

<sub>1</sub> An analysis of the North Sea International Bottom Trawl Survey

<sub>2</sub> Data

<sub>3</sub>

<sub>4</sub> **Abstract**

<sub>5</sub> In this research we present non-parametric estimation procedures for calculating abundance at age  
<sub>6</sub> indices, and investigate the sensitivity of these estimates with respect to the number of otoliths collected  
<sub>7</sub> at sea. The procedures presented are applied to the North Sea International Bottom Trawls Survey data  
<sub>8</sub> for cod (*Gadus morhua*) and saithe (*Pollachius virens*). We demonstrate how much information would  
<sub>9</sub> be lost if the survey design was defined such that fewer otoliths were collected. Age length keys (ALKs)  
<sub>10</sub> are used to map lengths to age, and we use ALKs with and without the assumption of constant age length  
<sub>11</sub> structures over relatively large areas. All abundance at age indices are presented with variance estimates.

<sub>12</sub>

<sub>13</sub> **1 Introduction**

<sub>14</sub> Fish stock assessments are used by fishery managers for making management decisions regarding catch  
<sub>15</sub> quotas. The assessments provide fundamental information about the status of the stock, for instance,  
<sub>16</sub> whether the stock is increasing and support for increased levels of harvest should be given, or whether the  
<sub>17</sub> stock is decreasing and stricter control on harvest should be implemented. Associated with the parameters  
<sub>18</sub> used in fish stock assessment is their uncertainty, which should not be ignored when formulating management  
<sub>19</sub> policies (Walters and Ludwig, 1981; Ludwig and Walters, 1981; Berg et al., 2014). This uncertainty can arise  
<sub>20</sub> from many sources including natural variability, estimation procedures and lack of knowledge regarding the  
<sub>21</sub> parameter (Ehrhardt and Legault, 1997). The North Sea International Bottom Trawl Survey (IBTS) data,

22 coordinated by the International Council for the Exploration of the Sea (ICES), provides information on  
23 seasonal distribution of stocks and estimates of abundance indices and catch in numbers of fish per age-class  
24 without an assessment of the accuracy of these estimates. As stated by Ludwig and Walters (1981) it is  
25 relevant for managers to take into the uncertainty related to stock size when making management policies. The  
26 indices of abundance at age from IBTS are based on data obtained from a stratified semi-random sampling  
27 approach of trawl stations, and it is essential to account for the sampling approach so as to produce reliable  
28 variance estimates (Lehtonen and Pahkinen, 2004). If the sampling approach is ignored, the effect on the  
29 variance of the parameters could be substantial. In particular, the variance could be greatly inflated due  
30 to the clustering effect, which involves intra-cluster correlation of the variables (Aanes and Vølstad, 2015;  
31 Lehtonen and Pahkinen, 2004).

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

43 Currently, abundance indices from IBTS are reported in DATRAS (ICES, 2018c) using an age-length  
44 key (ALK) (Fridriksson, 1934) which is assumed to be constant over relatively large areas. In this research  
45 we propose two ALKs which accounts for spatial variation: i) a nonparametric haul based ALK, and ii) a  
46 spatial model-based ALK. These ALKs are described in section 2, and the results from the model based ALK  
47 gives a strong case for assuming variation in the ALK within RFAs. A spatial model based ALK (Berg and  
48 Kristensen, 2012; Berg et al., 2014) known as the NS-IBTS Delta-GAM index (ICES, 2016b) is currently

49 being used to calculate standardized age-based survey indices used in assessment for the North Sea stock.  
50 And as far as we are aware the variance estimates of parameters estimated from NS-IBTS Delta-GAM index  
51 are *only* utilized for assessment of Herring (*Clupea harengus*) in the North Sea.

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

### 59 **1.1 Overview of the North Sea International Bottom Trawl Survey**

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

66 The North Sea IBTS began with quarterly surveys providing information on seasonal distribution of  
67 stocks sampled, hydrography and the environment, which allows changes in fish stock to be monitored and  
68 abundance of all fish species to be determined. These quarterly surveys, however became difficult to sustain  
69 as countries experienced budget cuts making it impossible to maintain high levels of research vessel effort.

70 As such, in 1997 countries carried out a survey only twice a year; a first quarter survey (January-February)  
71 and a third quarter survey (August-September). The target species of IBTS fished from 1991-2018 includes  
72 standard pelagic species: Herring (*Clupea harengus*), Sprat (*Sprattus sprattus*) and Mackerel (*Scomber*  
73 *scombrus*); and standard roundfish species: Cod (*Gadus morhua*), Haddock (*Melanogrammus aeglefinus*),  
74 Saithe (*Pollachius virens*), Norway Pout (*Trisopterus esmarkii*) and Whiting (*Merlangius merlangus*).

75        Research vessels from seven nations in the first quarter (Q1) and six nations in the third quarter (Q3) are  
76        used for conducting surveys on all finfish species in the North Sea during January–February and July–August,  
77        respectively, between 1997–2018 (Table 4 in Web appendix A gives details of the different research vessels).  
78        The sampling frame is defined by the ICES index or roundfish areas (RFA) as shown in Figure 1 numbered 1  
79        to 10, and which we refer to as superstrata (Nottestad et al., 2015; Fuller, 2011). These roundfish areas were  
80        substratified into small strata defined by non-overlapping statistical rectangles of roughly  $30 \times 30$  nautical  
81        miles ( $1^\circ$  Longitude  $\times 0.5^\circ$  Latitude), and were convenient to use for North Sea IBTS as they were already  
82        being used for fisheries management purposes. Most statistical rectangles contain a number of possible tows  
83        that are deemed free of obstructions, and vessels are free to choose any position in the rectangles as long as  
84        the hauls are separated by at least 10 nautical miles within and between rectangles. However, all countries  
85        select tows based on a semi-random approach from databases of national safe tows or DATRAS or com-  
86        mercial fishing data, except Sweden who uses fixed stations and in some cases depth-stratified semi-random  
87        sampling design (ICES, 2018b), and England who also uses fixed stations and only conduct surveys during  
88        the third quarter. In some rectangles, sampling may be further stratified due to significant changes in seabed  
89        depth which may, in turn, cause variations in the fish population. In particular, the North Sea herring,  
90        saithe and sprat data are weighted by depth strata in the statistical rectangle (see Table 6 in appendix C).  
91        It is also a requirement that countries avoid clustering their stations between adjacent rectangles in order to  
92        reduce positive serial correlation, and thereby maximize survey precision. The latest major reallocation of  
93        rectangles occurred in 1991, but since then the survey has tried to keep at least one vessel in every subarea  
94        in which it had fished in the most recent years. Minor reallocation of rectangles between Norway, Scotland  
95        and Germany was done in 2013. Each rectangle was typically sampled twice by two different countries before  
96        1997, but after that target coverage of two trawl hauls per rectangle per survey was introduced because of  
97        national financial constraints (ICES, 2015). But in some rectangles in the Eastern English Channel, South-  
98        ern North Sea and Central North Sea intensified sampling is carried out.

