

₁ An analysis of the North Sea International Bottom Trawl Survey

₂ Data

₃

₄ **Abstract**

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

₁₂

₁₃ **1 Introduction**

₁₄ Fish stock assessments are used by fishery managers for making management decisions regarding catch
₁₅ quotas. The assessments provide fundamental information about the status of the stock, for instance,
₁₆ whether the stock is increasing and support for increased levels of harvest should be given, or whether the
₁₇ stock is decreasing and stricter control on harvest should be implemented. Associated with the parameters
₁₈ used in fish stock assessment is their uncertainty, which should not be ignored when formulating management
₁₉ policies (Walters and Ludwig, 1981; Ludwig and Walters, 1981; Berg et al., 2014). This uncertainty can arise
₂₀ from many sources including natural variability, estimation procedures and lack of knowledge regarding the
₂₁ 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 2 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 30×30 nautical
81 miles (1° 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 4 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 ized 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 3 in Web appendix B gives the minimum
 111 sampling levels of otoliths for the target species.

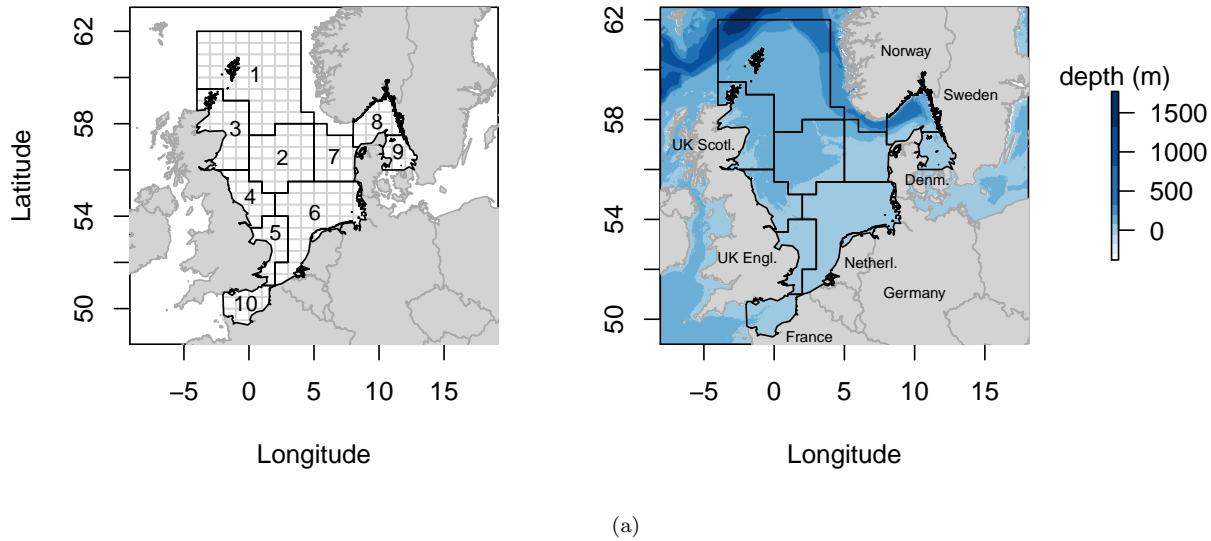


Figure 1: Standard roundfish areas used for roundfish since 1980, for all standard species since 1991 (left panel). Additional RFA 10 added in 2009. For example, the number 1 indicates ICES Index Area 1, and an ICES Statistical rectangle (ST) in IA 1 is 43F1. The map on the right panel shows norwegian 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 ω_s is a weight variable for each statistical rectangle. The weight variable ω_s varies between species.
 136 For some species ω 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 4
 138 in Web appendix C for weightings of statistical rectangles). The index for abundance at age in the whole
 139 study area, λ_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 ALK^D denote the DATRAS ALK. The 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*

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

165 Let the response variable of the age group of a fish be $a = M, \dots, A$ where M is the youngest age, and A
 166 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
 167 l caught at location \mathbf{s} . As in Berg and Kristensen (2012) we use a continuous ratio model for the spatial age
 168 given length model. However, in our application we assume for each species we know a length l_m such that
 169 all fish above length l_m are above age M , and all fish with length below l_m are of age below A . By including
 170 such a variable we reduce the number of parameters in the model by removing one linear predictor. Define
 171 the continuous ratio we are modelling as

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

172 where $p_a(l, s)$ is the probability of a fish with length l at location s to be of age a . Note that either $p_A(l, s)$
 173 or $p_M(l, s)$ is known to be equal to zero, and the other is selected such that $\sum_a p_a = 1$. We further assume
 174 the logit link

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

175 Here $f_a(l)$ is a continuous function of length and γ is a mean zero Gaussian spatial random field with Matérn
 176 covariance function. The spatial random field is intended to capture any spatial variation in the ALK.

