Optimising sampling effort of the North Sea International Bottom Trawl Survey Data

Abstract

In this research we present nonparametric 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 for generating abundance indices per age from the North Sea International Bottom Trawl Survey (IBTS) data. 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 research 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) and 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 base ALK, and ii) a spatial model base ALK. These ALKs are described in Section 2. A spatial model base 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. We apply these ALK methods to two case studies in Section 3, and a discussion is given in Section 4.

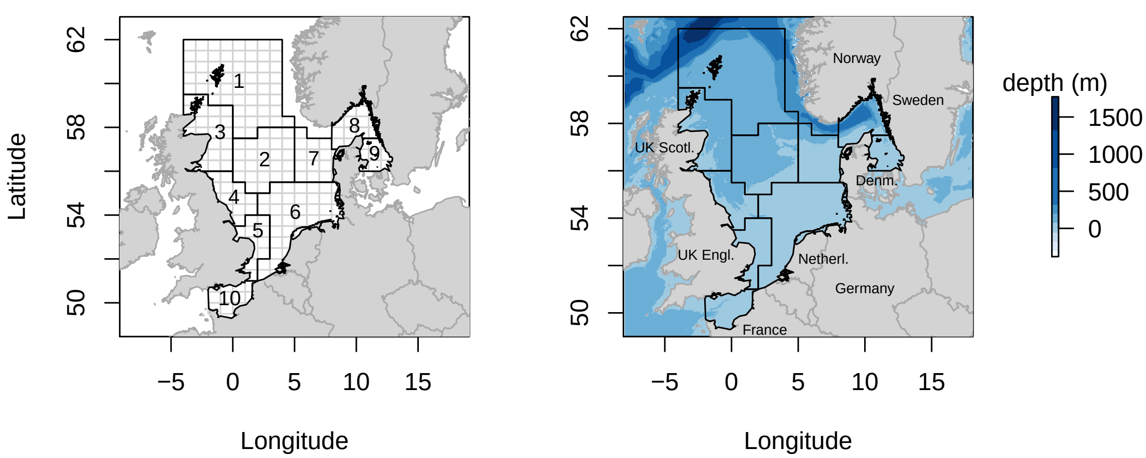
1.1 Overview of the North Sea International Bottom Trawl Survey

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 (July-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). There are also several by-catch species (see for example, ICES, 2006)

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 S1.1 in Supplementary Materials S1 gives details of the research vessels). The sampling frame is defined by the ICES index or roundfish areas (RFA) as shown in Figure 1 numbered 1 to 10. 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 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 S3.1 in appendix S3). 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 0.1 cm below for shellfish, to 0.5 cm below for herring and sprat and to 1 cm 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 S2.1 in Supplementary Materials S2 gives the minimum sampling levels of otoliths for the target species.



(a)

Figure 1: Standard roundfish areas (RFAs) used for roundfish since 1980 and for all standard species since 1991 (left panel). RFA 10 was added in 2009. The number 1, for example, indicates ICES RFA 1. The small grey rectangles in the left panel indicates the statistical rectangles of 30×30 nautical miles (1^o Longitude $\times$ 0.5^o Latitude). The map in the right panel shows the 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 (Kimura, 1977). 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 base ALK estimator that is haul dependent (Section 2.2.2) and a model base 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 S3.1 in Supplementary Materials S3). The mean catch per unit at age in the whole study area, λ_a, is defined by

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

This is know as the index of abundance at age, where $\mathbf{P}$ is the set of RFAs, A_p is the area of RFA p, and $A_{\text{total}} = \sum_{p \in \mathbf{P}} A_p$.

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 base ALK and *iii*) model base 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 Supplementary Materials S4.1. 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 base ALK

We define a haul dependent ALK by ALK^H. The $ALK^H_{a,l,h}$ used in equation (2.2) 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 (see Supplementary Materials S4.2 for the procedure).

2.2.3 Model base ALK

In this section we introduce a spatial model base ALK, which we define as ALK^{M}. 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 $y(l, \mathbf{s})$ 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. However, in our application we assume for each species we know a length l^* such that all fish above length l^* are above age M, and all fish with length below l^* are of age below A. By including such a variable we reduce the number of parameters in the model by removing one linear predictor. Define the continuous ratio we are modelling as

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

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 either $p_A(l, \mathbf{s})$ or $p_M(l, \mathbf{s})$ is known to be equal to zero, and the other is selected such that $\sum_a p_a = 1$. We further assume the logit link

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

Here $f_a(l)$ 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 species specific constant l^* is selected as the mid point between the shortest fish of age A and the longest fish of age M in the corresponding year and quarter. A sensitivity analysis of this constant were performed by adjusting it up and down 5 cm for cod in year 2018 in Q1. The point estimate of the mCPUEs then changed in the forth decimal, which is negligible.

The model base 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 (Puth et al., 2015; Karlsson, 2009). The bootstrap procedure is given in Supplementary Materials S5.

A bootstrap procedure for estimating the uncertainty of CPUEs in the North Sea is suggested in ICES (2013a). This procedure is given in Supplementary Materials S5. In the rest of this research, 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 would 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, 2013a) 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 should be remembered 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.

To estimate 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 both the haul based ALK and the model based ALK, we use the ages in the sampled hauls obtained when simulating the CPUE per length.