99        The recommended standard trawling gear of the North Sea IBTS is the mulitpurpose chalut à Grande  
100      Ouverture Verticale (GOV) trawl (ICES, 2012), which has been used on all participating vessels since 1992,  
101      while different pelagic and bottom trawls suitable for fishing finfish species were used before 1992. Standard-

102 trawling protocols were adopted with a towing speed of 4 knots but depending on vessel performance,  
 103 tide and weather conditions the average towing speed can be at minimum 3.5 and maximum 4.5 knots. From  
 104 2000-2018 trawling is done during the daylight hours, which are considered 15 minutes before sunrise to 15  
 105 minutes after sunset (ICES, 2012). After each trawl the total catch of the different species is weighed on  
 106 board and biological parameters such as length for all fish species caught (to 0.1cm below for shellfish, to  
 107 0.5cm below for herring and sprat and to 1cm below for all other species) are collected. Where the numbers  
 108 of individuals are too large for all of them to be measured to obtain the length distribution, a representative  
 109 subsample of 100 fish is selected. Otoliths are collected on board from a small fraction of all the target  
 110 species from all RFAs (Figure 1) to retrieve age reading. Table 5 in Web appendix B gives the minimum  
 111 sampling levels of otoliths for the target species.

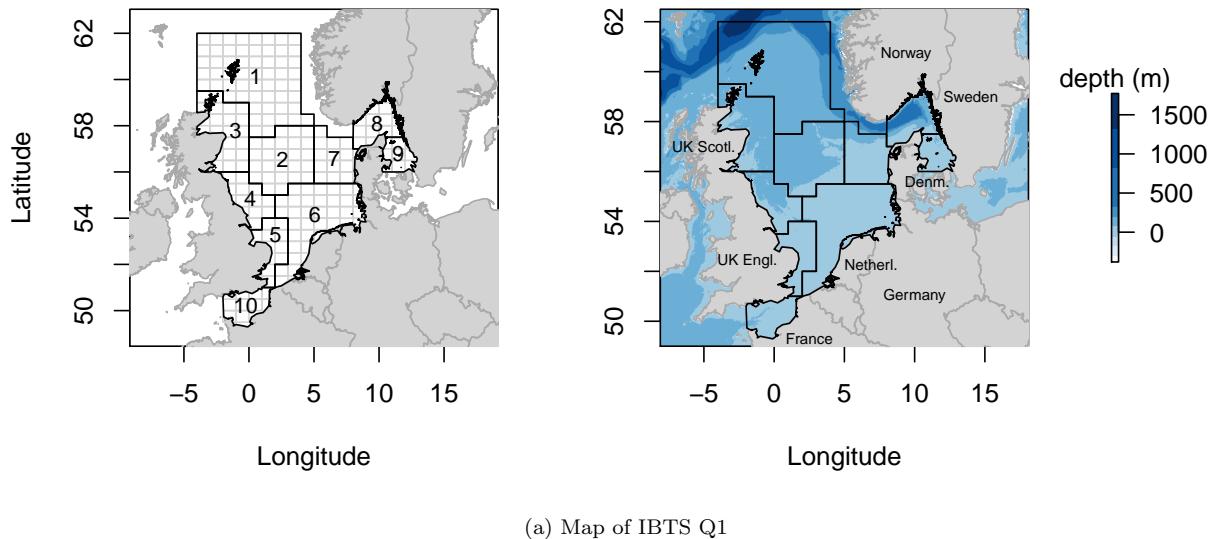


Figure 1: Standard roundfish areas used for roundfish since 1980, for all standard species since 1991 (left panel). Additional RFA 10 added in 2009. For example, the number 1 indicates ICES Index Area 1, and an ICES Statistical rectangle (ST) in IA 1 is 43F1. The map on the right panel shows norwegian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

112

## 2 METHODS

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

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

## 121 ***2.1 Catch per unit effort***

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

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

$$\text{CPUE}_{h,l} = \frac{n_{h,l}}{d_h}, \quad (2.1)$$

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

$$\text{CPUE}_{h,a} = \sum_{l \in \mathbf{L}} \text{CPUE}_{h,l} \times \text{ALK}_{a,l,h}, \quad (2.2)$$

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

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

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

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

134 where  $S_p$  is the set of all statistical rectangles in RFA  $p$ ,  $|S_p|$  is the number of statistical rectangles in RFA  
 135  $p$ , and  $\omega_s$  is a weight variable for each statistical rectangle. The weight variable  $\omega_s$  varies between species.  
 136 For some species  $\omega$  equals 1 (e.g. *Gadus morhua*) for all  $s$ , and for other species it is the proportion of the  
 137 statistical rectangle which has depth between 10 to 200 meters, for example *Pollachius virens* (see Table 6  
 138 in Web appendix C for weightings of statistical rectangles). The index for abundance at age in the whole  
 139 study area,  $\lambda_a$ , is further defined by

$$\lambda_a = \frac{\sum_{p \in \mathbf{P}} A_p m\text{CPUE}_{p,a}}{A_{\text{total}}}. \quad (2.5)$$

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

## 141 ***2.2 ALK estimators***

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

### 145 ***2.2.1 DATRAS ALK***

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

155    2.2.2 *Haul based ALK*

156    We define a haul dependent ALK by  $ALK^H$ . The  $ALK_{a,l,h}^H$  is defined as the average proportion of observed  
 157    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,  
 158    ages from the same length class in the haul close by is used (Web appendix D.2 describes the procedure in  
 159    detail).