177 The continuous function $f_a(l)$ in (2.7) is modelled with usage of P-splines (Wood, 2017), and these
 178 spline regression coefficients are included as a Gaussian random effect. The precision matrix for the spline

179 regression coefficients is constructed such that the variability (or wryggliness) in the spline is penalized, see
180 Wood (2017, page 239) for details. The R package mgcv (Wood, 2015) is used for extracting the precision
181 matrix needed for the spline regression coefficients.

182 We assume that the spatially Gaussian random field in (2.7), γ , follows a stationary Matérn covariance
183 structure:

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

184 where σ_γ^2 is the marginal variance, $\|\cdot\|$ is the Euclidean distance measure in kilometres, ν is a smoothing
185 parameter, κ_γ is a spatial scale parameter and $K_\nu(\cdot)$ is the modified Bessel function of the second kind with
186 $\nu = 1$. The spatial field is estimated with the stochastic partial differential equation (SPDE) procedure
187 described in Lindgren et al. (2011). The main concept behind the SPDE procedure is that the precision
188 matrix of a spatial field with Matérn covariance function can be approximated by a sparse matrix on a grid
189 covering the area of interest. Such a grid and sparse precision matrix are constructed with use of the R-INLA
190 package (Rue et al., 2009).

191 The species specific constant l_m is selected as the mid point between the shortest fish of age A and the
192 longest fish of age M in the corresponding year and quarter. As a sensitivity analysis of this constant we
193 adjusted it up and down 5 cm for cod in year 2018. The point estimate of the mCPUEs then changed in the
194 forth decimal, which is negligible.

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

200 2.3 Uncertainty estimation

201 In this section we describe how the uncertainty of the CPUE estimates are calculated. We use nonparametric
202 bootstrapping to quantify the uncertainty of the CPUEs. In nonparametric bootstrapping independent sam-
203 ples of lengths and age are drawn with replacement from the original data and approximate 95% confidence

204 intervals are obtained using bias-corrected percentile method (Carpenter and Bithell, 2000). Nonparametric
205 resampling allows us to estimate the sampling distribution of the CPUE empirically without making as-
206 sumptions concerning the data. The bias-Corrected method adjusts for the bias and skew of the sampling
207 distribution of the data. This method assumes that there is a monotonic increasing function and the estimator
208 $\hat{\lambda}_a$ has a monotonic increasing function $f()$ such that the transformed values $f(\hat{\lambda}_a)$ are normally distributed
209 with mean $f(\rho) - z_0$ and standard deviation one, where z_0 are the standard normal limits (Puth et al., 2015).

210 A bootstrap procedure for estimating the uncertainty of CPUEs in the North Sea is suggested in ICES
211 (2013a). In the rest of this research, we refer to this procedure as DATRAS bootstrap procedure. The
212 DATRAS procedure is divided into two parts; one part which samples CPUE per length (2.1), and another
213 part which samples the ALK used in (2.2). The DATRAS bootstrap procedure is based on the assumption
214 of homogeneous CPUE within RFAs. This assumption is likely to be wrong, and will typically cause an
215 overestimation of the uncertainty. Therefore, we have included a bootstrap procedure, defined as the stratified
216 bootstrap procedure, which instead assumes constant CPUE within each statistical rectangle.

217 *2.3.1 DATRAS and Stratified bootstrap procedure*

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

228 The Stratified bootstrap procedure is a modification of the DATRAS bootstrap procedure. Rather than
229 sampling hauls from the whole RFA, we sample the N_s trawl hauls from the list of hauls within the same

230 statistical rectangle. If there is only one trawl haul within a statistical rectangle, we sample either that haul
231 or the closest haul.

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

243 **2.4 Reducing sampling effort**

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

255

3 Case studies

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

269 In this research, the species of interest are cod and saithe. All samples are caught using the standard GOV
270 gear described in Section 1.1. Table 1 gives a brief description of the data for year 2018 in the first quarter
271 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
272 quarter; and saithe as old as 18 years in the first quarter and 17 years in the third quarter. In our analyses
273 we consider the age groups 1 to 6+ in Q1 and 0 to 6+ in Q3 for all ALK methods, where the last group
274 consists of fish of age 6 or older. Saithe are typically older than cod but smaller in length, particularly in Q1.
275 Catch rates are higher in the third quarter, 48% for cod and 164% for saithe, compared with the first quarter.