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 S2.1 in Supplementary Materials S2). 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

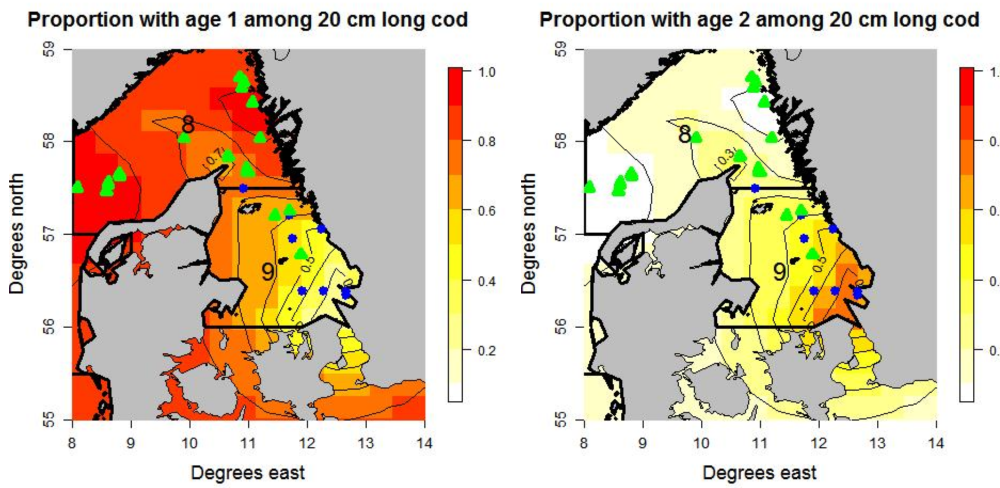
In this section we apply the methods described in Section 2 to data from the International Bottom Trawl Survey for the years 2017-2018, which is obtained from the DATRAS database (ICES, 2018c). These years are chosen for two reasons. The first is that in year 2018 new sampling procedures proposed by 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 S2.1 in Supplementary Materials S2 for 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 S6.1 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.

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 bias-corrected bootstrap method, given in Section 2.3.1 for estimating variability of estimated indices of abundance. The main assumption of DATRAS ALK is that the age-length compositions of species over large areas are the same. In this section we apply the three ALK methods given in section 2.2 for abundance estimation, and the bias-corrected bootstrap method, given in Section 2.3.1 for estimating variability of estimated indices of abundance. The main assumption of DATRAS ALK is that the age-length compositions of species over large areas are the same. Figure 2 illustrates the spatial model based ALK (2.7) for a 20 cm long cod, and gives a strong case on that the main assumption of DATRAS ALK is violated. From figure 2 gives an indication about the the proportion of two years old cod among 20 cm long cod is higher south and east in Skagerak.

Figures 3 gives estimates of indices of abundance for cod in years 2018 Q1 and for saithe in year 2017 Q3. Approximate 95% confidence intervals from the bias-corrected bootstrap method for 200 bootstrap replication are estimated from the three ALK methods. The stratified procedure described in 2.3.1 is used in



(a) (b)

Figure 2: Estimated proportion of age 1 and 2 among 20 cm long cod in Skagerak. The green triangles and blue points are observations of one and two year old cod respictively, which are in the length intervall 19 cm to 21 cm.

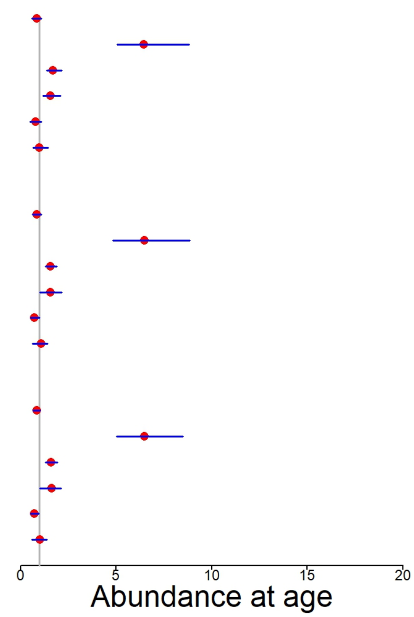
the sampling process to estimate bootstrap confidence intervals. Figures 3 shows that the resulting indices of abundance for cod and saithe turned out to be similar for all ALKs. IBTS is a complex multistage survey design, and since the ALKs are estimated from cluster-correlated data the resulting effective sample for estimating age-composition of fish would be lower than the number of fish measured (ICES, 2013b). Hence, the ALKs are subject to large sampling errors. For example, the estimated percentage relative standard errors from the spatial ALKs for the plus group (6+) for cod are $> 25\%$, suggesting high sampling error in the ALKs. Also, it should be remembered that DATRAS ALK is constant. Aanes and Vølstad (2015) showed that in such cases, and where only the variability of length compositions are allowed for, the estimated age-distributions may appear to be more precise than they truly are since the ALK itself is subject to sampling errors, see for example the estimated relative standard standard errors for ages 2, and the older fish (4, 5 and 6+) for both species.

As regards to which spatial ALK method to adopt, it is difficult to identify a method that gives the best performance over all age groups. While both methods seem to give reasonable estimates, the model base ALK generally gave shorter interval widths for both species. Uncertainty of relative abundance from the spatial ALK model is calculated using bootstrapping, as approximation of the joint distribution of the regression coefficient and spatial effect, in some cases, fails to account for the negative correlations between ages.

Also, estimating relative abundance at age and its precision from the spatial ALK model is computationalyl intensive. For these reasons we recommend the haul base ALK method for estimating age-distributions.