160    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  $a = M, \dots, A$  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

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

where  $p_a(l, \mathbf{s})$  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

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

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

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

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

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

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

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

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

### 183 2.3 *Uncertainty estimation*

184 In this section we describe how the uncertainty of the CPUE estimates are calculated. We use nonparametric  
185 bootstrapping to quantify the uncertainty of the CPUEs. In nonparametric bootstrapping independent sam-  
186 ples of lengths and age are drawn with replacement from the original data and approximate 95% confidence  
187 intervals are obtained using **the percentile method** and bias-corrected percentile method (Carpenter and  
188 Bithell, 2000). Nonparametric resampling allows us to estimate the sampling distribution of the CPUE em-  
189 pirically without making assumptions concerning the data. The bias-Corrected method adjusts for the bias  
190 and skew of the sampling distribution of the data. This method assumes that there is a monotonic increasing  
191 function and the estimator  $\hat{\lambda}_a$  has a monotonic increasing function  $f()$  such that the transformed values

192  $f(\hat{\lambda}_a)$  are normally distributed with mean  $f(\rho) - z_0$  and standard deviation one, where  $z_0$  are the standard  
193 normal limits (Puth et al., 2015).

194 A bootstrap procedure for estimating the uncertainty of CPUEs in the North Sea is suggested in ICES  
195 (2013). In the rest of this paper, we refer to this procedure as DATRAS bootstrap procedure. The DATRAS  
196 procedure is divided into two parts; one part which samples CPUE per length (2.1), and another part  
197 which samples the ALK used in (2.2). The DATRAS bootstrap procedure is based on the assumption  
198 of homogeneous CPUE within RFAs. This assumption is likely to be wrong, and will typically cause an  
199 overestimation of the uncertainty. Therefore, we have included a bootstrap procedure, defined as the stratified  
200 bootstrap procedure, which instead assumes constant CPUE within each statistical rectangle.

201 *2.3.1 DATRAS and Stratified bootstrap procedure*

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

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

216 For simulating the DATRAS ALK we sample with replacement age observations within each RFA strati-  
217 fied with respect to length. If there is only one observed age from a given length class, we sample either that

218 age or, at random, an age of the closest length class with observed ages. For the haul based ALK, we use  
219 the observed ages in the sampled hauls when simulating the CPUE per length (**is this from the stratified**  
220 **bootstrap procedure? what about explanations for the model based ALK?** We also have codes  
221 **for stratified procedure where DATRAS ALK is used, should we mention that this is done IN-**  
222 **STEAD or AS WELL and include this results below in results section?** Jon Helge mentioned  
223 **that it's not plausible to show results of DATRAS bootstrap as it does not account for the**  
224 **survey design. Better off using our proposed stratified procedure with DATRAS ALK; include**  
225 **resulst for DATRAS using their sugested bootstrap procedure in appendix and reference in**  
226 **paper concerning higher variances).**

## 227 **2.4 Reducing sampling effort**

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

## 239 **3 Case studies**

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

243 for the collection of otoliths were introduced in the surveys. For instance, one otolith per length group is  
244 sampled for most target species (see Table 5 in Web appendix B for details of the sampling procedures for  
245 each target species), and this data is appropriate for the application of our proposed sample optimization  
246 procedure described in Section 2.4. The second is that IBTS included Age 0 in Q3 surveys, and since data  
247 for year 2018 Q3 is not yet available, the data for years 2017 Q3 and 2018 Q1 will be used in our analyses.  
248 Also, some species such as saithe that occupies the deeper waters in the northern part of the North Sea  
249 and in the Skagerrak and Kattegat, along the shelf edge (ICES, 2018a), the IBTS Q3 data is relevant for  
250 analyses compared with data from IBTS Q1 surveys, which do not adequately cover these areas where saithe  
251 is distributed (see Figure 1). Note that both IBTS Q1 and Q3 surveys do not adequately cover the whole  
252 stock distribution of saithe but the data collected is considered generally representative (ICES, 2016a).

253 In this research, the species of interest are cod and saithe. All samples are caught using the standard GOV  
254 gear described in Section 1.1. Table 1 gives a brief description of the data for year 2018 in the first quarter  
255 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  
256 quarter; and saithe as old as 18 years in the first quarter and 17 years in the third quarter. In our analyses  
257 we consider the age groups 1 to 6+ in Q1 and 0 to 6+ in Q3 for all ALK methods, where the last group  
258 consists of fish of age 6 or older. Saithe are typically older than cod but smaller in length, particularly in Q1.  
259 Catch rates are higher in the third quarter, 48% for cod and 164% for saithe, compared with the first quarter.

260

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*

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

265 Recall that the main assumption of DATRAS ALK is that the age-length compositions of species over  
 266 large areas are the same. To illustrate that this assumption may not be valid, we used the spatial ALK  
 267 model (2.7) to predict probabilities of age given length of a 40 cm long cod and a .....cm long saithe in the  
 268 North Sea (Figure ??). These plots provide strong evidence against a null hypothesis of no spatial effect in  
 269 the ALKs, as the likelihood of age given length changes in some areas. The plots also suggest that cod is  
 270 distributed in all areas of the North Sea (Figure ?? (a)), whereas saithe is more likely to inhabit areas in  
 271 the northern North Sea, specifically RFA 1 (Figure ?? (b)).

272 Figures 2 and 3 give estimates of indices of abundance for cod in years 2017 Q3 and 2018 Q1, and saithe  
273 in year 2017 Q3. Approximate 95% confidence intervals from the percentile and bias-corrected bootstrap  
274 methods for 200 bootstrap replication estimated from the three ALK methods.

275 **discuss more about confidence intervals**

276 Figures 2 and 3 show that the resulting indices of abundance turned out to be similar for all ALKs.  
277 However, the implication of not accounting for variability over wider areas is higher uncertainty in the esti-  
278 mates, as shown by the DATRAS ALK. We would expect higher uncertainty estimates for older fishes from  
279 our ALK methods compared with DATRAS as variability is much higher for this group due to small sample  
280 sizes. However, in Q3 for both species estimated uncertainty is from our ALKs is similar or smaller compared  
281 with DATRAS ALK for the plus group.....

