Fisheries Research 129–130 (2012) 119–126

Contents lists available at SciVerse ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Spatial age-length key modelling using continuation ratio logits

Casper W. Berg[a,*], Kasper Kristensen[b]

[a] National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund Castle, 2920 Charlottenlund, Denmark
[b] Department of Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 305, 2800 Lyngby, Denmark

ARTICLE INFO

Article history:
Received 27 March 2012
Received in revised form 21 June 2012
Accepted 22 June 2012

Keywords:
Age length key comparison
Continuation ratio logits
Generalized Additive Models

ABSTRACT

Many fish stock assessments are based on numbers at age from research sampling programmes and samples from commercial catches. However, only a small fraction of the catch is typically analyzed for age as this is a costly and time-consuming process. Larger samples of the length distribution and a so-called age-length key (ALK) is then used to obtain the age distribution. Regional differences in ALKs are not uncommon, but stratification is often problematic due to a small number of samples. Here, we combine generalized additive modelling with continuation ratio logits to model the probability of age given length and spatial coordinates to overcome these issues. The method is applied to data gathered on North Sea haddock (*Melanogrammus aeglefinus*), cod (*Gadus morhua*), whiting (*Merlangius merlangus*) and herring (*Clupea harengus*) and its implications for a simple age-based survey index of abundance are examined. The spatial varying ALK outperforms simpler approaches with respect to AIC and BIC, and the survey indices created using the spatial varying ALK displays better internal and external consistency indicating improved precision.

1. Introduction

Estimation of catch at age from combined samples of length and age is standard procedure in analyses of fisheries data. Only a small fraction of the catch is typically analyzed for age as this is a costly and time-consuming process. Larger samples of the length distribution and a so-called age-length key (ALK) are then used to obtain the age distribution.

The ALK is typically estimated from length-stratified subsamples of the catch that are analyzed for age by examining the annual ring structure in the otholiths. Missing or few data points for a given combination of strata such as age, length group, and geographical area frequently occur due to unreadable otholiths or simply because no fish were caught. In this case the raw observed proportions of age-at-length are unsuitable for assigning age to fish in these length groups. A solution to this problem is to use a statistical model to create a smooth distribution of age given length and possibly other covariates, such that missing values can be interpolated in an objective and robust way, and the uncertainty due to the sampling variability can be taken into account. A statistical model also has the advantage of allowing formal testing of hypotheses such as whether two ALKs can be considered identical.

Continuation ratio logits (CRLs) is a type of model for ordered categorical responses (such as age groups) and it has previously been used for modelling ALKs (Kvist et al., 2000; Rindorf and Lewy, 2001). In addition to ALKs, Rindorf and Lewy (2001) also applied CRLs for estimating smooth length distributions.

CRLs have also been used to investigate spatial differences in ALKs (Gerritsen et al., 2006; Stari et al., 2010). In both cases, significant spatial differences were found in ALKs for North Sea haddock. Gerritsen et al. (2006) divided their data into 3 depth strata and examined the differences between using a single ALK and ALKs calculated for each stratum. The shallow stratum was significantly different from the deeper strata, with higher probabilities for younger fish in the shallow stratum. Using a combined ALK for all the strata resulted in nearly twice as many 1-year olds compared to a survey index calculated from the stratified ALKs. Stari et al. (2010) found significant differences between geographical areas, mature and immature fish, commercial and survey data, and fleets using different fishing gear.

In all previous applications of CRLs to ALK modelling, Generalized Linear Models (GLMs) have been used for estimation, and stratification has been used to model the effect of regional differences. Any type of stratification will exacerbate the problems with missing data, and the choice of strata will often be a somewhat subjective decision made by the modeller. In situations where detailed information about the geographical origin of the age samples is available, it is possible to consider alternatives to stratification by area.

One such alternative is to use Generalized Additive Models (GAMs) in place of GLMs. GAMs is a non-parametric tool for non-linear modelling, which allows smooth functions of the explanatory

* Corresponding author.
E-mail address: cbe@aqua.dtu.dk (C.W. Berg).

