Nonparametric resampling allows us to estimate the sampling distribution of the CPUE empirically without making assumptions concerning the data. The bias-Corrected method adjusts for the bias and skew of the sampling distribution of the data. This method assumes that there is a montonic increasing function and the estimator $\hat{\lambda}_a$ has a monotonic increasing function $f()$ such that the transformed values $f(\hat{\lambda}_a)$ are normally distributed with mean $f(\rho) - z_0$ and standard deviation one, where z_0 are the standard normal limits (Puth et al., 2015).

A bootstrap procedure for estimating the uncertainty of CPUEs in the North Sea is suggested in ICES (2013). In the rest of this paper, we refer to this procedure as DATRAS bootstrap procedure. The DATRAS procedure is divided into two parts; one part which samples CPUE per length (2.1), and another part which samples the ALK used in (2.2). The DATRAS bootstrap procedure is based on the assumption of homogeneous CPUE within RFAs. This assumption is likely to be wrong, and will typically cause an overestimation of the uncertainty. Therefore, we have included a bootstrap procedure, defined as the stratified bootstrap procedure, which instead assumes constant CPUE within each statistical rectangle.

2.3.1 DATRAS and Stratified bootstrap procedure

In this section we describe the bootstrap procedure for catch at length proposed by *DATRAS* (ICES, 2013) and the stratified procedure, and elaborate how the ALK is simulated. Assume there are N_s trawl hauls in a statistical rectangle. The DATRAS bootstrap procedure consists of sampling with replacement N_s of all trawl hauls in the corresponding RFA, and place them in the statistical rectangle. This procedure is performed independently across all statistical rectangles. It is worth reiterating that this procedure is based on the assumption that ALK is homogeneous in the whole RFA, and the implication of Datras bootstrap procedure on indices of abundance is two-fold. Firstly, Datras bootstrap procedure ignores the fine-scale stratification in the sampling process. This would lead to an overestimation of the uncertainty. Secondly, it ignores the sampling procedure of age-length data collected at the haul level. This would lead to an underestimation of the uncertainty. So there are biases in both directions, which are difficult to quantify.

The Stratified bootstrap procedure is a modification of the DATRAS bootstrap procedure. Rather than sampling hauls from the whole RFA, we sample the N_s trawl hauls from the list of hauls within the same statistical rectangle. If there is only one trawl haul within a statistical rectangle, we sample either that haul or the closest haul.

For simulating the DATRAS ALK we sample with replacement age observations within each RFA stratified with respect to length. If there is only one observed age from a given length class, we sample either that age or, at random, an age of the closest length class with observed ages. For the haul based ALK, we use the observed ages in the sampled hauls when simulating the CPUE per length (**is this from the stratified bootstrap procedure? what about explanations for the model based ALK? We also have codes for stratified procedure where DATRAS ALK is used, should we mention that this is done INSTEAD or AS WELL and include this results below in results section? Jon Helge mentioned that it's not plausible to show results of DATRAS bootstrap as it does not account for the survey design. Better off using our proposed stratified procedure with DATRAS ALK; include resulst for DATRAS using their sugested bootstrap procedure in appendix and reference in paper concerning higher variances**).

2.4 Reducing sampling effort

The current sampling procedure for the North Sea IBTS data is the sampling of one otolith from every observed length group in every trawl (see Table 3 in Web appendix B). We investigate the effect on the estimated mCPUE and its variance if the sampling procedure of otoliths changes such that fewer otoliths were collected. To determine this effect we remove otholits in a stratified manner, mimicking a sampling procedure where fewer otoliths are collected. For sampling fewer otoliths, we define wider length groups, for example 2 cm, or 3 cm, or 5 cm and so on, and simulate the otolith collection such that only one otolith is sampled from every wider length group. Estimated mCPUE's with summary statistics, based on the simulated reduced data sets are then compared with the parameters estimated from using all of data. In principle, we are free to define any length class to reduce the number of observed otoliths. For simplicity we propose two procedures: i) sample at random one otolith from every 2 cm length group, and ii) sample at random one otolith from every 5 cm length group.

3 Case studies

In this section we apply the methods described in Section 2 to data from the International Bottom Trawl Survey for the years 2017-2018, which is obtained from the DATRAS database (ICES, 2018c). These years are chosen for two reasons. The first is that in year 2018 new sampling procedures proposed by ICES for the collection of otoliths were introduced in the surveys. For instance, one otolith per length group is sampled for most target species (see Table 3 in Web appendix B for details of the sampling procedures for each target species), and this data is appropriate for the application of our proposed sample optimization procedure described in Section 2.4. The second is that IBTS included Age 0 in Q3 surveys, and since data for year 2018 Q3 is not yet available, the data for years 2017 Q3 and 2018 Q1 will be used in our analyses. Also, some species such as saithe that occupies the deeper waters in the northern part of the North Sea and in the Skagerrak and Kattegat, along the shelf edge (ICES, 2018a), the IBTS Q3 data is relevant for analyses compared with data from IBTS Q1 surveys, which do not adequately cover these areas where saithe is distributed (see Figure 1). Note that both IBTS Q1 and Q3 surveys do not adequately cover the whole stock distribution of saithe but the data collected is considered generally representative (ICES, 2016a).

In this research, the species of interest are cod and saithe. All samples are caught using the standard GOV gear described in Section 1.1. Table 1 gives a brief description of the data for year 2018 in the first quarter and year 2017 in the third quarter. Cod can be as old as 12 years in the first quarter and 11 years in the third quarter; and saithe as old as 18 years in the first quarter and 17 years in the third quarter. In our analyses we consider the age groups 1 to 6+ in Q1 and 0 to 6+ in Q3 for all ALK methods, where the last group consists of fish of age 6 or older. Saithe are typically older than cod but smaller in length, particularly in Q1. Catch rates are higher in the third quarter, 48% for cod and 164% for saithe, compared with the first quarter.

Table 1: Summary of North Sea IBTS cod and saithe (in parentheses) data for third quarter in year 2017 and first quarter in year 2018.