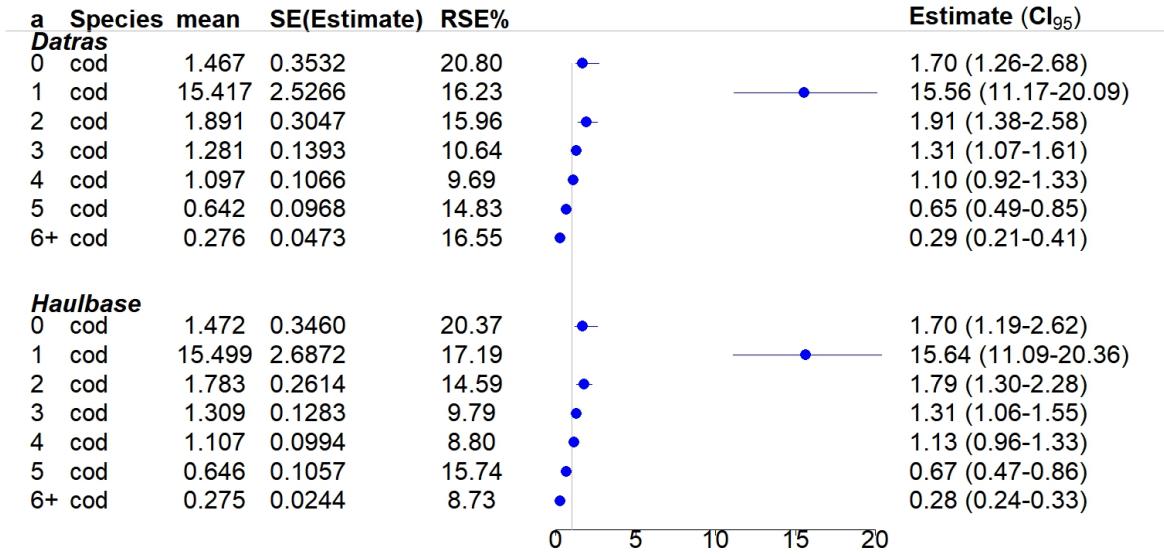
282

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

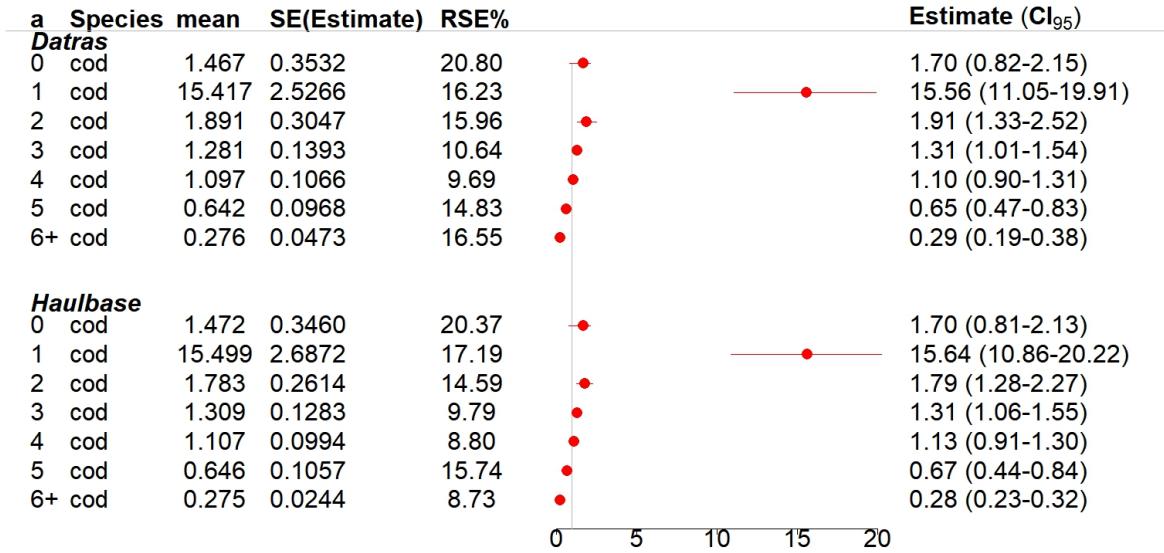
286

287 **need to generate this plot for appendix**

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

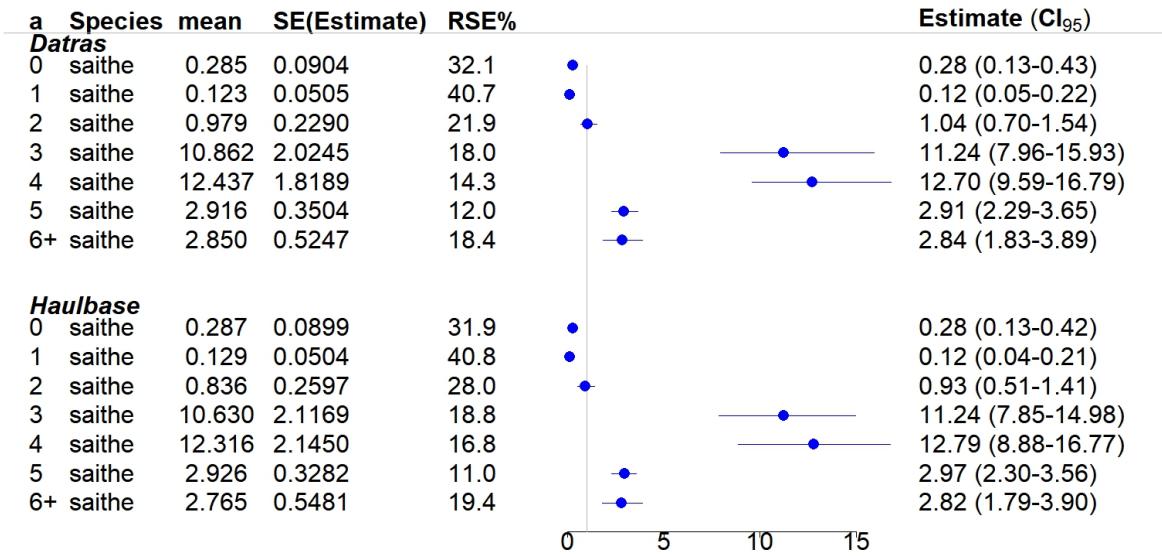


(a) Cod in years 2017 Q3 and 2018 Q1



(b) Percentile bootstrap method

Figure 2: Estimated confidence intervals (CI<sub>95</sub>) from bias-corrected bootstrap method for cod in years 2017 Q3 and 2018 Q1. Estimated indices of abundance (Estimate), and its standard error (SE(Estimate)), percentage relative standard error (RSE%), and bootstrap mean (mean) and median estimates are also given.



(a) Saithe in year 2017 Q3

Figure 3: Estimated confidence intervals (CI<sub>95</sub>) from bias-corrected bootstrap method for saithe in year 2017 Q3. Estimated indices of abundance (Estimate), and its standard error (SE(Estimate)), bootstrap mean (mean) and median estimates are also given.

292 • more data in Q3 particularly for larger older fishes compared with Q1 hence overestimation due to  
293 ignoring of fine-scale stratification rather than fewer samples, hence higher variance; 48% and 164%  
294 increases for cod and saithe, respectively?