0165-7836/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2012.06.016

variables in the specification of the mean value distribution, and numerous studies have used GAMs for modelling spatial effects. Toscas et al. (2009) used GAMs to fit spatio-temporal models of prawn catches with 2D thin plate regression splines for modelling spatial variation. The GAMs interpolated the data well, but extrapolation beyond data coverage was found to be problematic. Maxwell et al. (2012) compared GAMs to a stratified mean method (stratification in time and space) for estimating egg production of cod, plaice, and haddock in the Irish Sea. The methods gave relatively consistent estimates, but the GAM methodology offered higher precision and was better suited for handling missing observations. For a thorough introduction to GAMs see Wood (2006).

In this study we describe how GAMs can be used for fitting CRLs to model age as a smooth function of length and geographical position. In addition to the advantages offered by the GLM approach, it eliminates the need for spatial stratification by providing an ALK that varies smoothly with geographical position.

The methodology has been fully implemented in the `DATRAS` software package (Kristensen and Berg, 2012) for `R` (R Development Core Team, 2012), which offers an accessible way to create ALKs from all the data available in the DATRAS database (www.datras.ices.dk) as well as other data using this format.

Using ten years of survey data the new method is compared to the traditionally applied regional stratification of ALKs to determine whether a significantly better fit to data is obtained. Furthermore, internal and external consistencies are calculated to examine whether the new method leads to improved precision when used to create indices of abundance by age.

2. Methods

The response variable is the age group of a fish, $a = R \ldots A$, i.e., ordered categorical also known as ordinal response, where R denotes the youngest age category and A is the oldest. The latter category is often defined to be a "plus group" which consists of fish of age A or older. For each fish where the age has been determined, a set of covariates $\mathbf{x}$ is also available, which in this study includes the length l of the fish and the spatial coordinates of the fishery.

The continuation ratio model (Agresti, 2010) is well suited to model the distribution of ages $P_a(\mathbf{x}) = \{p_R \ldots p_A\}$. This is accomplished through A minus R models for the conditional probability of being of age a given that it is at least age a. That is, let

$$\pi_a = P(Y = a | Y \geq a) = \frac{p_a}{p_a + \cdots + p_A}, \quad a = R \ldots A - 1$$

be those conditional probabilities and let our set of continuation ratio logits be given by GAMs of the following type:

$$\text{logit}(\pi_a[\mathbf{x}_i]) = \mathbf{x}_i^* \boldsymbol{\theta}_a + f_{1a}(x_{1i}) + f_{2a}(x_{2i}, x_{3i}) + f_{3a}(x_{4i}) x_{5i} + \cdots, \quad a = R \ldots A - 1$$

where $\mathbf{x}_i$ is a vector of covariates, $\mathbf{x}_i^*$ is a subset of the covariates entering linearly in the model, $\boldsymbol{\theta}$ is the corresponding parameter vector, and f_j denotes some smooth function of the covariates x_k, which may be of one or more dimensions and also multiplied by known covariates. Given the set of $A - R$ models, we can calculate the estimated unconditional probabilities $\hat{p}_a$ from the conditional probabilities $\hat{\pi}_a$ (the dependence on covariates is omitted here):

$$\hat{p}_R = \hat{\pi}_R$$

$$\hat{p}_a = \hat{\pi}_a \left(1 - \sum_{j=R}^{a-1} \hat{p}_j \right) = \hat{\pi}_a \prod_{j=R}^{a-1} \left(1 - \hat{\pi}_j \right), \quad a > R$$

We choose to consider the following six formulations of the CRLs:

1. A single common ALK for the whole North Sea fitted using GLM methodology.
2. A stratified approach having separate ALKs within 3 subareas of the North Sea (see Fig. 1), also fitted using GLMs.
3. A smooth spatial varying ALK fitted using GAMs with smoothness selection by AIC. Only the intercept in the models are allowed to vary with location.
4. Same as model 4, but with smoothness selection by BIC instead of AIC.
5. A GAM where both the intercept and the regression coefficient on length are allowed to vary with geographical coordinates.
6. Like model 4, but with the same spatial effect in all years, as opposed to estimating a set of parameters for each year.