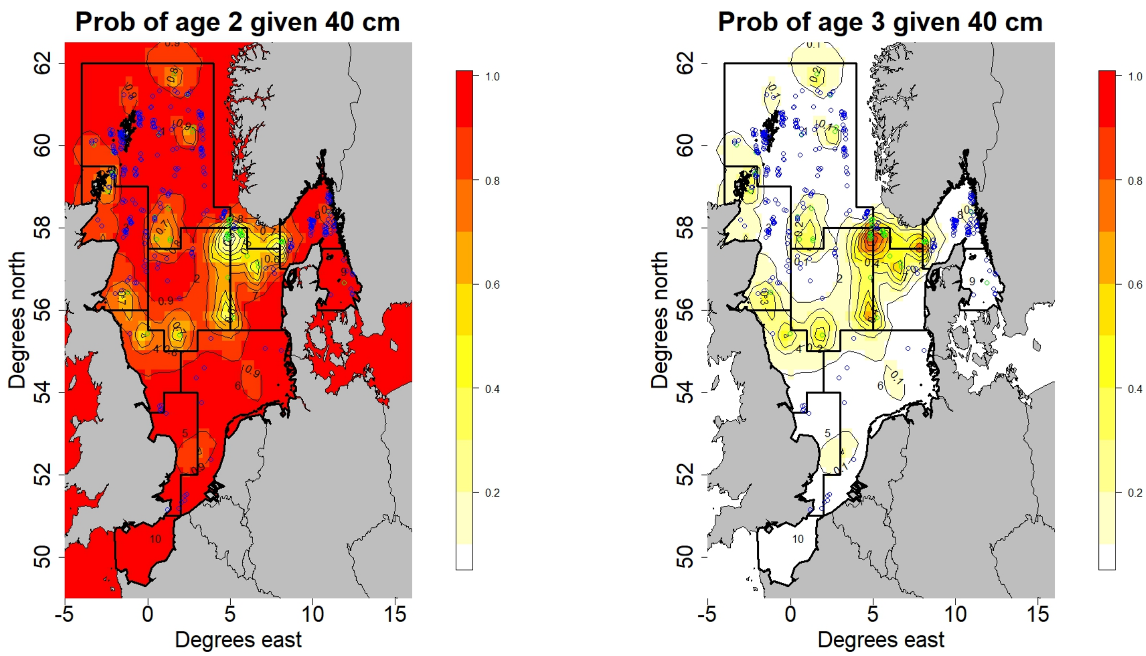
Data	**Description**
Trawl hauls	Total of 372 trawl hauls in year 2018 Q1; 238 (83) with length and 230 (81) with age information. In 2017 Q3, a total of 349 trawl hauls were taken; 238 (129) with length and 237 (128) with age information.
Age	The age varied between 1 (1) to 12 (18) years in year 2018 Q1 and 0 (0) to 11 (17) in year 2017 Q3.
Length	Length information in cm varied between 11 (13) to 114 (106) cm in year 2018 Q1 and between 6 (10) to 112 (109) cm in year 2017 Q3.
Date	Date of catch in year 2018 Q1 varied between 15.01.2018 to 28.02.2018 and in year 2017 Q3 between 18.07.2017 to 31.08.2018
Duration of haul	Mean duration is 29.37 minutes, with 30 minutes as 83% coverage interval in year 2018 Q1; and in 2017 Q3 with mean duration of 29.26 minutes with 30 minutes as 88% coverage .
Total count for all ages	1511 (793) in year 2018 Q1 and 2236 (2092) 2017 Q3.

3.1 Estimated indices of abundance and variability for cod and saithe

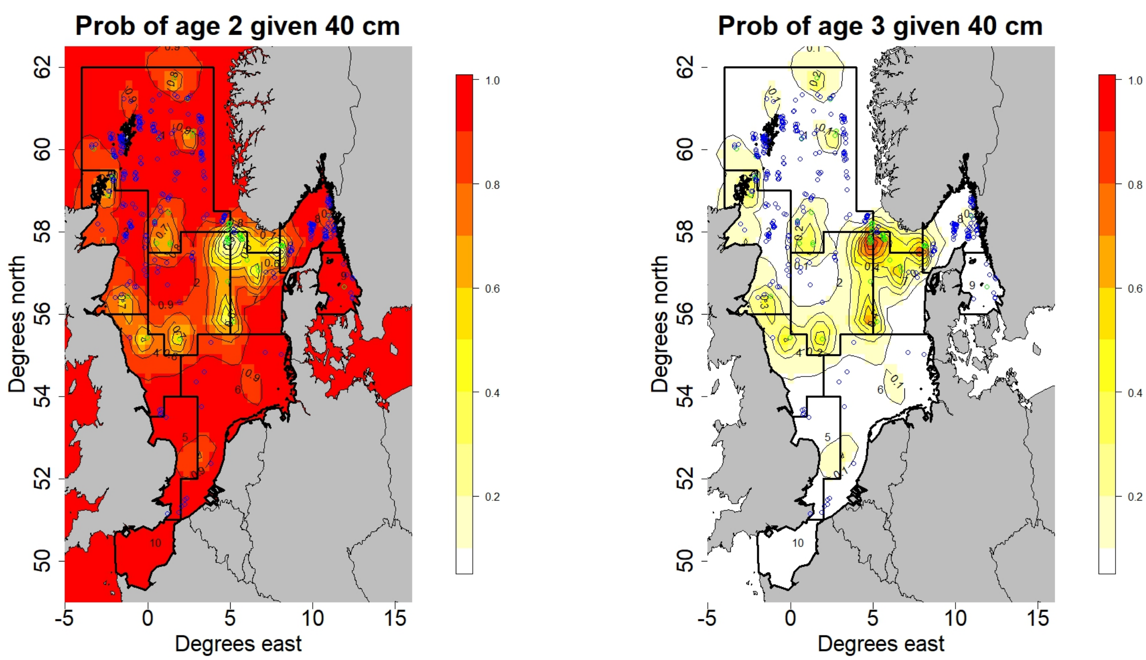
In this section we apply the three ALK methods given in section 2.2 for abundance estimation, and the two bootstrap methods, given in Section 2.3.1 for estimating variability of estimated indices of abundance. As discussed in Section 2.3.1 the Stratified bootstrap procedure is used for analyses in this research.

Recall that the main assumption of DATRAS ALK is that the age-length compositions of species over large areas are the same. To illustrate that this assumption may not be valid, we used the spatial ALK model (2.7) to predict probabilities of age given length of a 40 cm long cod and a 40 cm long saithe in the North Sea (Figure 2). These plots provide strong evidence against a null hypothesis of no spatial effect in the ALKs, as the likelihood of age given length changes in some areas. Figure 2 (a) shows that the eastern North Sea in RFAs 7 and 8 (the regions in yellow) is one of the areas where a 40 cm cod is more likely to be age 3. While 40 cm saithe is more likely to be......The plots also suggest that cod is distributed in all areas of the North Sea (Figure 2 (a)), whereas saithe is more likely to inhabit areas in the northern North Sea, specifically RFA 1 (Figure 2 (b)).

plot probability graph for saithe



(a) Probability plot of 40 cm cod in year 2018 Q1.



(b) Probability plot of 40 cm cod in year 2017 Q3

Figure 2: Predicted probabilities of age given length using model (2.6) and (2.7) for the year 2018 Q1. Graph (a) gives probabilities of predicted age of a 40 cm long cod, and graph (b) gives probabilities of predicted age of a ....cm saithe in RFAs 1 to 10 in the North Sea.

Figures 3 and 4 give estimates of indices of abundance for cod in years 2017 Q3 and 2018 Q1, and saithe in year 2017 Q3. Approximate 95% confidence intervals from the percentile and bias-corrected bootstrap methods for 200 bootstrap replication estimated from the three ALK methods.