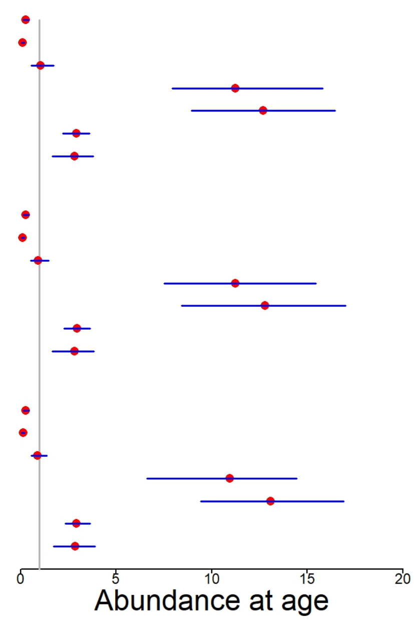
We also demonstrate the implications of using DATRAS bootstrap procedure for estimating the uncertainty around indices of abundance (see Figure S1 in Supplementary Materials S7). Compared with the stratified bootstrap procedure, DATRAS bootstrap procedure gives an overestimation of the uncertainty for all age groups, suggesting that it is highly relevant to account for the variation in the data over large areas.

Age	Species	Mean	Median	SE	RSE%	Width CI-95	Estimate (CI_{95})
Datras							
1	cod	0.902	0.915	0.133	15.5	0.468	0.85 (0.61-1.08)
2	cod	6.522	6.465	0.927	14.4	3.773	6.45 (5.05-8.82)
3	cod	1.587	1.562	0.202	12.0	0.781	1.68 (1.38-2.16)
4	cod	1.425	1.431	0.266	16.9	0.907	1.57 (1.18-2.09)
5	cod	0.712	0.704	0.156	20.0	0.566	0.78 (0.52-1.08)
6+	cod	0.879	0.845	0.227	22.7	0.782	1.00 (0.67-1.45)
Haul base							
1	cod	0.921	0.928	0.119	13.8	0.437	0.86 (0.63-1.07)
2	cod	6.513	6.509	0.984	15.2	4.010	6.50 (4.84-8.85)
3	cod	1.561	1.554	0.160	10.2	0.598	1.57 (1.31-1.91)
4	cod	1.494	1.510	0.319	20.1	1.118	1.58 (1.02-2.14)
5	cod	0.697	0.698	0.122	16.4	0.438	0.74 (0.54-0.98)
6+	cod	0.885	0.896	0.291	27.1	0.781	1.07 (0.62-1.40)
Model base							
1	cod	0.938	0.946	0.120	13.9	0.368	0.87 (0.66-1.03)
2	cod	6.623	6.607	0.964	14.8	3.461	6.49 (5.01-8.47)
3	cod	1.572	1.572	0.163	10.3	0.625	1.58 (1.30-1.92)
4	cod	1.503	1.519	0.326	20.1	1.114	1.62 (1.02-2.13)
5	cod	0.685	0.682	0.118	16.1	0.404	0.73 (0.54-0.95)
6+	cod	0.887	0.885	0.270	26.1	0.770	1.03 (0.60-1.37)



(a) Cod in year 2018 Q1

Age	Species	Mean	Median	SE	RSE%	Width CI-95	Estimate (CI_{95})
Datras							
0	saithe	0.276	0.280	0.0880	31.2	0.297	0.28 (0.14-0.43)
1	saithe	0.123	0.124	0.0527	42.5	0.173	0.12 (0.04-0.21)
2	saithe	1.005	0.997	0.2635	25.3	1.140	1.04 (0.57-1.71)
3	saithe	10.899	10.813	2.0905	18.6	7.845	11.24 (7.94-15.78)
4	saithe	12.424	12.425	1.8745	14.8	7.491	12.70 (8.95-16.45)
5	saithe	2.927	2.905	0.3528	12.1	1.384	2.91 (2.23-3.61)
6+	saithe	2.895	2.894	0.5621	19.8	2.138	2.84 (1.66-3.80)
Haul base							
0	saithe	0.284	0.283	0.0875	31.0	0.298	0.28 (0.13-0.43)
1	saithe	0.125	0.123	0.0501	40.6	0.169	0.12 (0.04-0.21)
2	saithe	0.872	0.871	0.2740	29.5	0.935	0.93 (0.52-1.46)
3	saithe	10.952	11.119	2.0775	18.5	7.903	11.24 (7.52-15.42)
4	saithe	12.535	12.647	2.1635	16.9	8.595	12.79 (8.41-17.00)
5	saithe	3.002	3.001	0.3634	12.2	1.362	2.97 (2.26-3.63)
6+	saithe	2.846	2.838	0.5860	20.8	2.166	2.82 (1.67-3.83)
Model base							
0	saithe	0.286	0.282	0.0875	31.2	0.296	0.28 (0.13-0.43)
1	saithe	0.130	0.132	0.0504	38.1	0.174	0.13 (0.04-0.22)
2	saithe	0.864	0.860	0.2083	23.1	0.807	0.90 (0.56-1.36)
3	saithe	10.872	11.037	2.0807	19.0	7.809	10.96 (6.62-14.43)
4	saithe	12.608	12.729	2.0261	15.5	7.456	13.09 (9.44-16.89)
5	saithe	2.888	2.875	0.3502	11.9	1.300	2.94 (2.34-3.64)
6+	saithe	2.772	2.752	0.5965	20.9	2.139	2.85 (1.74-3.88)



(b) Saithe in year 2017 Q3

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

3.2 *Optimum sampling effort for North Sea Cod and Saithe*

To determine optimum sampling levels of otoliths for saithe and cod in the North Sea, ALKs are estimated using the haul base method. As shown in Figure 3 the haul base ALK and the model base ALK gave similar estimates of abundance indices and precision as both approaches are attempting to capture spatial variation in the data. But, the spatial model base ALK 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. For this reason, the model base ALK approach will not be used in this analysis. Also, the assumption of no difference in regional compositions of age-length structures is invalid, as shown in Figure **??**, so DATRAS ALK 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. We sample one otolith per length group: 1 cm, 2 cm, 3 cm, 4 cm, 5 cm, 6 cm or 7 cm. Recall that the standardized IBTS sampling procedure is one otolith per length group for standard round fish as of year 2018 Q1, except for haddock and Norway Pout where 2 otoliths per cm is to be sampled (see Table S2.1 in Supplementary Materials S2).