Using mathematical notation these six models can be written as follows:

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \beta_{ayq} l_i \tag{1}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \delta_{ayq}(Area_i) + \omega_{ayq}(Area_i) l_i \tag{2}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \beta_{ayq} l_i + s_{ayq,AIC}(lon_i, lat_i) \tag{3}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \beta_{ayq} l_i + s_{ayq,BIC}(lon_i, lat_i) \tag{4}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + s_{ayq,BIC}(lon_i, lat_i) l_i + s_{ayq,BIC}(lon_i, lat_i) \tag{5}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \beta_{ayq} l_i + s_{aq,BIC}(lon_i, lat_i) \tag{6}$$

where i denotes the ith fish, l denotes the length of the fish, (lon, lat) the geographical coordinates where the haul was taken (longitude and latitude), $\delta_a(Area_i)$ maps the ith observation to one of 3 categorical effects for a division of the North Sea into 3 areas (see Fig. 1), and similarly denotes ω_a a regression parameter for each of the 3 areas, s_a is a thin plate spline in two dimensions, where subscripts AIC and BIC denote which criterion is used for smoothness selection, and (α_a, β_a) are ordinary regression parameters to be estimated. Subscripts y and q have been included here to indicate that each combination of year and quarter should have a distinct set of parameters to account for differences in population structure.

Note, that model 2 is equivalent to dividing the data set according to the 3 areas and fitting model 1 with individual parameters for each area. Models 3 and 4 include a spatial varying intercept for each continuation ratio logit but a common regression parameter on length, whereas model 5 is a varying-coefficients model (Hastie and Tibshirani, 1993), where both the intercept and the regression parameter are allowed to vary with geographical coordinates. Model 6 is like model 4 except that the spatial effect is constrained to be identical for all years. All the parameters in the model has the subscript a indicating that each logit has a distinct set of parameters. This implies that the likelihood equation can be partitioned into separate terms for each logit (Agresti, 2010; Kvist et al., 2000), and hence each logit can be fitted separately. Similarly, the total deviance for the model is simply the sum of deviances from the individual fits. This feature makes it possible to fit the continuation ratio logit model using standard software that can handle binomial responses.

Our GAM models are based on the implementation in the `mgcv` package for `R` (Wood, 2006), which offers a variety of types including multi-dimensional splines and automatic smoothness selection. We follow the recommendation in Wood (2006), who suggests using thin plate regression splines for inputs on same scale and where isotropy is relevant such as spatial coordinates. All the thin plate splines used in this study for geographical effects are splines with shrinkage smoothing (Wood, 2006, p. 160), which