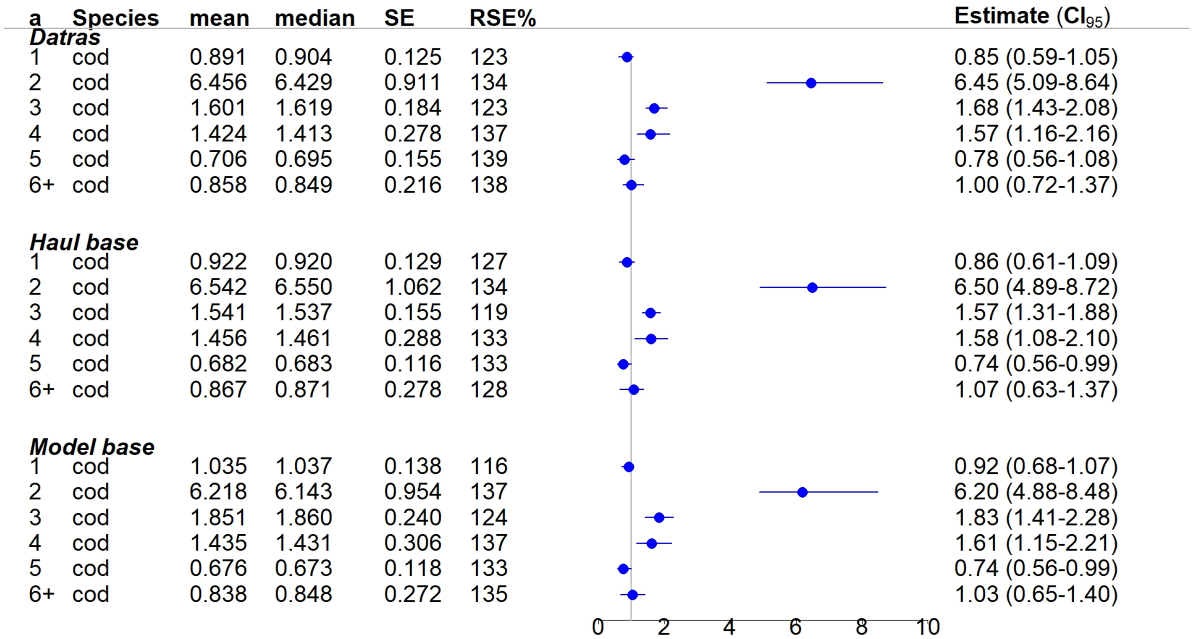
discuss more about confidence intervals

Figures 3 and 4 show that the resulting indices of abundance turned out to be similar for all ALKs. However, the implication of not accounting for variability over wider areas is higher uncertainty in the estimates, as shown by the DATRAS ALK. We would expect higher uncertainty estimates for older fishes from our ALK methods compared with DATRAS as variability is much higher for this group due to small sample sizes. However, in Q3 for both species estimated uncertainty is from our ALKs is similar or smaller compared with DATRAS ALK for the plus group......

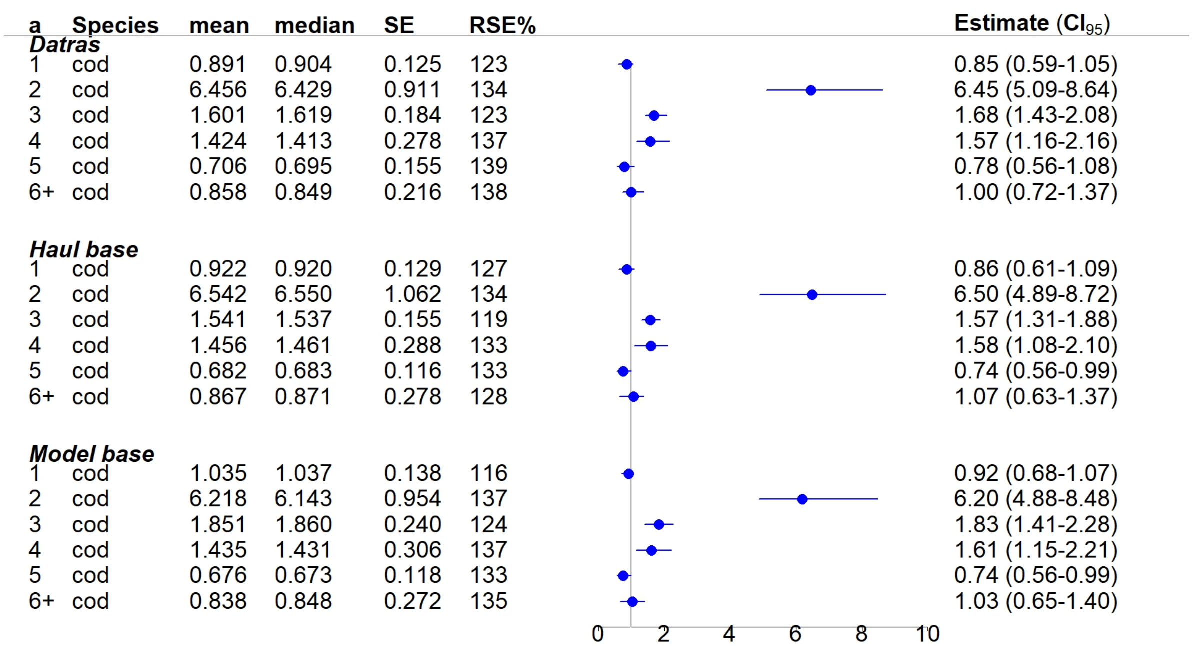
Discuss model base ALK results here....problems with variance-covariance structure etc using the required number of linear predictors e.g. A-1 or A-2 or an alternative approach to the fisher information or Hessian matrix, such as bootstrapping, for uncertainty estimation

For illustration we show the implications of using DATRAS bootstrap procedure for estimating the uncertainty around indices of abundance in Figure **??** in Web appendix E. Compared with the stratified bootstrap procedure, DATRAS bootstrap procedure gives an overestimation of the uncertainty for all age groups, suggesting that it is highly relevant to account for the variation in the data over large areas (**need to generate this plot for appendix**).

re-run codes with updated ALK model for cod and saithe (include 2017 Q3 plots)



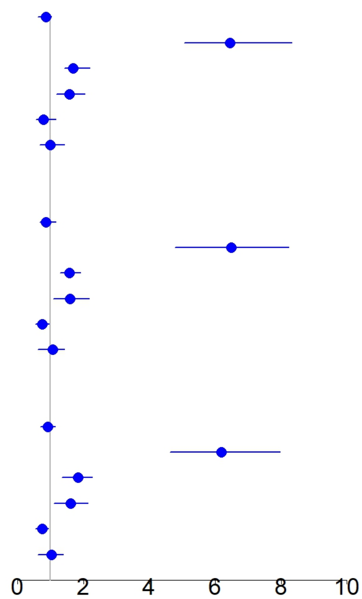
(a) Cod in year 2018 Q1



(b) Cod in year 2017 Q3

Figure 3: Estimated confidence intervals (CI_{95}) from bias-corrected bootstrap method for cod in years 2017 Q3 and 2018 Q1. Estimated indices of abundance (Estimate), and its standard error (SE), percentage relative standard error (RSE%), and bootstrap mean (mean) and median estimates are also given.

a	Species	mean	median	SE	RSE%	Estimate (CI_{95})
Datras						
1	saithe	0.898	0.899	0.112	122	0.85 (0.62-1.04)
2	saithe	6.632	6.575	0.948	129	6.45 (5.07-8.33)
3	saithe	1.598	1.599	0.192	130	1.68 (1.42-2.20)
4	saithe	1.446	1.432	0.273	130	1.57 (1.18-2.04)
5	saithe	0.720	0.706	0.150	148	0.78 (0.56-1.15)
6+	saithe	0.871	0.858	0.223	143	1.00 (0.68-1.42)
Haul base						
1	saithe	0.913	0.893	0.140	135	0.86 (0.67-1.16)
2	saithe	6.650	6.615	1.053	127	6.50 (4.78-8.22)
3	saithe	1.554	1.536	0.168	121	1.57 (1.29-1.90)
4	saithe	1.439	1.415	0.312	137	1.58 (1.09-2.17)
5	saithe	0.678	0.677	0.121	128	0.74 (0.55-0.95)
6+	saithe	0.845	0.844	0.306	132	1.07 (0.61-1.42)
Model base						
1	saithe	1.023	1.020	0.154	124	0.92 (0.69-1.14)
2	saithe	6.110	6.057	0.933	129	6.20 (4.63-7.97)
3	saithe	1.847	1.847	0.244	124	1.83 (1.35-2.26)
4	saithe	1.417	1.417	0.324	132	1.61 (1.10-2.13)
5	saithe	0.677	0.677	0.129	126	0.74 (0.53-0.94)
6+	saithe	0.819	0.822	0.281	133	1.03 (0.61-1.38)



(a) Saithe in year 2018 Q1

Figure 4: Estimated confidence intervals (CI_{96}) from bias-corrected bootstrap method for saithe in year 2018 Q1. Estimated indices of abundance (Estimate), and its standard error (SE), percentage relative standard error (RSE%), bootstrap mean (mean) and median estimates are also given.