Table1 compares estimates of relative abundance ($\hat{\lambda}_a$) for the original sample of otolith data with estimated indices for the reduced sample of otolith data ($\hat{\lambda}_a^*$) where a pair of otoliths are sampled from the following sampling procedures: 1 cm, 2 cm, 3 cm, 4 cm, 5 cm, 6 cm or 7 cm length group. Estimated standard errors of the estimated indices of abundance from the reduced data, with approximate bias-corrected bootstrap 95% confidence intervals and percentage relative standard error are given for 200 simulations and 200 bootstrap replication. Similar estimates are given for saithe in Table 2. A total of 1600 pairs of otoliths were sampled for cod in year 2018 Q1, while 2163 pairs of otoliths were sampled for saithe in year 2017 Q3 (see Table S6.1 in Supplementary Materials S6). The proportion of otoliths removed for cod from each of the sampling procedures stated above is: 14.4%, 28.6%, 38.4%, 44.5%, 49.3%, 52.6% or 55.6%, respectively, while for saithe the following proportions of otolihts are removed: 27.1%, 48.9%, 59.5%, 65.6%, 69.8%, 73.1% or 75.2%, respectively.

The results (Table 1) show that the difference between ($\hat{\lambda}_a$) and ($\hat{\lambda}_a^*$) is marginal for all sampling procedures for ages 1-4. For the older fishes (ages 5 and 6+) this difference is clear, particularly for sampling procedures $\geq$ 5 cm. This is supported by the percentage relative standard error, but for all ages RSE%$<$ 25%, suggesting that the variability in the estimates is low. A similar pattern emerges for satihe (Table 2), except for estimated relative abundance of age 2 saithe. It can be seen that the RSE% is substantially larger than those of the neighbouring age groups, for example 1 or 3 year-olds. Almost all the RSE% $>$ 30% for two-year olds, suggesting high variability in the parameter estimates (see $\hat{\lambda}_a$ and $\hat{\lambda}_a^*$ for age 2 saithe). As shown in Figure (**insert probability plot**) the predicted probabilities of age-given-length of 2-year old saithe overlaps with older fishes, for example 3 or 4-year olds, and the proportion of otoliths of 2-year olds sampled during IBTS 2017 Q3 survey is a mere 3% compared with 25% and 35% of 3-year and 4-year olds (see Table S6.2 in Supplementary Materials S6). And, given that a pair of otoliths (irrespective of age) is randomly selected from each of the following sampling procedures: 1 cm, 2 cm,...,7 cm, the older fishes (3 or 4-year olds) are more likely to be sampled, resulting in fewer samples utilised in the optimization procedure, and hence higher variability in the estimates. Furthermore, the sampling variability in lengths of 2-year olds are high compared with 1-year olds and are comparable with 3-year and 4-year olds, so the standard error estimates of the estimated abundance indices may be dominated by this uncertainty.

It can be seen also (Tables 1 and 2) for the older fishes ($\geq 3-$ years) the confidence intervals are wider compared with the younger fishes. This could be for several reasons. The nonparametric bias-corrected bootstrap method is adopted for estimating confidence intervals of indices of abundance, and although this method has the advantage of correcting for the bias and skew of the sampling distribution of the data; accounting for some of the variability in the sampling distribution of the CPUE; and does not assume any distribution for the data, there are some limitations of the bootstrap approach. 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 for IBTS surveys 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 2000-2018. Hence, the lack of a fully randomized sampling process has the potential to result in biased estimates of parameters and their uncertainty. Additionally, prior to 2013, all nations were sampling 8 pairs of otoliths per 1 cm length group for our focal species (Table S2.1 in Supplementary Materials S2), and these samples could be acquired from, for example the first haul (or first few trawl hauls), resulting in an unrepresentative sample of the population. From 2013, some nations adopted the current sampling procedure outlined by ICES for IBTS 2018 surveys of 1 pair of otolith per 1 cm length group from each haul, while other nations continued with sampling 8 pairs of otoliths per 1 cm length group. So, bias was still introduced via the sampling procedure. Another limitation of the bootstrap is the smaller the original sample the less likely it is to represent the entire population, thus the more difficult it becomes to compute valid confidence intervals. Note that the bootstrap relies heavily on the tails of the estimated sampling distribution when computing confidence intervals, and using small samples may jeopardize the validity of this computation. As explained above, in 2017 Q3 data for 2-year old saithe, for example, 65 pairs of otoliths out of 2163 were sampled (a mere 3%) and because of overlapping-age-length keys among neighbouring age groups, the probability of selecting 2-year olds is smaller, resulting in larger estimates of the variance, and hence wider confidence intervals.

need to conclude on which sampling procedure is best based on estimated uncertainty (RSE%), for example 5%?