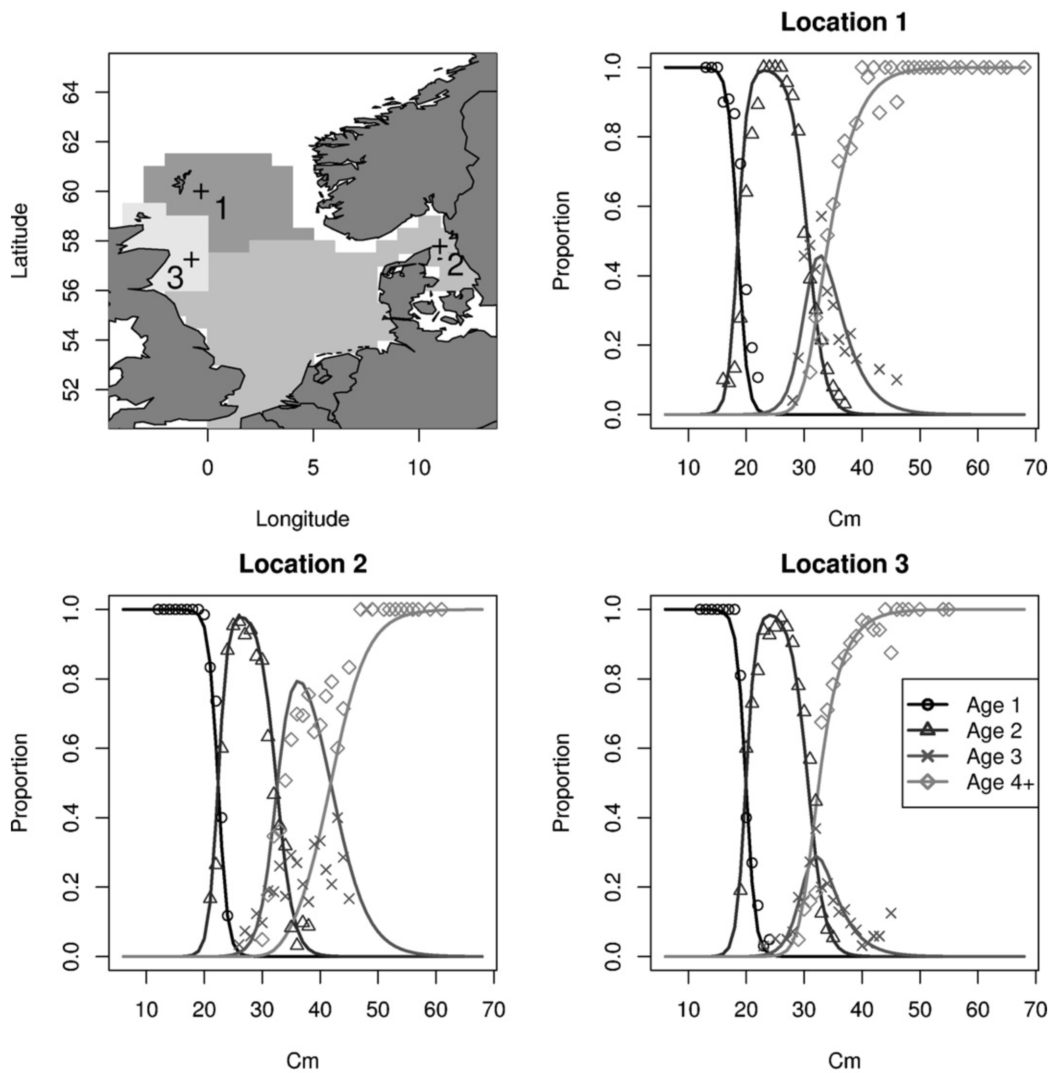

Fig. 1. Map of the three areas used in model 2 and of three selected locations 1,2 and 3 (top left). Predicted probabilities of age given length using model 4 at each location (solid lines) as well as the raw observed proportions (points) in each of the boxed areas for the year 2011 Q1 (top right and bottom figures).

allows them to be completely eliminated from the model in the sense of having all the parameters estimated to be zero.

2.1. Evaluating the implication of the ALK

Previous works (Gerritsen et al., 2006; Stari et al., 2010) have utilized the generalized likelihood ratio test for measuring whether two ALKs could be considered identical. This test requires that the smaller model is nested within the larger, which is not the case for all our models. Instead, we use the AIC and BIC values to investigate which models are more appropriate. For the GAMs, the number of parameters, which is needed to calculate the AIC and BIC, is replaced by the *effective degrees of freedom* (edf), see Wood (2006) for details.

However, since the AIC and BIC only applies to the age data, it does not tell us whether applying the estimated ALKs to all the length data will result in significant changes in an index of abundance by age. We will therefore create such an index to investigate the implications of our proposed method for creating ALKs. If a spatial ALK, in addition to providing a better fit to the age data, also results in improved precision for a derived index of abundance, this can be seen as further evidence that the spatial ALK is more appropriate. If the spatial effect in the ALKs were really noise rather than a true signal, one would expect the precision of an index of abundance to deteriorate when applying a spatial ALK as opposed to a non-spatial ALK.