3.2 Optimum sampling effort for North Sea Cod and Saithe

In order to determine optimum sampling levels of otoliths for saithe and cod in the North Sea, ALKs are estimated using the haul-based method. The haul based ALK and spatial ALK model gave similar estimates of abundance indices and precision (Figure 2) as both approaches account for spatial variation in the data. The spatial ALK model is quite complex, and model fitting would be computer-intensive since the model must be fitted for each bootstrap run and each simulated sampling procedure that mimics the real data collection procedure. Also, the assumption of no difference in regional compositions of age-length structures is invalid, as shown in Figure 2, so DATRAS ALK method is not use for further analyses. The removal procedure for otolith sampling described in Section 2.4 is applied to data in year 2018 Q1 for cod and year 2017 Q3 for saithe.

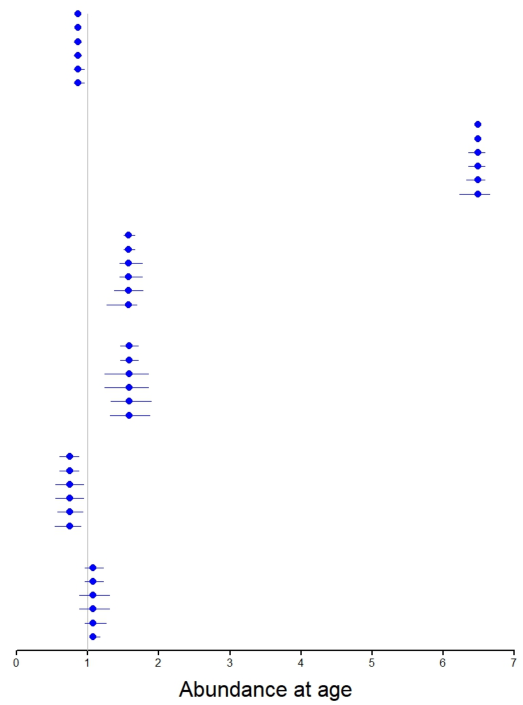
In year 2018 Q1, the number of otoliths sampled for cod was 1511, of which ....were removed in the experiments 1 cm, 2 cm, 3 cm, 4 cm, 5 cm or 7 cm length group, respectively. While 2092 otoliths were sampled for saithe in year 2017 Q3 and ....were removed in the experiments 1 cm, 2 cm, 3 cm, 4 cm, 5 cm or 7 cm length group, respectively. Figure5 gives estimates of abundance for the original data, estimated indices for the reduced data, and their standard errors, and approximate 95% confidence intervals for 200 simulations and 200 bootstrap replications.

The effect on abundance indices for 3 cm length or less is minimal for all age groups ....

Note that the nonparametric bootstrap method is advantageous because it does not assume any distribution for the data, and it also accounts for some of the variability in the sampling distribution of the CPUE, however, there are some limitations of this method. The most important limitation is the assumption that the distribution of the data represented by the sample is a reasonable estimate of the population function from which the data are sampled. If this assumption is violated the random sampling performed in the bootstrap procedure may add another level of sampling error, resulting in invalid statistical estimations (Haukoos and Lewis, 2005). As discussed in Section 1.1 the selection of the trawling locations in IBTS is semi-random where cruise leaders selects "clear" tow locations or "blind" tow locations if no clear tow exists by checking the proposed trawl track for hazardous seabed obstructions with acoustic methods. More recently selection of tow locations is based on pre-proposed valid tow locations with start and end positions executed in the period 2000-2017. Hence, the lack of a fully randomized sampling process has the potential to result in biased estimates of parameters and their uncertainty. Random sampling performed in the bootstrap procedure also adds another level of potential sampling error, which is reflected in variation and biased estimates commonly performed in the bootstrap analysis. Note that the sampling distribution of the bootstrapped statistics is frequently not symmetric and computing point estimates from in this manner may reflect biased estimation from the samples. This can be seen in the estimated bootstrap mean values in Figure 5

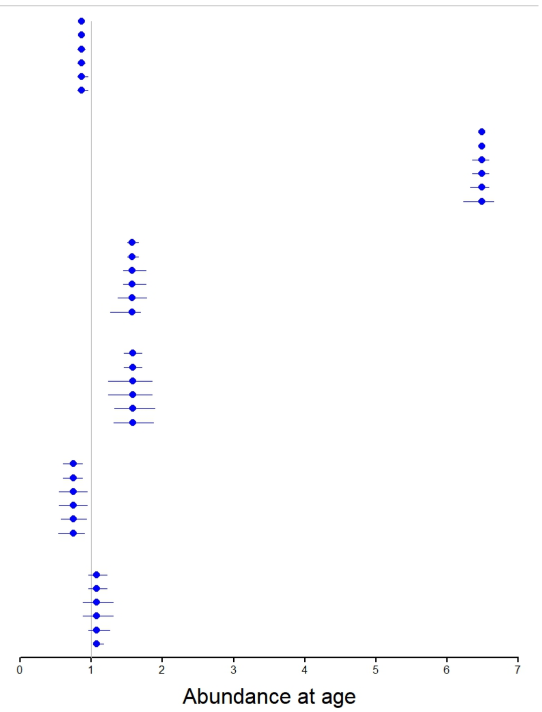
include correct plot for saithe

Length_group	Fulldata (mCPUE)	Reducedata (mCPUE)	SE(Reducedata)	CI-95 (Reducedata)
age 1				
1 cm	0.863	0.863	0.00000	(0.86-0.86)
2 cm	0.863	0.863	0.00000	(0.86-0.86)
3 cm	0.863	0.856	0.02740	(0.81-0.91)
4 cm	0.863	0.856	0.02740	(0.81-0.91)
5 cm	0.863	0.848	0.03317	(0.81-0.95)
7 cm	0.863	0.851	0.03764	(0.80-0.95)
age 2				
1 cm	6.496	6.497	0.00546	(6.49-6.50)
2 cm	6.496	6.497	0.00546	(6.49-6.50)
3 cm	6.496	6.506	0.05507	(6.36-6.59)
4 cm	6.496	6.506	0.05507	(6.36-6.59)
5 cm	6.496	6.510	0.06715	(6.33-6.59)
7 cm	6.496	6.497	0.09926	(6.23-6.65)
age 3				
1 cm	1.571	1.567	0.04181	(1.50-1.66)
2 cm	1.571	1.567	0.04181	(1.50-1.66)
3 cm	1.571	1.549	0.08874	(1.45-1.76)
4 cm	1.571	1.549	0.08874	(1.45-1.76)
5 cm	1.571	1.627	0.12090	(1.37-1.77)
7 cm	1.571	1.731	0.15068	(1.26-1.69)
age 4				
1 cm	1.584	1.589	0.06326	(1.46-1.70)
2 cm	1.584	1.589	0.06326	(1.46-1.70)
3 cm	1.584	1.613	0.15122	(1.24-1.85)
4 cm	1.584	1.613	0.15122	(1.24-1.85)
5 cm	1.584	1.559	0.14443	(1.32-1.89)
7 cm	1.584	1.518	0.15900	(1.31-1.88)
age 5				
1 cm	0.742	0.749	0.07014	(0.60-0.87)
2 cm	0.742	0.749	0.07014	(0.60-0.87)
3 cm	0.742	0.758	0.10749	(0.54-0.94)
4 cm	0.742	0.758	0.10749	(0.54-0.94)
5 cm	0.742	0.820	0.11949	(0.57-0.93)
7 cm	0.742	0.833	0.12728	(0.53-0.90)
age 6+				
1 cm	1.074	1.066	0.06985	(0.96-1.22)
2 cm	1.074	1.066	0.06985	(0.96-1.22)
3 cm	1.074	1.049	0.12244	(0.88-1.31)
4 cm	1.074	1.049	0.12244	(0.88-1.31)
5 cm	1.074	0.966	0.11135	(0.95-1.25)
7 cm	1.074	0.900	0.11003	(1.03-1.17)