Table 1: Estimated abundance ($\hat{\lambda}_a$) for cod from the original data in year 2018 Q1 compared with estimated abundance ($\hat{\lambda}_a^*$) from the reduced data for different sampling procedures of length groups (l). The estimated standard error of $\hat{\lambda}_a^*$ (SE($\hat{\lambda}_a^*$)) and the percentage relative standard error (RSE%) are also given.

	l	$\hat{\lambda}_a$	$\hat{\lambda}_a^*$	SE($\hat{\lambda}_a^*$)	RSE%	CI-95 ($\hat{\lambda}_a^*$)
age 1						
	1 cm	0.863	0.866	0.00910	1.051	(0.84, 0.88)
	2 cm	0.863	0.866	0.00969	1.119	(0.84, 0.88)
	3 cm	0.863	0.858	0.02476	2.886	(0.81, 0.90)
	4 cm	0.863	0.854	0.02993	3.507	(0.81, 0.90)
	5 cm	0.863	0.848	0.03712	4.379	(0.81, 0.92)
	6 cm	0.863	0.858	0.03646	4.249	(0.80, 0.93)
	7 cm	0.863	0.856	0.03920	4.582	(0.80, 0.93)
age 2						
	1 cm	6.496	6.485	0.02055	0.317	(6.47, 6.53)
	2 cm	6.496	6.485	0.01968	0.303	(6.46, 6.52)
	3 cm	6.496	6.503	0.05023	0.772	(6.38, 6.60)
	4 cm	6.496	6.498	0.05852	0.901	(6.39, 6.62)
	5 cm	6.496	6.507	0.07121	1.094	(6.36, 6.64)
	6 cm	6.496	6.510	0.07415	1.139	(6.34, 6.64)
	7 cm	6.496	6.491	0.08395	1.293	(6.33, 6.64)
age 3						
	1 cm	1.571	1.574	0.06955	4.418	(1.46, 1.73)
	2 cm	1.571	1.579	0.06785	4.429	(1.45, 1.71)
	3 cm	1.571	1.559	0.09708	6.228	(1.41, 1.75)
	4 cm	1.571	1.641	0.10051	6.124	(1.41, 1.84)
	5 cm	1.571	1.627	0.12505	7.686	(1.31, 1.88)
	6 cm	1.571	1.643	0.12598	7.670	(1.31, 1.91)
	7 cm	1.571	1.753	0.13991	7.979	(1.44, 2.02)
age 4						
	1 cm	1.584	1.603	0.12104	7.550	(1.37, 1.85)
	2 cm	1.584	1.592	0.11493	7.218	(1.37, 1.82)
	3 cm	1.584	1.610	0.14425	8.960	(1.31, 1.91)
	4 cm	1.584	1.571	0.15284	9.729	(1.31, 1.88)
	5 cm	1.584	1.585	0.15010	9.468	(1.32, 1.89)
	6 cm	1.584	1.588	0.15178	9.560	(1.26, 1.88)
	7 cm	1.584	1.503	0.16675	11.094	(1.38, 1.82)
age 5						
	1 cm	0.742	0.740	0.10729	14.504	(0.56, 095)
	2 cm	0.742	0.744	0.11047	14.848	(0.57, 0.98)
	3 cm	0.742	0.764	0.11675	15.287	(0.55, 1.02)
	4 cm	0.742	0.752	0.11398	15.158	(0.58, 0.99)
	5 cm	0.742	0.814	0.12034	14.789	(0.53, 1.08)
	6 cm	0.742	0.782	0.10975	14.043	(0.57, 1.02)
	7 cm	0.742	0.813	0.12545	15.425	(0.59, 1.09)
age 6+						
	1 cm	1.074	1.063	0.10489	9.865	(0.90, 1.28)
	2 cm	1.074	1.065	0.10657	10.009	(0.90, 1.29)
	3 cm	1.074	1.037	0.12215	11.775	(0.88, 1.25)
	4 cm	1.074	1.015	0.11855	11.679	(0.89, 1.24)
	5 cm	1.074	0.950	0.12210	12.849	(0.93, 1.21)
	6 cm	1.074	0.951	0.11562	12.154	(0.92, 1.19)
	7 cm	1.074	0.914	0.12577	13.760	(0.96, 1.18)

Table 2: Estimated abundance ($\hat{\lambda}_a$) for saithe from the original data in year 2017 Q3 compared with estimated abundance ($\hat{\lambda}^*_a$) from the reduced data for different sampling procedures of length groups (l).