295 • For all ALK methods the stratified bootstrap procedure is employed to estimate the uncertainty around  
296 indices of abundance. The estimated indices of abundance of cod are similar for all ALKs,  
297 but the the spatial ALKs generally gave higher estimates. Also, the spatial ALKs provide  
298 a better fit to the data in terms of precision. Uncertainty estimates (RSE) for older fishes  
299 ( $\geq 4$ ) are higher, as expected, for the spatial ALKs as fewer samples are generally collected  
300 for these, and given that spatial variation in the data is accounted for, variability would  
301 be higher.

302 • generate plots for year 2018 Q1 cod and 2017 Q3 saithe. Possibly show same age given length group  
303 for both species? change plots to our show information in them- see the structure directly from the raw  
304 data that they correspond to the colours in the figure; TMB has problems when few observations are  
305 available-issues with the joint covariance matrix

306 • have different plots...show fish of different ages (contours) corresponds to raw data, include points on  
307 the map of the different ages

308 • discuss differences in estimates from ALK methods (Table ?? - in terms of relative standard error  
309 estimates; overestimation and underestimation from DATRAS (bootstrap) method not knowing the  
310 strength in either direction) - Issues:

311 – Ignores fine scale stratification at the first stage, hence overestimation of the uncertainty  
312 – Ignores age-length data collected at the haul level, hence underestimates the uncertainty  
313 – Biases in both direction

314 • test for significant differences between estimates from ALK methods for age groups?

315 • As discussed in Section 2.2.1 and shown in the results in Table ?? the assumption of no difference  
316 in regional compositions of age-length structures is invalid and DATRAS ALK have introduced bias

317       in both direction, and the extent of this bias in either direction is unknown. Hence, this ALK is not  
318       appropriate to perform further analyses.

- 319       • codes do not run for saithe Q1 2018

320       • Saithe and haddock tend to have a northerly distribution. The abundance of fish predators is generally  
321       lower in the German bight area. Within the northern area, saithe is more abundant in the eastern  
322       areas.

323       • IBTS Q1 doesn't cover the distribution of saithe adequately. Saithe are found deeper than the survey  
324       extends and any fluctuations in abundance within the survey are not related to stock size, but due to  
325       movement up and down the slope at that time. Explanation is in the 2016 benchmark report for saithe,  
326       available on the ICES website

327       • The saithe assessment went through an ICES benchmark process in 2016 (ICES, 2016b). The scientific  
328       survey used in the assessment does not cover the whole stock distribution; however, it is considered  
329       generally representative. The number of observations (trawl stations) with saithe is low and the resulting  
330       survey index is uncertain. Commercial catch per unit effort information for French, German, and  
331       Norwegian trawlers was combined into a single index of biomass of fishable saithe. There are conflicting  
332       signals between the survey and fishable biomass index. The fraction of age 3 saithe migrating into the  
333       survey area (and the fishery) is low and varying between years with no obvious trend. Observations of  
334       saithe at age 3 are not suitable for predicting year-class strength. This means that assumed recruitment  
335       values are highly uncertain and a substantial portion (30%) of the advised wanted catch in 2017 is based  
336       on the recruitment assumptions for 2016 and 2017 (ICES, 2016a)

337 **3.2 Optimum sampling effort for North Sea Cod and Saithe**

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

- 347 • We can include a similar plot with 2017 (saithe) and 2018 (cod) data here to show that it's reasonable  
 348 to group at 2cm and 5 cm? Explain why 2 cm or 5 cm is chosen as illustrations?  
 349 • include (cm) on the x-axis to demonstrate the unit used for length; write out the word "probability" on  
 350 y- axis; remove bold titles on the plots; show real data plot alongside these plots

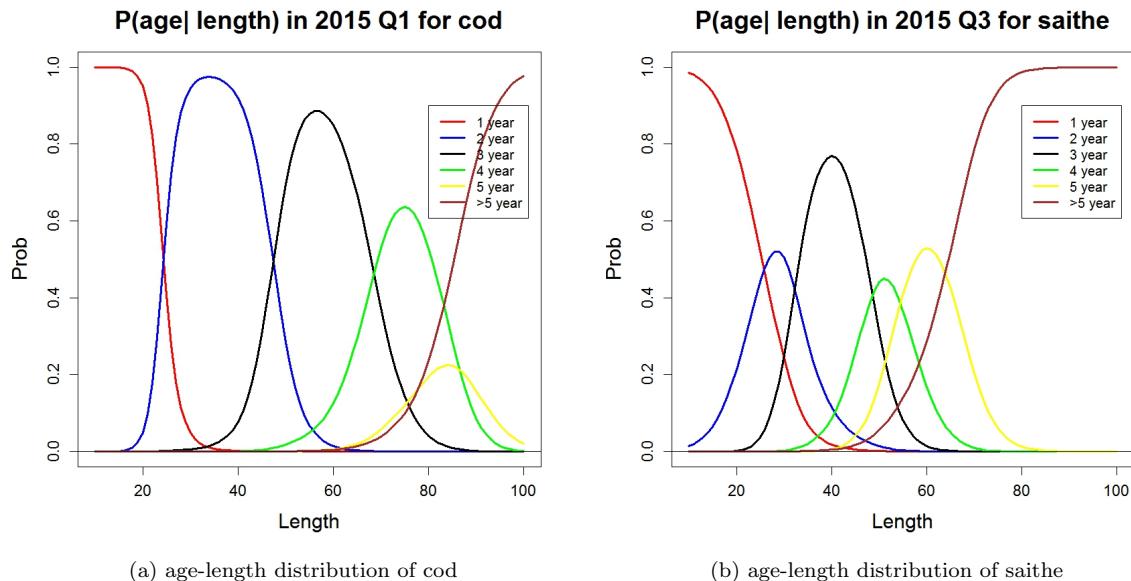


Figure 4: Predicted probabilities of age given length using the model described in (2.6) and (2.7) for cod (left panel) and saithe (right panel) in Q1 and Q3, respectively in year 2015.

351 In year 2018 Q1,  $\mathbf{x}$  otoliths for cod and  $\mathbf{y}$  otoliths for saithe were sampled in year 2017 Q3 (Table 1).