(a) Cod in year 2018 Q1

Length_group	Fulldata (mCPUE)	Reducedata (mCPUE)	SE(Reducedata)	CI-95 (Reducedata)
age 1				
1 cm	0.863	0.863	0.00000	(0.86-0.86)
2 cm	0.863	0.863	0.00000	(0.86-0.86)
3 cm	0.863	0.856	0.02740	(0.81-0.91)
4 cm	0.863	0.856	0.02740	(0.81-0.91)
5 cm	0.863	0.848	0.03317	(0.81-0.95)
7 cm	0.863	0.851	0.03764	(0.80-0.95)
age 2				
1 cm	6.496	6.497	0.00546	(6.49-6.50)
2 cm	6.496	6.497	0.00546	(6.49-6.50)
3 cm	6.496	6.506	0.05507	(6.36-6.59)
4 cm	6.496	6.506	0.05507	(6.36-6.59)
5 cm	6.496	6.510	0.06715	(6.33-6.59)
7 cm	6.496	6.497	0.09926	(6.23-6.65)
age 3				
1 cm	1.571	1.567	0.04181	(1.50-1.66)
2 cm	1.571	1.567	0.04181	(1.50-1.66)
3 cm	1.571	1.549	0.08874	(1.45-1.76)
4 cm	1.571	1.549	0.08874	(1.45-1.76)
5 cm	1.571	1.627	0.12090	(1.37-1.77)
7 cm	1.571	1.731	0.15068	(1.26-1.69)
age 4				
1 cm	1.584	1.589	0.06326	(1.46-1.70)
2 cm	1.584	1.589	0.06326	(1.46-1.70)
3 cm	1.584	1.613	0.15122	(1.24-1.85)
4 cm	1.584	1.613	0.15122	(1.24-1.85)
5 cm	1.584	1.559	0.14443	(1.32-1.89)
7 cm	1.584	1.518	0.15900	(1.31-1.88)
age 5				
1 cm	0.742	0.749	0.07014	(0.60-0.87)
2 cm	0.742	0.749	0.07014	(0.60-0.87)
3 cm	0.742	0.758	0.10749	(0.54-0.94)
4 cm	0.742	0.758	0.10749	(0.54-0.94)
5 cm	0.742	0.820	0.11949	(0.57-0.93)
7 cm	0.742	0.833	0.12728	(0.53-0.90)
age 6+				
1 cm	1.074	1.066	0.06985	(0.96-1.22)
2 cm	1.074	1.066	0.06985	(0.96-1.22)
3 cm	1.074	1.049	0.12244	(0.88-1.31)
4 cm	1.074	1.049	0.12244	(0.88-1.31)
5 cm	1.074	0.966	0.11135	(0.95-1.25)
7 cm	1.074	0.900	0.11003	(1.03-1.17)



(b) Saithe in year 2017 Q3

Figure 5: Comparing estimated abundance from the original data where one otolith per length group: 1cm, 2cm, 3cm, 4cm, 5cm or 7cm is sampled for cod in year 2018 Q1 and saithe in year 2017 Q3. Estimated confidence intervals (CI $-$ 95), and its standard error (SE) estimates of the reduced samples are also given.

4 DISCUSSION

In this research we have determined minimum sampling efforts of otoliths for target species of the North Sea International Bottom Trawl Survey. This was achieved by testing sampling procedures that mimic the real data collection procedure but with a reduced number of otoliths. Several sampling procedures were tested and the effect on estimated abundance indices and their variance were investigated. Abundance indices were estimated using age-length keys (ALKs). The database for trawl surveys (DATRAS) manned by ICES includes an ALK that uses the raw proportions of age given length assuming constant age-length compositions over relatively large areas. We have developed two spatial ALK methods to estimate abundance indices and their variance that accounts for spatial variation in the data: 1) a haul based ALK that produces an ALK for each trawl haul, and which uses the raw proportions of age given length, and 2) a spatial ALK model that uses logits for modelling the age distribution in catch data from the length-stratified subsamples. Several studies have used spatial ALK modelling for estimating abundance indices of the North Sea stocks used in assessments (Berg and Kristensen, 2012; Berg et al., 2014; Gerritsen et al., 2006). These studies used continuous ratio logits with General Linear Model (GLM) or General Additive Models (GAMs) to model the spatial effects and found regional effects...... We propose to use Gaussian Random Field Theory to model the spatial effects as a smooth surface.........

- *discuss positives of using GRT to model spatial effects: what problems are eliminated when using this in terms of missing data*
- *compare the effects of our method with GAMs (Berg and Kristensen, 2012) and (Berg et al., 2014) and the NS-IBTS Delta-GAM index for estimating standardized age-based indices and the species theses are used for to include in assessment*
- *what does our model allows in terms of the age groups (samller or higher age groups (6+) possible with our model); covariates such as haul effect (included as a random effect)*

Also, both spatial ALK methods proposed in this paper provided a much better fit to the data compared with DATRAS ALK....

Reducing the number of otoliths by **x** percent had **no** significant effect on estimated abundance

- *discuss sampling procedure: limitation and advantages; and possibly more advanced selection procedures?*

- *new approach adopted in surveys from 2018*

- *IBTS has a standardized survey indices? -(yes Berg's NS-IBTS Delta-GAM index). so changes in catch rates are due to changes in population size? Berg et al. (2014) developed a standardized index for IBTS data but only applied to some species e.g., haddock? cod- last year 2017. Is the designed based age index on DATRAS not a standardized index?*

- *how does changes in survey design or other factors affect changes in catch rates? If so are these changes significant?*

5 General comments

- Decide on whether we say, "in this research or paper"

- Decide on whether to say, "In this subsection or section"

- Decide on year of data for case studies

- Decide on writing "haul(model)-based or haul (model) based

- Decide on calling the survey "The North Sea IBTS or IBTS"

- Decide on writing "Cod or cod, and Saithe or saithe"

- Decide on a title for the paper

- **what are issues with including haul effect in model based ALK? (Olav)**

References

Aanes, S. and Vølstad, J. H. (2015). Efficient statistical estimators and sampling strategies for estimating the age composition of fish. *Canadian journal of fisheries and aquatic sciences*, 72(6):938–953.

Berg, C. W. and Kristensen, K. (2012). Spatial age-length key modelling using continuation ratio logits. *Fisheries Research*, 129:119–126.

Berg, C. W., Nielsen, A., and 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.

Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in medicine*, 19(9):1141–1164.

Ehrhardt, N. M. and Legault, C. M. (1997). The role of uncertainty in fish stock assessment and management: a case study of the spanish mackerel, scomberomorus maculatus, in the us gulf of mexico. *Fisheries research*, 29(2):145–158.

Fridriksson, A. (1934). On the calculation of age-distribution within a stock of cod by means of relatively few age-determinations as a key to measurements on a large scale. *Rapports Et Proces-Verbaux Des Reunions, Conseil International Pour lExploration De La Mer*, 86:1–5.