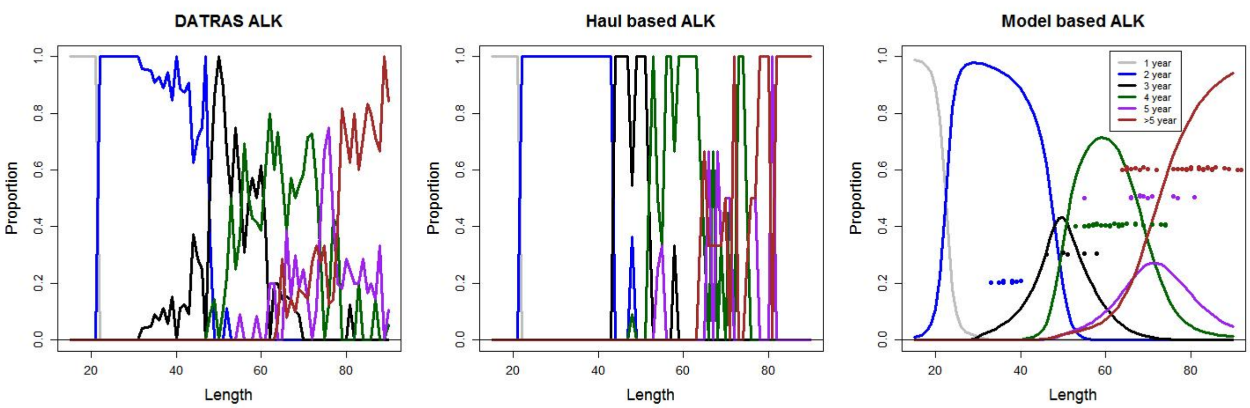
	l	$\hat{\lambda}_a$	$\hat{\lambda}^*_a$	SE($\hat{\lambda}^*_a$)	RSE%	CI-95 ($\hat{\lambda}^*_a$)
age 0						
	1 cm	0.282	0.282	0.00000	0.00	(0.28, 0.28)
	2 cm	0.282	0.282	0.00000	0.00	(0.28, 0.28)
	3 cm	0.282	0.288	0.00628	2.18	(0.28, 0.29)
	4 cm	0.282	0.291	0.00578	1.99	(0.28, 0.29)
	5 cm	0.282	0.282	0.00000	0.00	(0.28, 0.28)
	6 cm	0.282	0.297	0.00949	3.19	(0.28, 0.31)
	7 cm	0.282	0.290	0.00595	2.05	(0.28, 0.29)
age 1						
	1 cm	0.123	0.123	0.00000	0.00	(0.12, 0.12)
	2 cm	0.123	0.123	0.00000	0.00	(0.12, 0.12)
	3 cm	0.123	0.117	0.00628	5.36	(0.11, 0.12)
	4 cm	0.123	0.117	0.00641	5.47	(0.11, 0.13)
	5 cm	0.123	0.125	0.00139	1.11	(0.12, 0.13)
	6 cm	0.123	0.112	0.00942	8.43	(0.11, 0.13)
	7 cm	0.123	0.115	0.00630	5.46	(0.11, 0.13)
age 2						
	1 cm	0.929	0.917	0.17774	19.39	(0.58, 1.25)
	2 cm	0.929	0.892	0.28437	31.89	(0.47, 1.50)
	3 cm	0.929	0.985	0.33392	33.90	(0.51, 1.65)
	4 cm	0.929	0.982	0.33738	34.36	(0.48, 1.55)
	5 cm	0.929	1.003	0.34661	34.56	(0.50, 1.61)
	6 cm	0.929	0.945	0.33643	35.60	(0.49, 1.77)
	7 cm	0.929	0.982	0.33668	34.28	(0.42, 1.62)
age 3						
	1 cm	11.238	11.272	0.53160	4.72	(10.30, 12.15)
	2 cm	11.238	11.161	0.93489	8.38	(9.56, 13.11)
	3 cm	11.238	11.091	1.17954	10.64	(9.15, 13.39)
	4 cm	11.238	10.903	1.05854	9.71	(9.45, 13.27)
	5 cm	11.238	10.857	1.12152	10.33	(9.16, 12.95)
	6 cm	11.238	10.937	1.10301	10.09	(9.77, 12.88)
	7 cm	11.238	10.724	1.04157	9.71	(8.80, 13.07)
age 4						
	1 cm	12.789	12.773	0.52290	4.09	(11.89, 13.74)
	2 cm	12.789	12.874	0.91838	7.13	(10.80, 14.54)
	3 cm	12.789	12.877	1.14444	8.89	(10.76, 14.84)
	4 cm	12.789	13.019	1.09987	8.45	(10.69, 15.09)
	5 cm	12.789	13.033	1.15195	8.84	(10.51, 15.25)
	6 cm	12.789	13.111	1.09241	8.33	(10.57, 15.13)
	7 cm	12.789	13.278	1.07854	8.12	(10.23, 15.64)
age 5						
	1 cm	2.971	2.967	0.12220	4.12	(2.76, 3.24)
	2 cm	2.971	3.036	0.29591	9.75	(2.48, 3.58)
	3 cm	2.971	2.970	0.29382	9.89	(2.47, 3.61)
	4 cm	2.971	3.045	0.38412	12.61	(2.31, 3.99)
	5 cm	2.971	3.077	0.43326	14.08	(2.31, 3.87)
	6 cm	2.971	2.936	0.38625	13.16	(2.23, 3.77)
	7 cm	2.971	2.971	0.42992	14.14	(2.12, 3.81)
age 6+						
	1 cm	2.819	2.817	0.05126	1.82	(2.73, 2.90)
	2 cm	2.819	2.783	0.08525	3.06	(2.70, 2.94)
	3 cm	2.819	2.823	0.11729	4.15	(2.64, 3.04)
	4 cm	2.819	2.794	0.12590	4.15	(2.61, 3.05)
	5 cm	2.819	2.775	0.13938	5.02	(2.54, 3.02)
	6 cm	2.819	2.813	0.14435	5.13	(2.53, 3.12)
	7 cm	2.819	2.790	0.15922	5.71	(2.54, 3.06)

4 DISCUSSION

In this research we have determined optimal sampling efforts of otoliths for target species of the North Sea International Bottom Trawl Survey (IBTS). This was achieved by testing different sampling procedures that mimic the real data collection procedure but with a reduced number of otoliths. The estimated indices of abundance and their estimated uncertainty were investigated to determine if there is any real change in the precision of the estimates. 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 base 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 large spatio-temporal variability of the ALK and relative abundance at age. We proposed to use Gaussian Random Field Theory to model the spatial effects as a smooth surface to estimate age-at-length and relative abundance for the IBTS data. The spatial model base ALK and the design base spatial ALK (haul base) gave similar estimates as DATRAS estimator for relative abundance at age but the spatial ALK estimators gained better precision.