352 Sampling at random, one otolith per 2 cm length group removes  $\mathbf{x}$  percent of otoliths for cod and  $\mathbf{y}$  percent

353 of otoliths for saithe; while  $\mathbf{x}\%$  of otoliths for cod and  $\mathbf{y}\%$  of otoliths for saithe are removed when one otolith

354 per 5 cm length group is sampled. Table 3 gives estimates of abundance,  $\hat{\lambda}_a$ , its standard error,  $\widehat{SE}_{\hat{\lambda}_a}$  and

355 the upper bound (UP) and lower bound (LB) of the parameter estimates when all the otolith data is used

356 and when either one otolith is sampled per 2 cm or 5 cm length group for  $n = 100$  simulations and  $B = 200$

357 bootstrap runs. The coefficient of variation ( $Cv\%$ ) is also given.

358 Sampling one otolith per 2 cm or 5 cm length group provide .....

359 • *The time needed for estimating the model varies slightly between species and year. A laptop with*  
360 *processor intel(R) Core(TM) i5-6300 CPU @ 2,40 GHz, used e.g. approximately 2 minutes to estimate*  
361 *the parameters for cod in year 2018.*

362 • *discuss effects of this reduction on estimate and variance*

363 • *extract data for number or percentage of otoliths removed when 2 cm or 5 cm removal procedure is used*  
364 *and fill  $\mathbf{x}$  and  $\mathbf{y}$  in text with values*

365 • include no otoliths for each section —all data, reduced data by 2 cm etc

Table 2: Estimates of abundance indices of cod in year 2017 Q3.

	All data estimate	Removed data estimate	Lower	Upper	SE
1	1.70	1.70	1.69	1.72	0.01
2	15.64	15.60	15.55	15.71	0.05
3	1.79	1.79	1.73	1.83	0.03
4	1.31	1.29	1.22	1.36	0.05
5	1.13	1.12	1.09	1.17	0.04
6	0.67	0.69	0.64	0.75	0.07
7	0.28	0.28	0.26	0.30	0.06

Table 3: Estimates of abundance indices and standard error estimates for the North Sea cod and saithe species when all the otolith data is used and when one otolith per 2 cm or 5 cm length group is used. The lower bounds (LB) and upper bounds (UB) of approximate 95% confidence intervals are given.

Species	$a$	All data			Reduced data by 2 cm				Reduced data by 5 cm				
		No. otoliths	$\hat{\lambda}_a$	$\widehat{SE}_{\hat{\lambda}_a}$	$\hat{\lambda}_a$		$\widehat{SE}_{\hat{\lambda}_a}$		No. otoliths	LB	UB	$\hat{\lambda}_a$	$\widehat{SE}_{\hat{\lambda}_a}$
					No. otoliths	LB	UB	LB					
<b>cod</b>													
2017 Q3	0	102	0	0	0			0	0				
	1	1332	0.764	0.26		0.60	0.24			0.70	0.36		
	2	269	21.989	6.76		22.21	4.15			22.11	4.28		
	3	207	11.285	2.19		10.58	1.20			10.99	1.77		
	4	174	3.265	0.71		3.67	1.28			3.50	0.87		
	5	102	1.147	0.34		1.27	0.42			1.20	0.48		
	6+	50	1.276	0.38		1.40	0.70			1.21	0.42		
2018 Q1	1	144	0.764	0.26		0.60	0.24			0.70	0.36		
	2	767	21.989	6.76		22.21	4.15			22.11	4.28		
	3	214	11.285	2.19		10.58	1.20			10.99	1.77		
	4	168	3.265	0.71		3.67	1.28			3.50	0.87		
	5	100	1.147	0.34		1.27	0.42			1.20	0.48		
	6+	118	1.276	0.38		1.40	0.70			1.21	0.42		
<b>saithe</b>													
2017 Q3	0	21	0	0		0	0	0	0				
	1	26	0.764	0.26		0.60	0.24			0.70	0.36		
	2	65	21.989	6.76		22.21	4.15			22.11	4.28		
	3	494	11.285	2.19		10.58	1.20			10.99	1.77		
	4	734	3.265	0.71		3.67	1.28			3.50	0.87		
	5	334	1.147	0.34		1.27	0.42			1.20	0.48		
	6+	418	1.276	0.38		1.40	0.70			1.21	0.42		
2018 Q1	1	11	0.764	0.26		0.60	0.24			0.70	0.36		
	2	9	21.989	6.76		22.21	4.15			22.11	4.28		
	3	36	11.285	2.19		10.58	1.20			10.99	1.77		
	4	210	3.265	0.71		3.67	1.28			3.50	0.87		
	5	262	1.147	0.34		1.27	0.42			1.20	0.48		
	6+	264	1.276	0.38		1.40	0.70			1.21	0.42		

## 4 DISCUSSION

367 In this research we have determined minimum sampling efforts of otoliths for target species of the North  
 368 Sea International Bottom Trawl Survey. This was achieved by testing sampling procedures that mimic  
 369 the real data collection procedure but with a reduced number of otoliths. Two sampling procedures, 1)  
 370 sampling at random, one otolith per 2 cm and 2) sampling at random, one otolith per 5 cm length group,  
 371 were tested and the effect on estimated abundance indices and their variance were investigated. Abundance  
 372 indices were estimated using age-length keys (ALKs). The database for trawl surveys (DATRAS) manned  
 373 by ICES includes an ALK that uses the raw proportions of age given length assuming constant age-length  
 374 compositions over relatively large areas. We have developed two spatial ALK methods to estimate abundance  
 375 indices and their variance that accounts for spatial variation in the data: 1) a haul based ALK that produces  
 376 an ALK for each trawl haul, and which uses the raw proportions of age given length, and 2) a spatial ALK  
 377 model that uses logits for modelling the age distribution in catch data from the length-stratified subsamples.  
 378 Several studies have used spatial ALK modelling for estimating abundance indices of the North Sea stocks  
 379 used in assessments (Berg and Kristensen, 2012; Berg et al., 2014; Gerritsen et al., 2006). These studies used  
 380 continuous ratio logits with General Linear Model (GLM) or General Additive Models (GAMs) to model the  
 381 spatial effects and found regional effects..... We propose to use Gaussian Random Field Theory to model  
 382 the spatial effects as a smooth surface.....

383     • *discuss positives of using GRT to model spatial effects: what problems are eliminated when using this  
 384       in terms of missing data*

385     • *compare the effects of our method with GAMs (Berg and Kristensen, 2012) and (Berg et al., 2014) and  
 386       the NS-IBTS Delta-GAM index for estimating standardized age-based indices and the species theses are  
 387       used for to include in assessment*

388     • *what does our model allows in terms of the age groups (smaller or higher age groups (6+) possible with  
 389       our model); covariates such as haul effect (included as a random effect)*

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

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

- 393 • discuss sampling procedure: limitation and advantages; and possibly more advanced selection proce-  
394 dures?