276

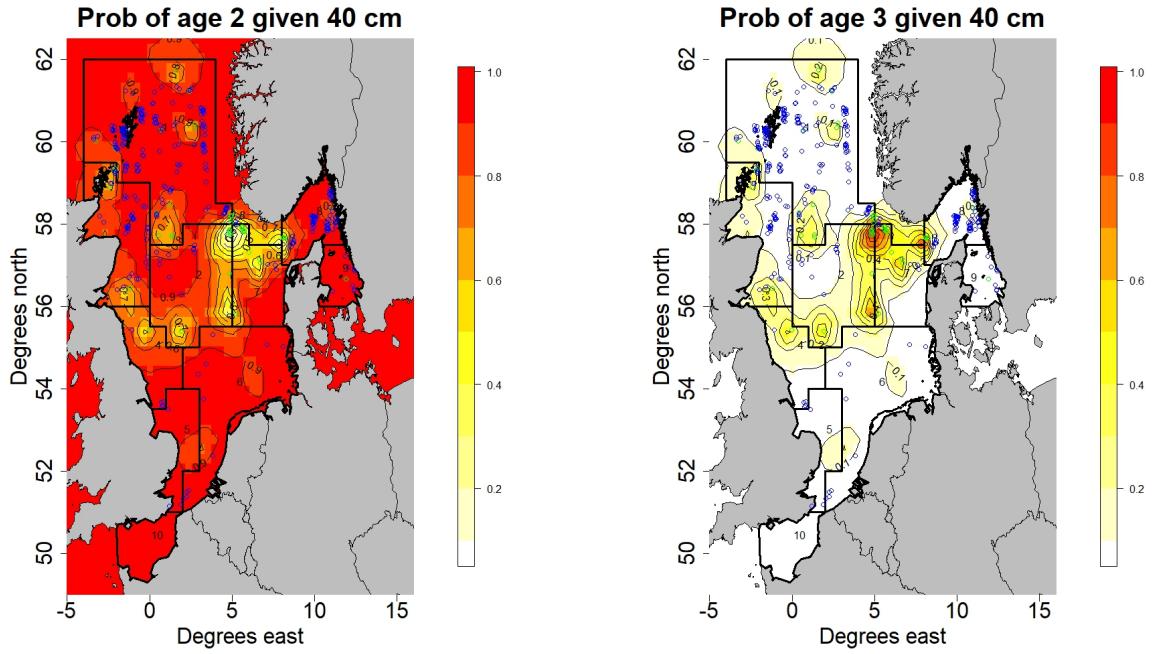
277 **check that number of otoliths are correct, removal simulation suggests 2163 otoliths for**
278 **saithe and 1600 for cod**

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.

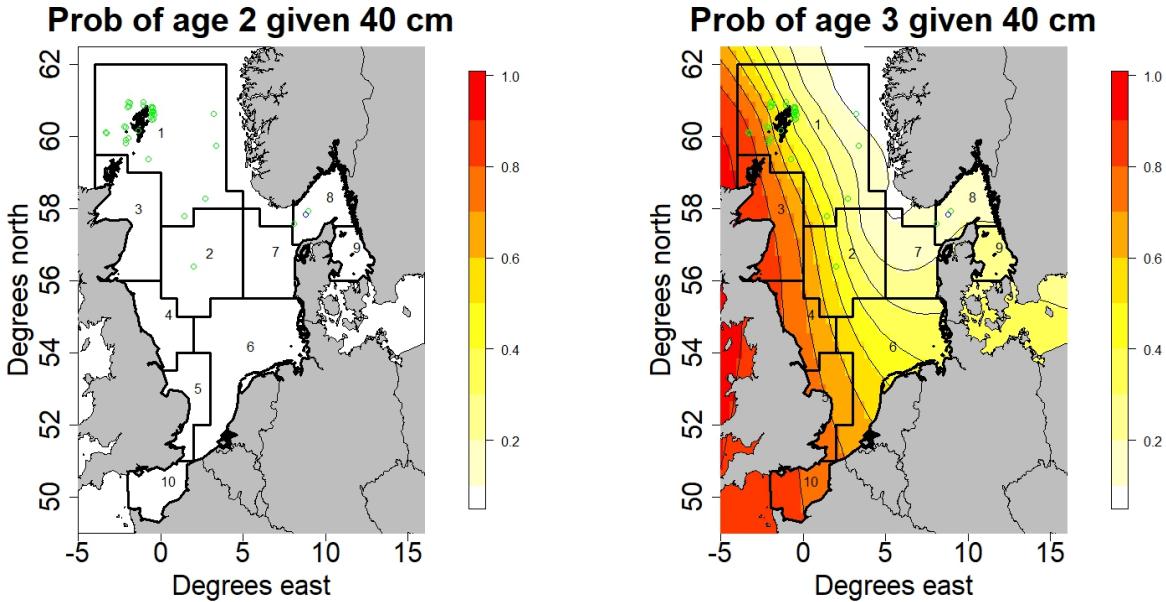
Data	Description
Trawl hauls	Total of 372 trawl hauls in year 2018 Q1; 238 (83) with length and 230 (81) with age information. In 2017 Q3, a total of 349 trawl hauls were taken; 238 (129) with length and 237 (128) with age information.
Age	The age varied between 1 (1) to 12 (18) years in year 2018 Q1 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 .