The spatial ALK model base estimator appears to be a useful tool to detect significant differences between ALKs over large areas, although estimation of the uncertainty in the ALK from the joint precision matrix is problematic. Including the uncertainty of the ALK in the model requires an approximation of the joint distribution of the regression coefficient and the spatial effect, but this approximation is only as good as the quality of the data in a given year and quarter, for example Figure (**insert probability plot**) shows that the approximation of the ALK for a cod of length 90 cm is likely to be 2 years old in year 2018 Q1. This occurs presumably because the approximation fails to account for the negative correlation structures between ages. So the uncertainty in the relative abundance was, therefore, calculated using bootstrapping as done by Berg and Kristensen (2012); Berg et al. (2014). In future, the model might be expanded to include the probability of recording inaccurate age-at-length data, so that uncertainty in the ALK could be estimated using the joint precision matrix. The model might also be expanded to include covariates such as trawl hauls to capture any haul variation, for example a trawl haul may "hit" a school of fish of a certain age.

Figure 4 illustrates the estimated age compositions as a function of length for a given haul in RFA 1. The haul selected is the haul with the most number of observed ages of cod in 2018 Q1. Notice that the the model based ALK is smooth, while the DATRAS ALK and the haul based ALK are not. This is an important advantage of the model based ALK, and we found it surprising that it did not result in a larger difference in the estimated mCPUE shown in table 1 and 2. An intuitive reason for that the model based ALK procedure did not result in different mCPUE estimates, is that there are enough observed ages per length group for the haul based ALK to be representative.



(a) (b) (c)

Figure 4: Estimated age compositions of cod as function of length in a given haul in RFA 1 with a) DATRAS ALK, b) haul based ALK and c) model based ALK. Note that explenation of the colors are only given in c). Each colored point in c) illustrates that there was an observed cod with the correspondning length and age in the haul. The haul selected is the haul with most observed ages of cod in 2018 Q1.

discuss DATRAS and Haul base ALK and recommended optimum sampling level of otoliths per length group

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.

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 (2006). Report of the workshop on implementation in datras of confidence limits estimation of abundance indices from bottom trawl survey data. ices datras report. *International Council for the Exploration of the Sea*, DRAFT.

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

ICES (2013b). Report of the third workshop on practical implementation of statistical sound catch sampling programmes. *International Council for the Exploration of the Sea*, WKPICS3.

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.

Karlsson, A. (2009). Bootstrap methods for bias correction and confidence interval estimation for nonlinear quantile regression of longitudinal data. *Journal of Statistical Computation and Simulation*, 79(10):1205–1218.

Kimura, D. K. (1977). Statistical assessment of the age–length key. *Journal of the Fisheries Board of Canada*, 34(3):317–324.

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.

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.

Supplemental Materials: Optimizing sampling effort of the North Sea International Bottom Trawl Survey.

S1 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 S1.1 gives the countries and research vessels participating the North Sea IBTS.

Table S1.1: 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

S2 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 S2.1 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 S2.1: 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

S3 Weightings of Statistical Rectangles

The weightings of the some statistical rectangles are alloted to species such as sprat, saithe and herring by depth strata. Table S3.1 gives these weights, which are used in the analysis of the saithe data.

Table S3.1: 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		

S4 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 base ALK estimator, and 3) Spatial model base ALK estimator.

S4.1 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 RFA.

S4.2 Haul base 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 (S4.1) in step 1.**

S5 Nonparametric Bootstrap Sampling procedure

Nonparametric bootstrapping is attractive as it makes no distributional assumption, and is suitable for estimating confidence interval for indices of abundance. Suppose we have a vector $\mathbf{x}$ of m independent observations, and we are interested in estimating a parameter $\hat{\theta}(\mathbf{x})$ and its variance. The general nonparametric bootstrap algorithm is as follows:

1. Sample m observations randomly with replacement from $\mathbf{x}$ to obtain a bootstrap data set, denoted by $\mathbf{x}^*$.
2. Calculate the bootstrap version of the statistic of interest, $\theta^* = \hat{\theta}(\mathbf{x}^*)$.
3. Repeat steps 1 and 2 a large number of times, say B, to obtain an estimate of the bootstrap distribution
4. calculate the average of the bootstrapped statistics, $\sum_{b=1}^{B} \theta^*{}_{(b)}/B$
5. compute the variance of the estimator $\hat{\theta}(\mathbf{x})$ through the variance of the set $\theta^*{}_{(b)},\ b = 1, 2, ..., B$, given by
$$\frac{\sum_{b=1}^{B} \left(\theta^*{}_{(b)} - \theta^*{}_{(\cdot)}\right)^2}{(B-1)} \tag{S5.1}$$
where $\theta^*{}_{(\cdot)} = \sum_{b=1}^{B} \theta^*{}_{(b)}/B$.

The Bias-Corrected 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 $f(\lambda_a) - z_0$ and standard deviation one, where z_0 are the standard normal limits (Puth et al., 2015; Karlsson, 2009). Now, let $P^*\left(\hat{\theta}(\mathbf{x}^*) \leq \hat{\theta}(\mathbf{x})\right)$ denote the proportion of $\hat{\theta}(\mathbf{x}^*)'$s in the bootstrap sample that have a value lower than the value of the parameter estimate $\hat{\theta}(\mathbf{x})$, and le z_0 be defined as

$$z_0 = \Phi^{-1}\left\{P^*\left(\hat{\theta}(\mathbf{x}^*) \leq \hat{\theta}(\mathbf{x})\right)\right\}, \tag{S5.2}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. Also let $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ be defined as

$$\tilde{\alpha}_1 = \Phi(2z_0 + z_\alpha), \tag{S5.3}$$

and

$$\tilde{\alpha}_2 = \Phi(2z_0 + z_{1-\alpha}), \tag{S5.4}$$

respectively. A $100(1-2\alpha)$ percent confidence interval for $\theta(\mathbf{x})$ is then given by

$$\hat{\theta}_{(\tilde{\alpha}_1(B+1))}(\mathbf{x}^*) \leq \hat{\theta}(\mathbf{x}) \leq \hat{\theta}_{((\tilde{\alpha}_2-1)(B+1))}(\mathbf{x}^*). \tag{S5.5}$$

S6 IBTS data set for cod and saithe

Table S6.1: 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 345 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	1600 (822) in year 2018 Q1 and 2330 (2163) 2017 Q3.