Fuller, W. A. (2011). *Sampling statistics*, volume 560. John Wiley & Sons.

Gerritsen, H. D., McGrath, D., and Lordan, C. (2006). A simple method for comparing age–length keys reveals significant regional differences within a single stock of haddock (melanogrammus aeglefinus). *ICES Journal of Marine Science*, 63(6):1096–1100.

Haukoos, J. S. and Lewis, R. J. (2005). Advanced statistics: bootstrapping confidence intervals for statistics with difficult distributions. *Academic emergency medicine*, 12(4):360–365.

Hirst, D., Storvik, G., Rognebakke, H., Aldrin, M., Aanes, S., and Vølstad, J. H. (2012). A bayesian modelling framework for the estimation of catch-at-age of commercially harvested fish species. *Canadian journal of fisheries and aquatic sciences*, 69(12):2064–2076.

ICES (2012). Manual for the international bottom trawl surveys, revision viii. series of ices survey protocols. *International Council for the Exploration of the Sea*, SISP 1-IBTS VIII.

ICES (2013). Ns-ibts indices calculation procedure. datras procedure document. *International Council for the Exploration of the Sea*, 1.1 NS-IBST indices-2013.

ICES (2015). Manual for the international bottom trawl surveys, revision ix. series of ices survey protocols. *International Council for the Exploration of the Sea*, SISP 10-IBTS IX.

ICES (2016a). Ices advice on fishing opportunities, catch, and effort greater north sea and celtic seas ecoregions. *International Council for the Exploration of the Sea*, Book 6.

ICES (2016b). Report of the benchmark workshop on north sea stocks (wknsea), 14–18 march, 2016. *International Council for the Exploration of the Sea*, ICES CM 2016/ACOM:37.

ICES (2018a). Ices fishmap: Atlas of the north sea fish, including fact sheets of key species and distribution maps. *International Council for the Exploration of the Sea (ICES FishMap website, 2018*, 1.

ICES (2018b). Manual for the international bottom trawl surveys, revision x. series of ices survey protocols. *International Council for the Exploration of the Sea*, SISP X-IBTS X.

ICES, D. (2018c). Datras.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. (2015). Tmb: automatic differentiation and laplace approximation. *arXiv preprint arXiv:1509.00660*.

Lehtonen, R. and Pahkinen, E. (2004). *Practical methods for design and analysis of complex surveys*. John Wiley & Sons.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Ludwig, D. and Walters, C. J. (1981). Measurement errors and uncertainty in parameter estimates for stock and recruitment. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(6):711–720.

Nottestad, L., Utne, K. R., 'Oskarsson, G. J., J'onsson, S. T., Jacobsen, J. A., Tangen, O., Anthonypillai, V., Aanes, S., Vølstad, J. H., Bernasconi, M., et al. (2015). Quantifying changes in abundance, biomass, and spatial distribution of northeast atlantic mackerel (scomber scombrus) in the nordic seas from 2007 to 2014. *ICES Journal of Marine Science*, 73(2):359–373.

Puth, M.-T., Neuhäuser, M., and Ruxton, G. D. (2015). On the variety of methods for calculating confidence intervals by bootstrapping. *Journal of Animal Ecology*, 84(4):892–897.

Rindorf, A. and Lewy, P. (2001). Analyses of length and age distributions using continuation-ratio logits. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(6):1141–1152.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Walters, C. J. and Ludwig, D. (1981). Effects of measurement errors on the assessment of stock–recruitment relationships. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(6):704–710.

Wood, S. (2015). Package mgcv. *R package version*, pages 1–7.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

A Areas fished by different countries in the North Sea IBTS

Typically, two different countries fish each rectangle so that at least two trawl hauls are made per rectangle, but intensified sampling is carried out in the following areas: at least 3 hauls per rectangle are taken in statistical rectangles 31F1, 31F2, 32F1, 33F4, 34F2, 34F3, 34F4, 35F3, 35F4; while six or more hauls per rectangle are taken in statistical rectangles 30F1, 32F2, 32F3, 33F2, 33F3 (ICES 1999). The Skagerrak and Kattegat is fished solely by Sweden, who sample more than once in every rectangle while the west of Shetland (in Q1 and Q3) and inshore areas (Q3) is fished solely by Scotland. The edge of the Norwegian Trench is fished solely by Norway, but inshore areas near Denmark is fished by Denmark. The southern North Sea is fished by Denmark, Germany and England. France, typically, is the only country that surveys the western English Channel. Areas are surveyed by a single country because of the large proportion of untrawalable area (and subsequent gear damage issues experienced by other nations) for efficient logistical purposes. Table 2 gives the countries and research vessels participating the North Sea IBTS.

Table 2: Survey countries, vessel name, and period research vessels participating in first quarter (Q1) and third quarter (Q3) during 1997-2017.

	First Quarter (Q1)		**Third Quarter (Q3)**	
Country	Vessel name	Period	Vessel name	Period
Denmark	Dana	January-February	Dana	July-August
France	Thalassa II	January-February	-	-
Germany	Walther Herwig III	January-February	Walther Herwig III	July-August
Netherlands	Tridens 2	January-February	-	-
Norway	G.O. Sars	January-February	Johan Hjort	July
UK England	-	-	Endeavour	August-September
UK Scotland	Scotia III	January-February	Scotia III	July-August
Sweden	Dana	January-February	Dana	August

B Otolith sampling per fish species

From 1991-2017, most countries conducted quota sampling of otoliths per length group in a RFA. But from 2013 Norway has been sampling one otolith per length class from each trawl haul (to 0.1cm below for shellfish, to 0.5cm below for herring and sprat and to 1cm below for all other species). From the first quarter in 2018 all countries are required to sample one otolith per length class per trawl haul. Table 3 gives the minimum sampling levels of otoliths for the target species. However, for the smallest size groups, that presumably contain only one age group, the number of otoliths per length class may be reduced, and more otoliths per length are required for the larger length classes.

Table 3: Minimum sampling levels of otoliths by species for RFA or per trawl haul.

Period	Species	Minimum sampling levels of otoliths per length class
1991-2017		**Number of otoliths per length class in a RFA**
	herring	8 otoliths per $\frac{1}{2}$ cm group
	sprat	16 otoliths per $\frac{1}{2}$ cm length class $8.0-11.0$ cm
		12 otoliths per $\frac{1}{2}$ cm length class ≥ 11.0 cm
	mackerel	8 otoliths per $\frac{1}{2}$ cm length class
	cod	8 otoliths per 1 cm length class
	haddock	8 otoliths per 1 cm length class
	whiting	8 otoliths per 1 cm length class
	Norway pout	8 otoliths per 1 cm length class
	saithe	8 otoliths per 1 cm length class
	All target species	From 2013 Norway and Scotland, and Netherlands from 2016 have been sampling 1 otolith per length class from each trawl haul (to 0.1cm below for shellfish, to 0.5cm below for herring and sprat, and to 1cm below for all other species).
2018		**Number of otoliths per length class per trawl haul**
	herring	1 otolith per $\frac{1}{2}$ cm group
	sprat	1 otolith per $\frac{1}{2}$ cm length class $8.0-11.0$ cm
		1 otolith per $\frac{1}{2}$ cm length class ≥ 11.0 cm
	mackerel	1 otolith per 1 cm length class
	cod	1 otolith per 1 cm length class
	haddock	2 otoliths per 5 cm length class $11-15,\ 16-20,\ 21-25,\ 26-30$ cm
	Norway pout	2 otoliths per 5 cm length class $5-10,\ 11-15$ cm
		2 otoliths per 1 cm length class > 15 cm
	saithe	1 otolith per 1 cm length class
	plaice	1 otolith per 1 cm length class

C Weightings of Statistical Rectangles

Table 4: Weights used for Pollachius virens in equation (2.3).