279 *3.1 Estimated indices of abundance and variability for cod and saithe*

280 In this section we apply the three ALK methods given in section 2.2 for abundance estimation, and the
 281 bias-corrected bootstrap method, given in Section 2.3.1 for estimating variability of estimated indices of
 282 abundance. The main assumption of DATRAS ALK is that the age-length compositions of species over
 283 large areas are the same. To illustrate that this assumption may not be valid, we used the spatial ALK
 284 model (2.7) to predict probabilities of age given length of a 40 cm long cod and a 40 cm long saithe in the
 285 North Sea (Figure 2). These plots provide strong evidence against a null hypothesis of no spatial effect in
 286 the ALKs, as the likelihood of age given length changes in some areas. Figure 2 (a) shows that the eastern
 287 North Sea in RFAs 7 and 8 (the regions in yellow) is one of the areas where a 40 cm cod is more likely to
 288 be age 3. A saithe of 40 cm is more likely to be 3 years or older. The plots also show that cod is distributed
 289 in all areas of the North Sea, except RFA 10 (Figure 2 (a)), whereas saithe is more likely to inhabit deeper
 290 areas in the northern North Sea, specifically RFA 1, and Skagerrak in RFA 8 (Figure 2 (b) and Figure 1,
 291 right panel).



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



(b) Probability plot of 40 cm saithe in year 2018 Q1

Figure 2: Predicted probabilities of age given length using model (2.6) and (2.7) for the year 2018 Q1. Graph (a) gives probabilities of predicted age of a 40 cm long cod, and graph (b) gives probabilities of predicted age of a 40 cm saithe in RFAs 1 to 10 in the North Sea. The small coloured circles (\circ , blue or green) are the trawl hauls with cod or saithe data.

292 Figures 3 gives estimates of indices of abundance for cod in years 2018 Q1 and for saithe in year 2017
293 Q3. Approximate 95% confidence intervals from the bias-corrected bootstrap method for 200 bootstrap
294 replication are estimated from the three ALK methods. The stratified procedure described in 2.3.1 is used in
295 the sampling process to estimate bootstrap confidence intervals. Figures 3 shows that the resulting indices
296 of abundance for cod and saithe turned out to be similar for all ALKs. IBTS is a complex multistage survey
297 design, and since the ALKs are estimated from cluster-correlated data the resulting effective sample for
298 estimating age-composition of fish would be lower than the number of fish measured (ICES, 2013b). Hence,
299 the ALKs are subject to large sampling errors. For example, the estimated percentage relative standard
300 errors from the spatial ALKs for the plus group (6+) for cod are > 25%, suggesting high sampling error in
301 the ALKs. Also, it should be remembered that DATRAS ALK is constant. Aanes and Vølstad (2015) showed
302 that in such cases, and where only the variability of length compositions are allowed for, the estimated age-
303 distributions may appear to be more precise than they truly are since the ALK itself is subject to sampling
304 errors, see for example the estimated relative standard standard errors for ages 2, and the older fish (4, 5
305 and 6+) for both species.

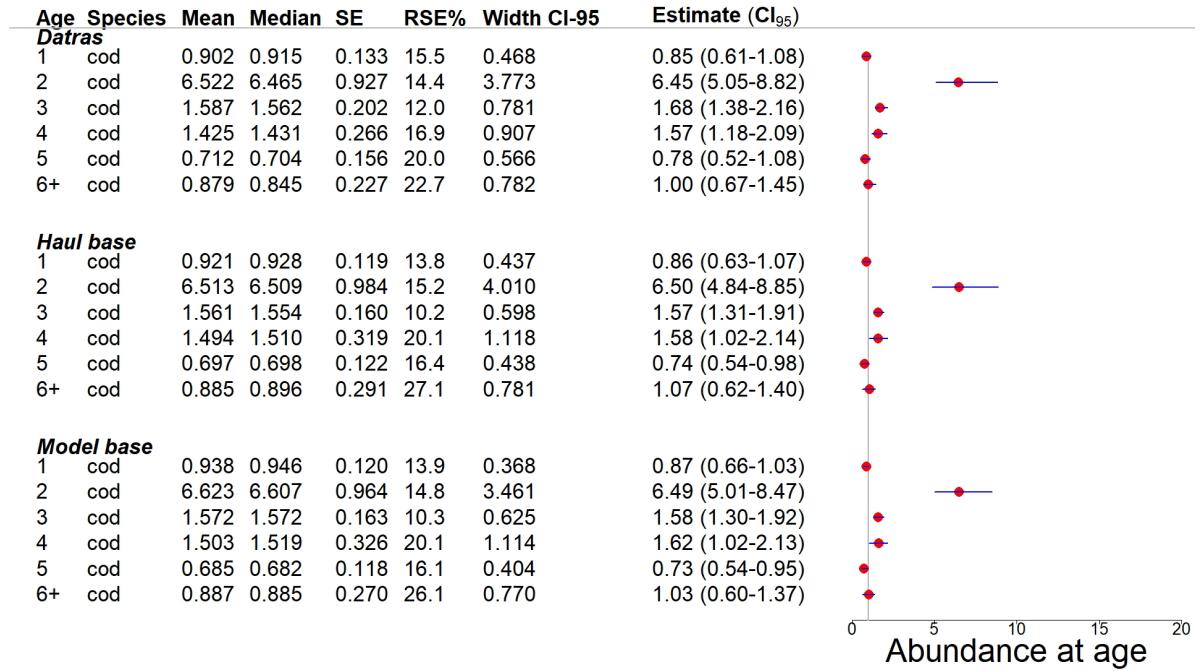
306 As regards to which spatial ALK method to adopt, it is difficult to identify a method that gives the
307 best performance over all age groups. While both methods seem to give reasonable estimates, the model
308 base ALK generally gave shorter interval widths for both species. It should be noted that the estimated
309 variance of abundance indices from the spatial model base ALK are computed using bootstrapping. The
310 variance-covariance matrix (fisher information) (**discuss issues with model based ALK and variance**
311 **estimation)**