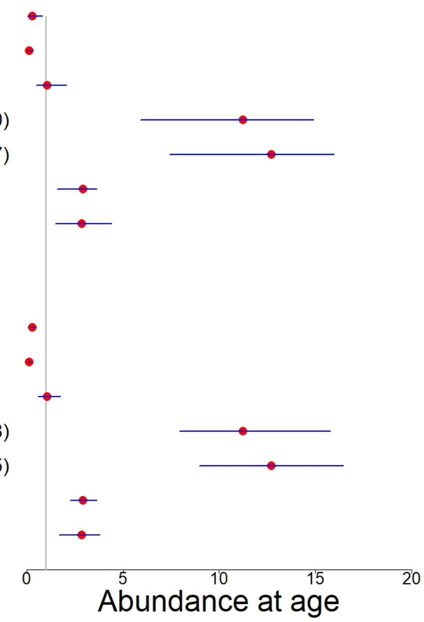
Table S6.2: Age and length data for saithe in year 2017 Q3 and cod in year 2018 Q1. Data collected in the first quarter (Q1) has no age 0 group but this is collected in quarter 3 (Q3) surveys.

	saithe in year 2017 Q3						cod in year 2018 Q1					
Age	Numbers aged	$\mathbf{L_{min}}$	$\mathbf{L_{max}}$	$\mathbf{L_{mean}}$	Sd($\mathbf{L}$)	CV($\mathbf{L}$)	Otoliths	$\mathbf{L_{min}}$	$\mathbf{L_{max}}$	$\mathbf{L_{mean}}$	Sd($\mathbf{L}$)	CV($\mathbf{L}$)
0	21	10	14	12.143	1.195	0.098						
1	26	23	32	27.654	2.297	0.083	149	11	30	18.407	3.693	0.201
2	65	27	47	38.077	3.337	0.088	814	17	53	33.180	6.290	0.190
3	531	34	56	42.041	3.785	0.090	222	30	81	50.654	10.185	0.202
4	767	35	73	48.261	4.521	0.094	189	43	92	64.479	8.399	0.130
5	334	46	78	56.876	6.105	0.107	102	54	96	76.627	9.594	0.125
6	159	50	91	66.025	7.137	0.108	84	54	100	80.871	9.456	0.117
7	127	57	93	73.976	7.163	0.097	28	58	110	84.086	11.308	0.134
8	69	63	94	77.725	7.010	0.090	4	80	94	85.500	6.455	0.075
9	18	64	97	85.333	7.499	0.088	5	66	96	83.400	11.305	0.061
10	22	84	107	92.364	5.803	0.063	1	87	87	-	-	-
11	5	79	102	92.800	9.311	0.100	1	106	106	-	-	-
12	7	91	109	99.429	6.554	0.066						
13	5	94	104	98.800	4.550	0.046						
14	1	108	108	108	-	-						
15	1	105	105	108	-	-						
16	4	93	106	100.250	5.439	0.054						
17	1	109	109	109	-	-						

S7 Estimates from DATRAS and Stratified bootstrap procedures

The bootstrap procedure proposed 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 underestimation of the uncertainty. The results (FigureS1) 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.

Age	Species	Mean	Median	SE	RSE%	Width CI-95	Estimate (CI_{95})
Datras							
0	saithe	0.431	0.368	0.3072	108.9	0.784	0.28 (0.01-0.80)
1	saithe	0.125	0.120	0.0668	53.8	0.261	0.12 (0.04-0.30)
2	saithe	1.126	1.049	0.4036	38.7	1.565	1.04 (0.47-2.03)
3	saithe	12.548	12.555	2.8061	25.0	8.997	11.24 (5.91-14.90)
4	saithe	14.196	14.227	2.7229	21.4	8.551	12.70 (7.42-15.97)
5	saithe	3.268	3.286	0.5770	19.8	2.054	2.91 (1.57-3.62)
6+	saithe	3.030	2.934	0.7421	26.1	2.941	2.84 (1.45-4.40)
Stratified							
0	saithe	0.276	0.280	0.0880	31.2	0.297	0.28 (0.14-0.43)
1	saithe	0.123	0.124	0.0527	42.5	0.173	0.12 (0.04-0.21)
2	saithe	1.005	0.997	0.2635	25.3	1.140	1.04 (0.57-1.71)
3	saithe	10.899	10.813	2.0905	18.6	7.845	11.24 (7.94-15.78)
4	saithe	12.424	12.425	1.8745	14.8	7.491	12.70 (8.95-16.45)
5	saithe	2.927	2.905	0.3528	12.1	1.384	2.91 (2.23-3.61)
6+	saithe	2.895	2.894	0.5621	19.8	2.138	2.84 (1.66-3.80)



(a) Datras and Stratified bootstrap Procedures

Figure S1: Comparison of estimated confidence intervals (CI_{95}) from DATRAS and stratified bootstrap procedures. The bias-corrected bootstrap method is used to give estimates for saithe in year 2017 Q3. Estimated indices of abundance (Estimate), and its standard error (SE), bootstrap mean (Mean), Median estimates, percentage relative standard error (RSE %) and width of confidence intervals are also given.