StatRec	Weight	StatRec	Weight	StatRec	Weight	StatRec	Weight	StatRec	Weight
31F1	0.6	38F0	1	41F7	1	44F3	1	48E7	1
31F2	0.8	38F1	1	41F8	0.1	44F4	1	48E8	0.9
31F3	0.05	38F2	1	41G0	0.2	44F5	0.9	48E9	1
32F1	0.8	38F3	1	41G1	0.97	44F8	0.25	48F0	1
32F2	1	38F4	1	41G2	0.53	44F9	0.8	48F1	1
32F3	0.8	38F5	1	42E7	0.4	44G0	0.94	48F2	1
32F4	0.01	38F6	1	42E8	1	44G1	0.6	48F3	0.5
33F1	0.3	38F7	1	42E9	1	45E6	0.4	48G0	0.02
33F2	1	38F8	0.3	42F0	1	45E7	1	49E6	0.8
33F3	1	39E8	0.5	42F1	1	45E8	1	49E7	1
33F4	0.4	39E9	1	42F2	1	45E9	1	49E8	0.4
34F1	0.4	39F0	1	42F3	1	45F0	1	49E9	1
34F2	1	39F1	1	42F4	1	45F1	1	49F0	1
34F3	1	39F2	1	42F5	1	45F2	1	49F1	1
34F4	0.6	39F3	1	42F6	1	45F3	1	49F2	1
35F0	0.8	39F4	1	42F7	1	45F4	0.6	49F3	0.5
35F1	1	39F5	1	42F8	0.2	45F8	0.3	50E6	0.1
35F2	1	39F6	1	42G0	0.32	45F9	0.02	50E7	0.6
35F3	1	39F7	1	42G1	0.89	45G0	0.24	50E8	0.7
35F4	0.9	39F8	0.4	42G2	0.64	45G1	0.55	50E9	0.9
35F5	0.1	40E7	0.04	43E7	0.03	46E6	0.4	50F0	1
36F0	0.9	40E8	0.8	43E8	0.9	46E7	0.9	50F1	1
36F1	1	40E9	1	43E9	1	46E8	1	50F2	1
36F2	1	40F0	1	43F0	1	46E9	1	50F3	0.2
36F3	1	40F1	1	43F1	1	46F0	1	51E6	0
36F4	1	40F2	1	43F2	1	46F1	1	51E7	0
36F5	1	40F3	1	43F3	1	46F2	1	51E8	0.5
36F6	0.9	40F4	1	43F4	1	46F3	0.8	51E9	1
36F7	0.4	40F5	1	43F5	1	46F9	0.3	51F0	1
36F8	0.5	40F6	1	43F6	1	46G0	0.52	51F1	1
37E9	0.2	40F7	1	43F7	1	46G1	0.2	51F2	0.5
37F0	1	40F8	0.1	43F8	0.94	47E6	0.8	51F3	0
37F1	1	41E6	0.03	43F9	0.41	47E7	0.6	52E6	0
37F2	1	41E7	0.8	43G0	0.21	47E8	1	52E7	0
37F3	1	41E8	1	43G1	0.7	47E9	1	52E8	0
37F4	1	41E9	1	43G2	0.3	47F0	1	52E9	0.1
37F5	1	41F0	1	44E6	0.5	47F1	1	52F0	0.2
37F6	1	41F1	1	44E7	0.5	47F2	1	52F1	0.5
37F7	1	41F2	1	44E8	0.9	47F3	0.6	52F2	0.1
37F8	0.8	41F3	1	44E9	1	47F9	0.01		
38E8	0.2	41F4	1	44F0	1	47G0	0.3		
38E9	0.9	41F5	1	44F1	1	47G1	0.02		
52F3	0	41F6	1	44F2	1	48E6	1		

D Imputation for missing age samples

Catches of the target species are sampled (or subsampled with a size of 100 if the catches are too large) for length, and otoliths are typically collected from a subsample of the individuals sampled for length in the RFA, or per trawl haul as in the case of Norway for determining age of the fish (see Table **??**). In the case of

Norway where all trawl hauls are sampled for otoliths, missing age samples would still occur for the following two reasons: 1) the fish is below minimum length for otolith sampling (unreadable otoliths) or 2) otoliths are misplaced. Abundance indices by age group are estimated based on three age-length-keys (ALK): 1) DATRAS ALK estimator, 2) Haul dependent ALK estimator, and 3) Spatial model-based ALK estimator.

D.1 *DATRAS ALK Borrowing Approach*

The ALK proposed in DATRAS (ICES 2013), which is an aggregation of individual samples from a haul combined over a round fish area (RFA), and missing age samples are imputed as follows:

1. If there is no ALK for a length in the CPUE dataframe, age information is obtained accordingly
    - If length class (CPUE) $<$ minimum length class (ALK), then age=1 for the first quarter and age=0 for all other quarters
    - If minimum length class (ALK) $<$ length class (CPUE) $<$ maximum length (ALK) then age is set to the nearest ALK. If the ALK file contains values at equal distance, a mean is taken from both values.
2. If length class (CPUE) $>$ maximum length (ALK) age is set to the plus group.

The underlying assumption of this ALK approach is that age-length compositions are homogeneous within the superstrata.

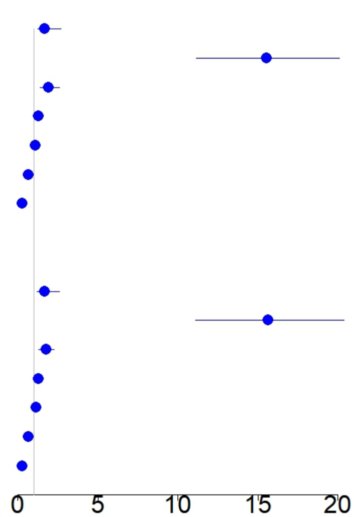
D.2 *Haul-based ALK Borrowing Approach*

The second is an a haul dependent ALK estimator which we propose, and is denoted by ALK^H. Since the age-length composition of fish may be space-variant, that is, there may be variation in age-length compositions between trawl stations within a superstrata, the spatial dependence of the age-length composition must be accounted for to produce reliable estimates of the CPUE per age estimates. If this spatial dependence is ignored not only will estimates of abundance be biased but the impact on the variance may be substantial. So for each trawl haul an ALK^H is produced. *Since there are few or none observations of ages for each length class in a trawl haul, length classes are therefore pooled in increasing order such that there are five* *length classes in each pooled length group. To replace missing values for the age distribution in the pooled length groups the method of "borrowing" ages from length groups in trawl hauls closest in air distance within the RFA is used. If there are no observed ages in the pooled length group in the RFA, missing values for the age distribution are replaced following the procedure outlined in the DATRAS ALK procedure (D.1) in step 1.*

E Estimates from DATRAS and Stratified bootstrap procedures

The naive bootstrap procedure prosed by DATRAS lacks the potential to account for the spatial variation in the data. The DATRAS bootstrap procedure ignores the fine-scale stratification in the sampling process, leading to an overestimation of the uncertainty; and ignores the age-length data collected at the haul level, resulting in an underestiamtion of the uncertainty. The results (Table 6) shows an overestimation of the uncertainty for all age groups, suggesting that it is relevant to account for the fine-scale stratification when resampling the data.