312

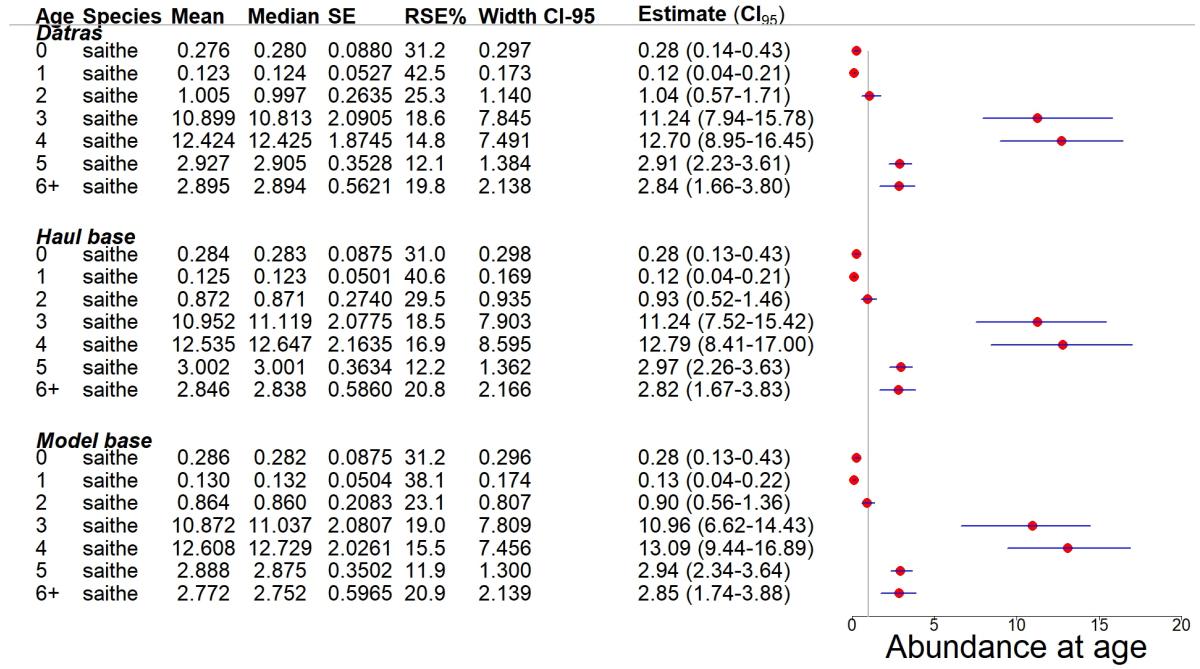
313 For these reasons we recommend the haul base ALK method for estimating age-distributions??

314

315 The implications of using DATRAS bootstrap procedure for estimating the uncertainty around indices
316 of abundance are shown for illustration in Figure 5 in Web appendix E. Compared with the stratified
317 bootstrap procedure, DATRAS bootstrap procedure gives an overestimation of the uncertainty for all age
318 groups, suggesting that it is highly relevant to account for the variation in the data over large areas.



(a) Cod in year 2018 Q1



(b) Saithe in year 2017 Q3

Figure 3: Estimated confidence intervals (CI₉₅) 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.