We choose one of the simplest estimates of abundance:

$$I_{ayq} = \frac{1}{h_{yq}} \sum_{i=1}^{n_{yq}} \hat{p}_a(\mathbf{x}_i) \tag{7}$$

where I_{ayq} is the average predicted number of fish caught in age group a per haul in year and quarter (y, q), n is the total number of fish caught, and h is the number of hauls.

An appropriate way to test whether one index of abundance is more accurate than another would be to run full assessment models using the different indices as well as commercial catch data and compare their estimated observation variances. However, since this is a quite complicated task we choose a simpler way of comparing our different indices of abundance based on the concepts of internal and external consistency (e.g., Payne et al., 2009). Under the assumptions that an index is proportional to the abundance without error and of constant catchability and constant total mortality over time, the logarithm of the abundance at time t should be perfectly correlated with the logarithm of abundance of the same cohort at time $t + \Delta t$. Although all these assumptions are not correct, we should still be able to obtain significant correlations for values of Δt within the range of a year, given that the signal in the time-series outweighs the variability from sampling noise and violations of our assumptions. Recapping from Payne et al. (2009), if we assume that we have a survey index with a log-normal error structure and substitute this into the Baranov catch equation we get

$$I_{a(t)} = q_{a(t)} N_{a(t)} \epsilon_{a(t)}, \quad \epsilon \sim LN(0, \sigma_{a(t)})$$

$$\log(I_{a(t)}) = \log(I_{a(t+\Delta t)}) + \log\left(\frac{q_{a(t+\Delta t)}}{q_{a(t)}}\right) + \log(\epsilon_{a(t+\Delta t)}) - \log(\epsilon_{a(t)}) - Z_a(t, t+\Delta t)$$

where $I_{a(t)}$ refers to the index of abundance for some age group a at time t, q denotes catchability, Z the total mortality over the considered time interval, and ϵ is a random log-normal distributed component.

Internal consistency refers to correlations between Is within the same survey index (e.g., Age 1 in quarter 1 year y versus Age 2 in quarter 1 in year $y+1$), whereas external consistency refers to comparing two independent survey indices, such as those for quarter 1 and 3 (Q1 and Q3). We will refer to internal consistency between age a and $a+1$ ($\Delta t = 1$ year) in quarter q as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 ($\Delta t = 0.5$ years) as $EC(a)$.

3. Case studies

In this section the method will be applied to ten years (2001–2011) of data from the International Bottom Trawl Survey (IBTS) obtained from the DATRAS database (www.datras.ices.dk). The samples are collected in the first and third quarters of the year and all samples are caught using the same gear type. For further details about the IBTS survey see (ICES, 2012). In section 3.1 we will investigate an application of models 1 through 6 on North Sea haddock data. Section 3.2 will deal with a less detailed rerun of models 1 and 4 on multiple species focusing on consistencies only.

3.1. Haddock using models 1–6

For the area stratified model 2 we divide the North Sea into 3 areas (Fig. 1) with roughly the same number of age samples per year (see tables in online supplemental material[1]). Area 2 is much larger than the others, but this is due to the fact that haddock is primarily caught in the northern parts of the North Sea. For all models except model 2, it was possible to consider up to age group 8 without estimation problems. However, for simplicity we consider the age groups 1 to 4+ for all models, where the last group consists of fish of age 4 or older. As age 0 appears for the first time in the IBTS survey in Q3, it must also be included when creating the ALKs for this time-series, but results of this estimation are not included in the further analysis.

Table 1 shows the AIC and BIC calculated for each combination of model and quarters. Since lower values of AIC and BIC are to be preferred, model 2 is consistently better than model 1, implying that there is significant geographical variation in the ALKs. Model 3 is consistently best with respect to AIC while models 6 and 5 are respectively best with respect to BIC for Q1 and Q3, but the differences are much smaller between models 3–5 than the rest. These values provide strong evidence against a null hypothesis of no spatial effect in the ALKs, and also that the stratified GLM approach did not sufficiently capture the spatial variation.