a	Species	mean	SE(Estimate)	RSE%	Estimate (CI_{95})
Datras					
0	cod	1.430	0.4408	25.96	1.70 (1.26-2.68)
1	cod	15.802	4.4119	28.35	15.56 (11.17-20.09)
2	cod	1.990	0.5101	26.72	1.91 (1.38-2.58)
3	cod	1.290	0.2067	15.78	1.31 (1.07-1.61)
4	cod	1.095	0.1834	16.68	1.10 (0.92-1.33)
5	cod	0.652	0.1534	23.50	0.65 (0.49-0.85)
6+	cod	0.289	0.0749	26.23	0.29 (0.21-0.41)
Stratified					
0	cod	1.467	0.3532	20.80	1.70 (1.19-2.62)
1	cod	15.417	2.5266	16.23	15.64 (11.09-20.36)
2	cod	1.891	0.3047	15.96	1.79 (1.30-2.28)
3	cod	1.281	0.1393	10.64	1.31 (1.06-1.55)
4	cod	1.097	0.1066	9.69	1.13 (0.96-1.33)
5	cod	0.642	0.0968	14.83	0.67 (0.47-0.86)
6+	cod	0.276	0.0473	16.55	0.28 (0.24-0.33)



(a) Datras and Stratified bootstrap Procedures

Figure 6: Comparison of estimated confidence intervals (CI_{96}) from Datras and stratified bootstrap procedures. The bias-corrected bootstrap method is used to give estimates for cod in year 2017 Q3. Estimated indices of abundance (Estimate), and its standard error (SE(Estimate), bootstrap mean (mean) and median estimates are also given.

Table 5 below will not be included in paper. For our purposes to look at estimates of saithe with and without weights and, when the code for weights was "incorrect"

Table 5: Average estimates of abundance indices for the North Sea cod and saithe species from 200 bootstrap samples in years 2017-2018. Standard error estimates (SE) (relative standard error, RSE in parentheses) and the lower bounds (LB) and upper bounds (UP) of approximate 95% confidence intervals from the three ALK methods are also given.

Species	a	**DATRAS ALK**				**Haul based ALK**				**Model based ALK**			
		$m\widehat{CPUE}_{N,a}$	SE(RSE)	LB	UB	$m\widehat{CPUE}_{N,a}$	SE (RSE)	LB	UB	$m\widehat{CPUE}_{N,a}$	SE (RSE)	LB	UB
saithe-weights													
2017 Q3	0	0.282	0.09 (31.9%)	0.135	0.431	0.282	0.09 (37.6%)	0.132	0.431		(%)		
	1	0.124	0.05 (40.3%)	0.043	0.217	0.123	0.05 (37.9%)	0.042	0.213		(%)		
	2	1.043	0.23 (22.1%)	0.557	1.463	0.929	0.26 (24.4%)	0.341	1.245		(%)		
	3	11.243	2.02 (18.0%)	7.076	15.03	11.238	2.12 (14.4%)	6.439	14.486		(%)		
	4	12.703	1.82 (14.3%)	9.022	16.013	12.789	2.14 (14.2%)	7.866	16.121		(%)		
	5	2.912	0.35 (12.0%)	2.317	3.655	2.971	0.33 (18.9%)	2.232	3.516		(%)		
	6+	2.845	0.52 (18.3%)	1.795	3.884	2.819	0.55 (17.6%)	1.744	3.855		(%)		
2018 Q1	1	0.035	0.01 (28.6%)	0.015	0.053	0.035	0.01 (28.6%)	0.017	0.052		(%)		
	2	0.049	0.01 (20.4%)	0.039	0.076	0.051	0.01 (19.6%)	0.038	0.074		(%)		
	3	1.679	0.91 (54.2%)	0.264	3.544	0.317	0.07 (22.1%)	0.166	0.423		(%)		
	4	9.997	5.26 (52.6%)	1.698	18.458	8.988	4.17 (46.4%)	1.861	15.639		(%)		
	5	7.476	3.59 (48.0%)	2.123	13.259	9.762	4.54 (46.5%)	2.319	16.734		(%)		
	6+	2.561	0.79 (30.8%)	1.305	3.983	2.646	0.76 (28.7%)	1.344	3.871		(%)		
saithe-weights old													
2017 Q3	0	0.744	0.26 (34.9%)	0.0326	1.154	0.744	0.28 (37.6%)	0.340	1.151		(%)		
	1	0.130	0.05 (38.4%)	0.046	0.213	0.132	0.05 (37.9%)	0.053	0.213		(%)		
	2	1.293	0.30 (23.2%)	0.658	1.812	1.190	0.29 (24.4%)	0.500	1.562		(%)		
	3	14.812	2.31 (15.6%)	9.282	18.088	14.979	2.16 (14.4%)	9.554	17.404		(%)		
	4	16.391	2.39 (14.6%)	10.705	20.116	16.256	2.31 (14.2%)	10.841	19.738		(%)		
	5	3.832	0.59 (15.4%)	2.782	4.979	3.913	0.74 (18.9%)	2.618	5.204		(%)		
	6+	3.190	0.54 (16.9%)	2.104	4.149	3.178	0.56 (17.6%)	2.183	4.300		(%)		
2018 Q1	1	0.059	0.03 (50.1%)	0.030	0.079	0.059	0.01 (16.9%)	0.031	0.041		(%)		
	2	0.06	0.02 (33.3%)	0.058	0.085	0.059	0.01 (16.9%)	0.057	0.067		(%)		
	3	2.805	2.63 (93.8%)	0.469	5.442	0.372	0.07 (18.8%)	0.305	0.427		(%)		
	4	16.752	14.06 (83.9%)	2.644	29.272	15.017	11.68 (69.7%)	3.081	24.999		(%)		
	5	12.117	10.35 (85.4%)	3.112	22.655	16.119	12.45 (77.8%)	3.457	26.877		(%)		
	6+	3.983	1.74 (43.7%)	2.126	5.408	4.150	2.03 (48.9%)	1.968	5.829		(%)		
saithe-no-weights													
2017 Q3	0	0.744	0.26 (34.9%)	0.0326	1.154	0.744	0.28 (37.6%)	0.340	1.151		(%)		
	1	0.130	0.05 (38.4%)	0.046	0.213	0.132	0.05 (37.9%)	0.053	0.213		(%)		
	2	1.293	0.30 (23.2%)	0.658	1.812	1.190	0.29 (24.4%)	0.500	1.562		(%)		
	3	14.812	2.31 (15.6%)	9.282	18.088	14.979	2.16 (14.4%)	9.554	17.404		(%)		
	4	16.391	2.39 (14.6%)	10.705	20.116	16.256	2.31 (14.2%)	10.841	19.738		(%)		
	5	3.832	0.59 (15.4%)	2.782	4.979	3.913	0.74 (18.9%)	2.618	5.204		(%)		
	6+	3.190	0.54 (16.9%)	2.104	4.149	3.178	0.56 (17.6%)	2.183	4.300		(%)		
2018 Q1	1	0.059	0.03 (50.1%)	0.030	0.079	0.059	0.01 (16.9%)	0.031	0.041		(%)		
	2	0.06	0.02 (33.3%)	0.058	0.085	0.059	0.01 (16.9%)	0.057	0.067		(%)		
	3	2.805	2.63 (93.8%)	0.469	5.442	0.372	0.07 (18.8%)	0.305	0.427		(%)		
	4	16.752	14.06 (83.9%)	2.644	29.272	15.017	11.68 (69.7%)	3.081	24.999		(%)		
	5	12.117	10.35 (85.4%)	3.112	22.655	16.119	12.45 (77.8%)	3.457	26.877		(%)		
	6+	3.983	1.74 (43.7%)	2.126	5.408	4.150	2.03 (48.9%)	1.968	5.829		(%)		