395 • new approach adopted in surveys from 2018

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

400 • how does changes in survey design or other factors affect changes in catch rates? If so are these changes  
401 significant?

402 **5 General comments**

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

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

405 • Decide on year of data for case studies

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

407 • Decide on calling the survey "The North Sea IBTS or IBTS"

408 • Decide on writing "Cod or cod, and Saithe or saithe"

409 • Decide on a title for the paper

410 • what are issues with including haul effect in model based ALK? (Olav)

411

## References

- 412 Aanes, S. and Vølstad, J. H. (2015). Efficient statistical estimators and sampling strategies for estimating  
 413 the age composition of fish. *Canadian journal of fisheries and aquatic sciences*, 72(6):938–953.
- 414 Berg, C. W. and Kristensen, K. (2012). Spatial age-length key modelling using continuation ratio logits.  
 415 *Fisheries Research*, 129:119–126.
- 416 Berg, C. W., Nielsen, A., and Kristensen, K. (2014). Evaluation of alternative age-based methods for  
 417 estimating relative abundance from survey data in relation to assessment models. *Fisheries Research*,  
 418 151:91–99.
- 419 Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide  
 420 for medical statisticians. *Statistics in medicine*, 19(9):1141–1164.
- 421 Ehrhardt, N. M. and Legault, C. M. (1997). The role of uncertainty in fish stock assessment and management:  
 422 a case study of the spanish mackerel, scomberomorus maculatus, in the us gulf of mexico. *Fisheries  
 423 research*, 29(2):145–158.
- 424 Fridriksson, A. (1934). On the calculation of age-distribution within a stock of cod by means of relatively few  
 425 age-determinations as a key to measurements on a large scale. *Rapports Et Proces-Verbaux Des Reunions,  
 426 Conseil International Pour l'Exploration De La Mer*, 86:1–5.
- 427 Fuller, W. A. (2011). *Sampling statistics*, volume 560. John Wiley & Sons.
- 428 Gerritsen, H. D., McGrath, D., and Lordan, C. (2006). A simple method for comparing age-length keys  
 429 reveals significant regional differences within a single stock of haddock (melanogrammus aeglefinus). *ICES  
 430 Journal of Marine Science*, 63(6):1096–1100.
- 431 Hirst, D., Storvik, G., Rognebakke, H., Aldrin, M., Aanes, S., and Vølstad, J. H. (2012). A bayesian  
 432 modelling framework for the estimation of catch-at-age of commercially harvested fish species. *Canadian  
 433 journal of fisheries and aquatic sciences*, 69(12):2064–2076.
- 434 ICES (2012). Manual for the international bottom trawl surveys, revision viii. series of ices survey protocols.  
 435 *International Council for the Exploration of the Sea*, SISP 1-IBTS VIII.

- 436 ICES (2013). Ns-ibts indices calculation procedure. datras procedure document. *International Council for*  
437 *the Exploration of the Sea*, 1.1 NS-IBST indices-2013.
- 438 ICES (2015). Manual for the international bottom trawl surveys, revision ix. series of ices survey protocols.  
439 *International Council for the Exploration of the Sea*, SISP 10-IBTS IX.
- 440 ICES (2016a). Ices advice on fishing opportunities, catch, and effort greater north sea and celtic seas  
441 ecoregions. *International Council for the Exploration of the Sea*, Book 6.
- 442 ICES (2016b). Report of the benchmark workshop on north sea stocks (wknsea), 14–18 march, 2016.  
443 *International Council for the Exploration of the Sea*, ICES CM 2016/ACOM:37.
- 444 ICES (2018a). Ices fishmap: Atlas of the north sea fish, including fact sheets of key species and distribution  
445 maps. *International Council for the Exploration of the Sea (ICES FishMap website, 2018*, 1.
- 446 ICES (2018b). Manual for the international bottom trawl surveys, revision x. series of ices survey protocols.  
447 *International Council for the Exploration of the Sea*, SISP X-IBTS X.
- 448 ICES, D. (2018c). Datras.
- 449 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. (2015). Tmb: automatic differentiation  
450 and laplace approximation. *arXiv preprint arXiv:1509.00660*.
- 451 Lehtonen, R. and Pahkinen, E. (2004). *Practical methods for design and analysis of complex surveys*. John  
452 Wiley & Sons.
- 453 Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian  
454 Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical  
455 Society: Series B (Statistical Methodology)*, 73(4):423–498.
- 456 Ludwig, D. and Walters, C. J. (1981). Measurement errors and uncertainty in parameter estimates for stock  
457 and recruitment. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(6):711–720.
- 458 Nottestad, L., Utne, K. R., 'Oskarsson, G. J., J'onsson, S. T., Jacobsen, J. A., Tangen, O., Anthonypillai,  
459 V., Aanes, S., Vølstad, J. H., Bernasconi, M., et al. (2015). Quantifying changes in abundance, biomass,

- 460 and spatial distribution of northeast atlantic mackerel (*scomber scombrus*) in the nordic seas from 2007  
461 to 2014. *ICES Journal of Marine Science*, 73(2):359–373.
- 462 Puth, M.-T., Neuhäuser, M., and Ruxton, G. D. (2015). On the variety of methods for calculating confidence  
463 intervals by bootstrapping. *Journal of Animal Ecology*, 84(4):892–897.
- 464 Rindorf, A. and Lewy, P. (2001). Analyses of length and age distributions using continuation-ratio logits.  
465 *Canadian Journal of Fisheries and Aquatic Sciences*, 58(6):1141–1152.
- 466 Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models  
467 by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B  
(Statistical Methodology)*, 71(2):319–392.
- 468 Walters, C. J. and Ludwig, D. (1981). Effects of measurement errors on the assessment of stock-recruitment  
469 relationships. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(6):704–710.
- 470 Wood, S. (2015). Package mgcv. *R package version*, pages 1–7.
- 471 Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

473

## A Areas fished by different countries in the North Sea IBTS

474 Typically, two different countries fish each rectangle so that at least two trawl hauls are made per rectangle,  
 475 but intensified sampling is carried out in the following areas: at least 3 hauls per rectangle are taken in  
 476 statistical rectangles 31F1, 31F2, 32F1, 33F4, 34F2, 34F3, 34F4, 35F3, 35F4; while six or more hauls per  
 477 rectangle are taken in statistical rectangles 30F1, 32F2, 32F3, 33F2, 33F3 (ICES 1999). The Skagerrak  
 478 and Kattegat is fished solely by Sweden, who sample more than once in every rectangle while the west of  
 479 Shetland (in Q1 and Q3) and inshore areas (Q3) is fished solely by Scotland. The edge of the Norwegian  
 480 Trench is fished solely by Norway, but inshore areas near Denmark is fished by Denmark. The southern  
 481 North Sea is fished by Denmark, Germany and England. France, typically, is the only country that surveys  
 482 the western English Channel. Areas are surveyed by a single country because of the large proportion of  
 483 untrawalable area (and subsequent gear damage issues experienced by other nations) for efficient logistical  
 484 purposes. Table 4 gives the countries and research vessels participating the North Sea IBTS.