319 **3.2 Optimum sampling effort for North Sea Cod and Saithe**

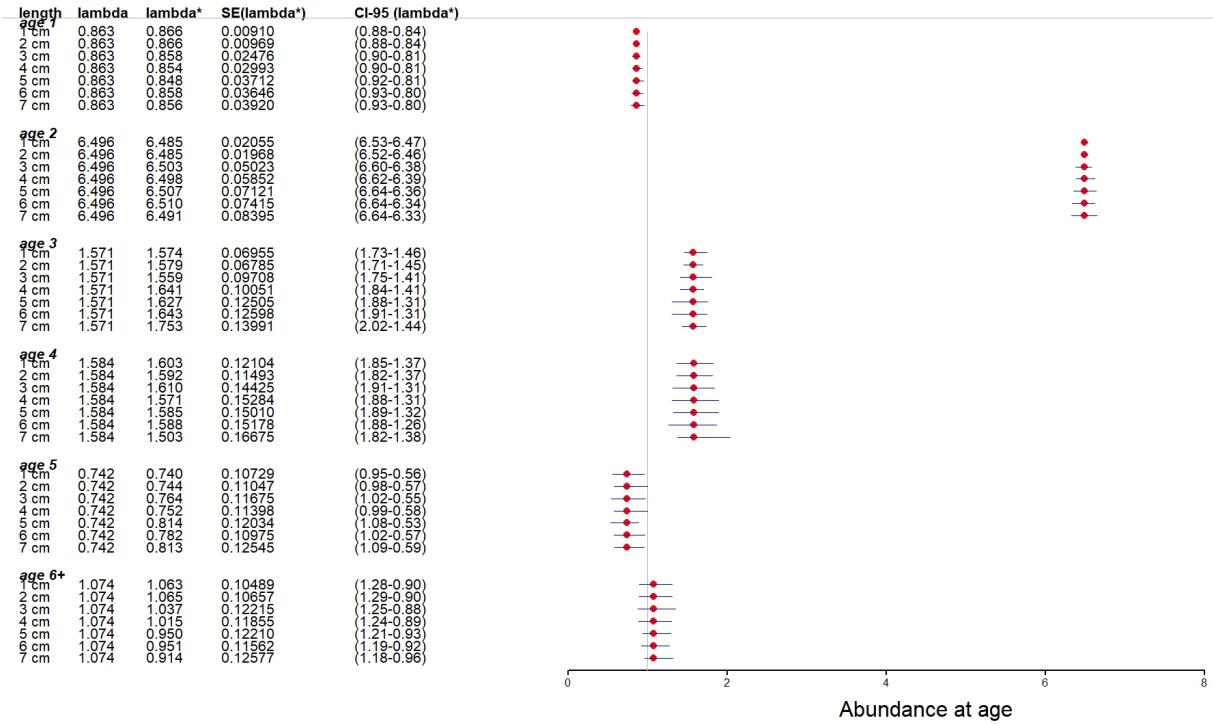
320 In order to determine optimum sampling levels of otoliths for saithe and cod in the North Sea, ALKs are
321 estimated using the haul based method. As shown in Figures 3 and ?? the haul base ALK and the model
322 base ALK gave similar estimates of abundance indices and precision as both approaches are attempting to
323 account for spatial variation in the data. But, the spatial ALK model is quite complex, and model fitting
324 would be computer-intensive since the model must be fitted for each bootstrap run and each simulated
325 sampling procedure that mimics the real data collection procedure. For this reason, the model base ALK
326 approach will not be used in this analysis. Also, the assumption of no difference in regional compositions of
327 age-length structures is invalid, as shown in Figure 2, so DATRAS ALK is not use for further analyses. The
328 removal procedure for otolith sampling described in Section 2.4 is applied to data in year 2018 Q1 for cod
329 and year 2017 Q3 for saithe.

330 In year 2018 Q1, the number of otoliths sampled for cod was **1600**, of which 231, 457, 615, 712, 788, 841,
331 and 889 are removed in the experiments 1 cm, 2 cm, 3 cm, 4 cm, 5 cm, 6 cm or 7 cm length group, respectively.

332 In year 2017 Q3, 2092 otoliths were sampled for saithe andwere removed in the experiments 1 cm, 2 cm,
333 3 cm, 4 cm, 5 cm or 7 cm length group, respectively. Figure4 gives estimates of abundance for the original
334 data, estimated indices for the reduced data, and their standard errors, and approximate 95% confidence
335 intervals for 200 simulations and 200 bootstrap replication.

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

337 Note that the nonparametric bootstrap method is advantageous because it does not assume any distribu-
338 tion for the data, and it also accounts for some of the variability in the sampling distribution of the CPUE,
339 however, there are some limitations of this method. The most important limitation is the assumption that
340 the distribution of the data represented by the sample is a reasonable estimate of the population function
341 from which the data are sampled. If this assumption is violated the random sampling performed in the boot-
342 strap procedure may add another level of sampling error, resulting in invalid statistical estimations (Haukoos
343 and Lewis, 2005). As discussed in Section 1.1 the selection of the trawling locations in IBTS is semi-random
344 where cruise leaders selects "clear" tow locations or "blind" tow locations if no clear tow exists by checking
345 the proposed trawl track for hazardous seabed obstructions with acoustic methods. More recently selection
346 of tow locations is based on pre-proposed valid tow locations with start and end positions executed in the
347 period 2000-2017. Hence, the lack of a fully randomized sampling process has the potential to result in biased
348 estimates of parameters and their uncertainty. Random sampling performed in the bootstrap procedure also
349 adds another level of potential sampling error, which is reflected in variation and biased estimates commonly
350 performed in the bootstrap analysis. Note that the sampling distribution of the bootstrapped statistics is
351 frequently not symmetric and computing point estimates from in this manner may reflect biased estimation
352 from the samples. This can be seen in the estimated bootstrap mean values in Figure 4