Table 4: Survey countries, vessel name, and period research vessels participating in first quarter (Q1) and third quarter (Q3) during 1997-2017.

Country	First Quarter (Q1)		Third Quarter (Q3)	
	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

485

486

## B Otolith sampling per fish species

487 From 1991-2017, most countries conducted quota sampling of otoliths per length group in a RFA. But from  
 488 2013 Norway has been sampling one otolith per length class from each trawl haul (to 0.1cm below for shellfish,

489 to 0.5cm below for herring and sprat and to 1cm below for all other species). From the first quarter in 2018  
 490 all countries are required to sample one otolith per length class per trawl haul. Table 5 gives the minimum  
 491 sampling levels of otoliths for the target species. However, for the smallest size groups, that presumably  
 492 contain only one age group, the number of otoliths per length class may be reduced, and more otoliths per  
 493 length are required for the larger length classes.

Table 5: Minimum sampling levels of otoliths by species for RFA or per trawl haul.

Period	Species	Minimum sampling levels of otoliths per length class
<b>1991-2017</b>		<b>Number of otoliths per length class in a RFA</b>
	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 $\geq 11.0$ 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).
<b>2018</b>		<b>Number of otoliths per length class per trawl haul</b>
	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 $\geq 11.0$ 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 6: Weights used for *Pollachius virens* in equation (2.3).

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

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

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

504 **D.1 DATRAS ALK Borrowing Approach**

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

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

- 513 2. If length class (CPUE) > maximum length (ALK) age is set to the plus group.

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

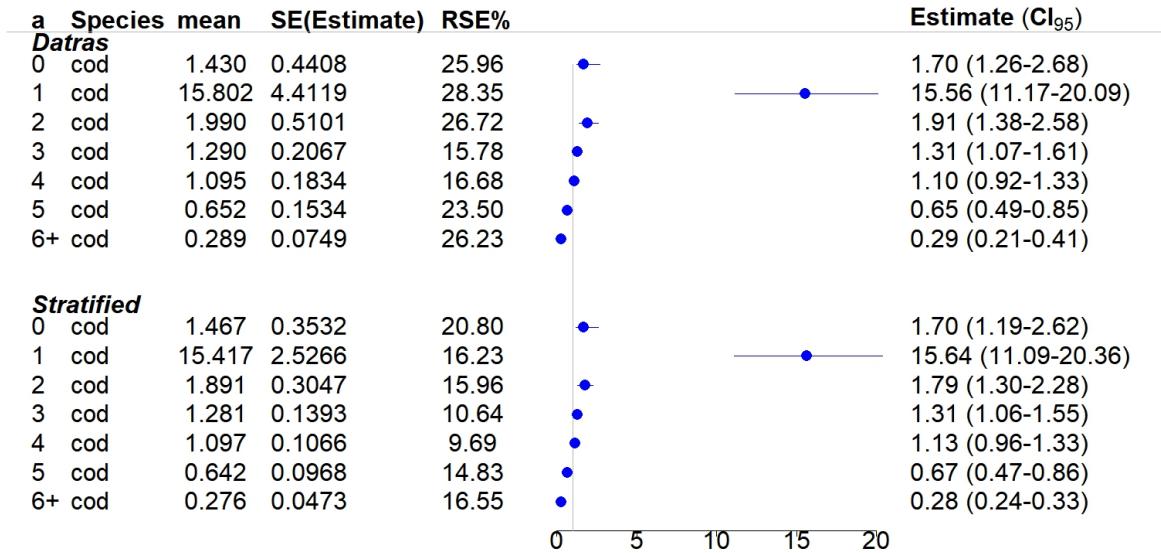
516 **D.2 Haul-based ALK Borrowing Approach**

517 The second is an a haul dependent ALK estimator which we propose, and is denoted by  $\text{ALK}^H$ . Since the age-  
518 length composition of fish may be space-variant, that is, there may be variation in age-length compositions  
519 between trawl stations within a superstrata, the spatial dependence of the age-length composition must be  
520 accounted for to produce reliable estimates of the CPUE per age estimates. If this spatial dependence is  
521 ignored not only will estimates of abundance be biased but the impact on the variance may be substantial.  
522 So for each trawl haul an  $\text{ALK}^H$  is produced. *Since there are few or none observations of ages for each*  
523 *length class in a trawl haul, length classes are therefore pooled in increasing order such that there are five*

524 length classes in each pooled length group. To replace missing values for the age distribution in the pooled  
 525 length groups the method of "borrowing" ages from length groups in trawl hauls closest in air distance within  
 526 the RFA is used. If there are no observed ages in the pooled length group in the RFA, missing values for the  
 527 age distribution are replaced following the procedure outlined in the DATRAS ALK procedure (D.1) in step  
 528 1.

## 529 E Estimates from DATRAS and Stratified bootstrap procedures

530 The naive bootstrap procedure prosed by DATRAS lacks the potential to account for the spatial variation  
 531 in the data. The DATRAS bootstrap procedure ignores the fine-scale stratification in the sampling process,  
 532 leading to an overestimation of the uncertainty; and ignores the age-length data collected at the haul level,  
 533 resulting in an underestiamtion of the uncertainty. The results (Table 5) shows an overestimation of the  
 534 uncertainty for all age groups, suggesting that it is relevant to account for the fine-scale stratification when  
 535 resampling the data.



(a) Datras and Stratified bootstrap Procedures

Figure 5: Comparison of estimated confidence intervals (CI<sub>95</sub>) 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.

536      **Table 7** below will not be included in paper. For our purposes to look at estimates of saithe  
537      with and without weights and, when the code for weights was "incorrect"

Table 7: 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 (UB) 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
<b>saithe-weights</b>													
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				(%)
<b>saithe-weights old</b>													
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				(%)
<b>saithe-no-weights</b>													
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				(%)