(a) Cod in year 2018 Q1

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

353

4 DISCUSSION

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

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

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

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

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

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

- 379 • discuss sampling procedure: limitation and advantages; and possibly more advanced selection proce-
380 dures?
- 381 • new approach adopted in surveys from 2018
- 382 • IBTS has a standardized survey indices? -(yes Berg's NS-IBTS Delta-GAM index). so changes in
383 catch rates are due to changes in population size? Berg et al. (2014) developed a standardized index for
384 IBTS data but only applied to some species e.g., haddock? cod- last year 2017. Is the designed based
385 age index on DATRAS not a standardized index?
- 386 • how does changes in survey design or other factors affect changes in catch rates? If so are these changes
387 significant?

388 5 General comments

- 389 • Decide on whether we say, "in this research or paper"
- 390 • Decide on whether to say, "In this subsection or section"
- 391 • Decide on year of data for case studies
- 392 • Decide on writing "haul(model)-based or haul (model) based"
- 393 • Decide on calling the survey "The North Sea IBTS or IBTS"
- 394 • Decide on writing "Cod or cod, and Saithe or saithe"
- 395 • Decide on a title for the paper
- 396 • what are issues with including haul effect in model based ALK? (Olav)

References

- 398 Aanes, S. and Vølstad, J. H. (2015). Efficient statistical estimators and sampling strategies for estimating
 399 the age composition of fish. *Canadian journal of fisheries and aquatic sciences*, 72(6):938–953.
- 400 Berg, C. W. and Kristensen, K. (2012). Spatial age-length key modelling using continuation ratio logits.
 401 *Fisheries Research*, 129:119–126.
- 402 Berg, C. W., Nielsen, A., and Kristensen, K. (2014). Evaluation of alternative age-based methods for
 403 estimating relative abundance from survey data in relation to assessment models. *Fisheries Research*,
 404 151:91–99.
- 405 Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide
 406 for medical statisticians. *Statistics in medicine*, 19(9):1141–1164.
- 407 Ehrhardt, N. M. and Legault, C. M. (1997). The role of uncertainty in fish stock assessment and management:
 408 a case study of the spanish mackerel, scomberomorus maculatus, in the us gulf of mexico. *Fisheries
 409 research*, 29(2):145–158.
- 410 Fridriksson, A. (1934). On the calculation of age-distribution within a stock of cod by means of relatively few
 411 age-determinations as a key to measurements on a large scale. *Rapports Et Proces-Verbaux Des Reunions,
 412 Conseil International Pour l'Exploration De La Mer*, 86:1–5.
- 413 Fuller, W. A. (2011). *Sampling statistics*, volume 560. John Wiley & Sons.
- 414 Gerritsen, H. D., McGrath, D., and Lordan, C. (2006). A simple method for comparing age-length keys
 415 reveals significant regional differences within a single stock of haddock (melanogrammus aeglefinus). *ICES
 416 Journal of Marine Science*, 63(6):1096–1100.
- 417 Haukoos, J. S. and Lewis, R. J. (2005). Advanced statistics: bootstrapping confidence intervals for statistics
 418 with difficult distributions. *Academic emergency medicine*, 12(4):360–365.
- 419 Hirst, D., Storvik, G., Rognebakke, H., Aldrin, M., Aanes, S., and Vølstad, J. H. (2012). A bayesian
 420 modelling framework for the estimation of catch-at-age of commercially harvested fish species. *Canadian
 421 journal of fisheries and aquatic sciences*, 69(12):2064–2076.

- 422 ICES (2012). Manual for the international bottom trawl surveys, revision viii. series of ices survey protocols.
- 423 *International Council for the Exploration of the Sea*, SISP 1-IBTS VIII.
- 424 ICES (2013a). Ns-ibts indices calculation procedure. datras procedure document. *International Council for*
425 *the Exploration of the Sea*, 1.1 NS-IBST indices-2013.
- 426 ICES (2013b). Report of the third workshop on practical implementation of statistical sound catch sampling
427 programmes. *International Council for the Exploration of the Sea*, WKPICS3.
- 428 ICES (2015). Manual for the international bottom trawl surveys, revision ix. series of ices survey protocols.
- 429 *International Council for the Exploration of the Sea*, SISP 10-IBTS IX.
- 430 ICES (2016a). Ices advice on fishing opportunities, catch, and effort greater north sea and celtic seas
431 ecoregions. *International Council for the Exploration of the Sea*, Book 6.
- 432 ICES (2016b). Report of the benchmark workshop on north sea stocks (wknsea), 14–18 march, 2016.
433 *International Council for the Exploration of the Sea*, ICES CM 2016/ACOM:37.
- 434 ICES (2018a). Ices fishmap: Atlas of the north sea fish, including fact sheets of key species and distribution
435 maps. *International Council for the Exploration of the Sea (ICES FishMap website, 2018*, 1.
- 436 ICES (2018b). Manual for the international bottom trawl surveys, revision x. series of ices survey protocols.
- 437 *International Council for the Exploration of the Sea*, SISP X-IBTS X.
- 438 ICES, D. (2018c). Datras.
- 439 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. (2015). Tmb: automatic differentiation
440 and laplace approximation. *arXiv preprint arXiv:1509.00660*.
- 441 Lehtonen, R. and Pahkinen, E. (2004). *Practical methods for design and analysis of complex surveys*. John
442 Wiley & Sons.
- 443 Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian
444 Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical
445 Society: Series B (Statistical Methodology)*, 73(4):423–498.

- ⁴⁴⁶ Ludwig, D. and Walters, C. J. (1981). Measurement errors and uncertainty in parameter estimates for stock
⁴⁴⁷ and recruitment. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(6):711–720.
- ⁴⁴⁸ Nottestad, L., Utne, K. R., 'Oskarsson, G. J., J'onsson, S. T., Jacobsen, J. A., Tangen, O., Anthonypillai,
⁴⁴⁹ V., Aanes, S., Vølstad, J. H., Bernasconi, M., et al. (2015). Quantifying changes in abundance, biomass,
⁴⁵⁰ and spatial distribution of northeast atlantic mackerel (*scomber scombrus*) in the nordic seas from 2007
⁴⁵¹ to 2014. *ICES Journal of Marine Science*, 73(2):359–373.
- ⁴⁵² Puth, M.-T., Neuhäuser, M., and Ruxton, G. D. (2015). On the variety of methods for calculating confidence
⁴⁵³ intervals by bootstrapping. *Journal of Animal Ecology*, 84(4):892–897.
- ⁴⁵⁴ Rindorf, A. and Lewy, P. (2001). Analyses of length and age distributions using continuation-ratio logits.
⁴⁵⁵ *Canadian Journal of Fisheries and Aquatic Sciences*, 58(6):1141–1152.
- ⁴⁵⁶ Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models
⁴⁵⁷ by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B
(Statistical Methodology)*, 71(2):319–392.
- ⁴⁵⁹ Walters, C. J. and Ludwig, D. (1981). Effects of measurement errors on the assessment of stock–recruitment
⁴⁶⁰ relationships. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(6):704–710.
- ⁴⁶¹ Wood, S. (2015). Package mgcv. *R package version*, pages 1–7.
- ⁴⁶² Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

463

A Areas fished by different countries in the North Sea IBTS

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

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

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

475

476

B Otolith sampling per fish species

477 From 1991-2017, most countries conducted quota sampling of otoliths per length group in a RFA. But from
478 2013 Norway has been sampling one otolith per length class from each trawl haul (to 0.1cm below for shellfish,

479 to 0.5cm below for herring and sprat and to 1cm below for all other species). From the first quarter in 2018
 480 all countries are required to sample one otolith per length class per trawl haul. Table 3 gives the minimum
 481 sampling levels of otoliths for the target species. However, for the smallest size groups, that presumably
 482 contain only one age group, the number of otoliths per length class may be reduced, and more otoliths per
 483 length are required for the larger length classes.

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

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

C Weightings of Statistical Rectangles

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

StatRec	Weight								
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

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

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

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

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

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

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

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

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

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

514 *length classes in each pooled length group. To replace missing values for the age distribution in the pooled*
515 *length groups the method of "borrowing" ages from length groups in trawl hauls closest in air distance within*
516 *the RFA is used. If there are no observed ages in the pooled length group in the RFA, missing values for the*
517 *age distribution are replaced following the procedure outlined in the DATRAS ALK procedure (D.1) in step*
518 *1.*

519 **E Estimates from DATRAS and Stratified bootstrap procedures**

520 The bootstrap procedure proposed by DATRAS lacks the potential to account for the spatial variation in
521 the data. The DATRAS bootstrap procedure ignores the fine-scale stratification in the sampling process,
522 leading to an overestimation of the uncertainty; and ignores the age-length data collected at the haul level,
523 resulting in an underestimation of the uncertainty. The results (Figure5) shows an overestimation of the
524 uncertainty for all age groups, suggesting that it is relevant to account for the fine-scale stratification when
525 resampling the data.

Figure 5: 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.