Fig. 1 shows the fitted distribution (model 4) of age given length at three selected locations, as well as the raw observed proportions within each stratum. The observed proportions seem to differ between areas, and the fit in the three chosen locations resembles the raw observations in the three strata. We should note, that we cannot expect the fitted distributions to be the best interpolation of the raw proportions since the raw proportions are calculated over the entire stratum, but the shown fitted distributions applies only to the points in space marked by the numbers on the map, and the fits will therefore vary over the strata due to the significant spatial effect in the model.

Table 1
Haddock: summary of models 1–6. The columns 'ΔAIC' and 'ΔBIC' contain the decrease in AIC and BIC from model 1, and the best values are shown in bold face. The column 'edf' contains the effective number of parameters.

Model	Quarter	edf	ΔAIC	ΔBIC
1	1	66	0	0
2	1	198	3028.36	1940.47
3	1	770.10	**10,155.53**	4352.64
4	1	361.13	9252.17	6819.83
5	1	396.28	9600.90	6878.87
6	1	125	7659.33	**7173.11**
1	3	66	0	0
2	3	198	2966.02	1858.38
3	3	936.62	**10,605.83**	3300.25
4	3	384.29	9268.28	6597.43
5	3	421.89	9720.27	**6733.91**
6	3	130.88	6685.59	6141.19

Fig. 2 shows the spatial pattern in the probability of being older than one year given a length of 20 cm in 2011 Q1. The figure illustrates that there is spatial contrast in the data with a peak east of the Scottish coast. Given that a 20 cm haddock is caught in this region, it is more likely to be 2 years or older than being 1 year old, whereas the opposite is true in the south-eastern parts of the North Sea.

In order to illustrate the differences between models 1–6, the estimated age probabilities for a 30 cm haddock along a selected route (Fig. 3) from each model in year 2001 Q3 are shown in Fig. 4. The same plots for all the years and quarters can be found in the

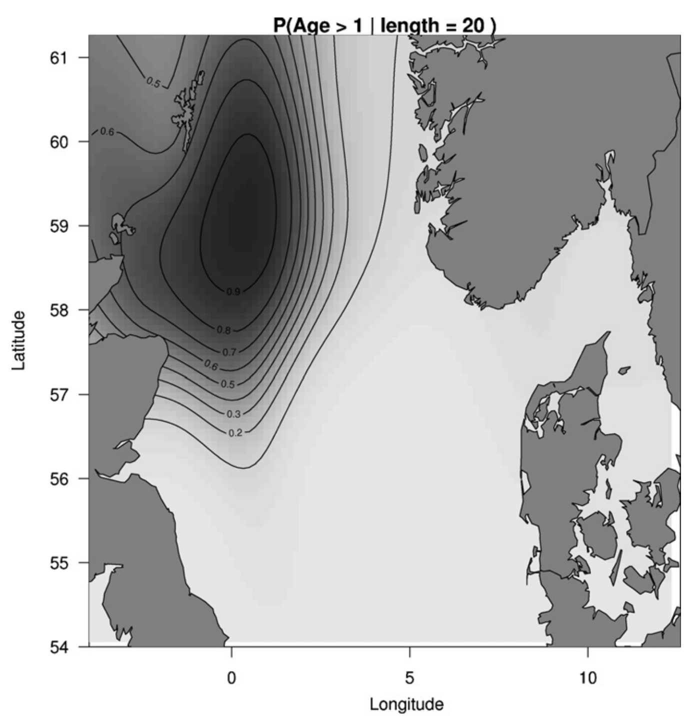

Fig. 2. Contour plot of the estimated probability (model 4) of being older than 1 year given a length of 20 cm in year 2011 Q1.

[1] See Appendix A.

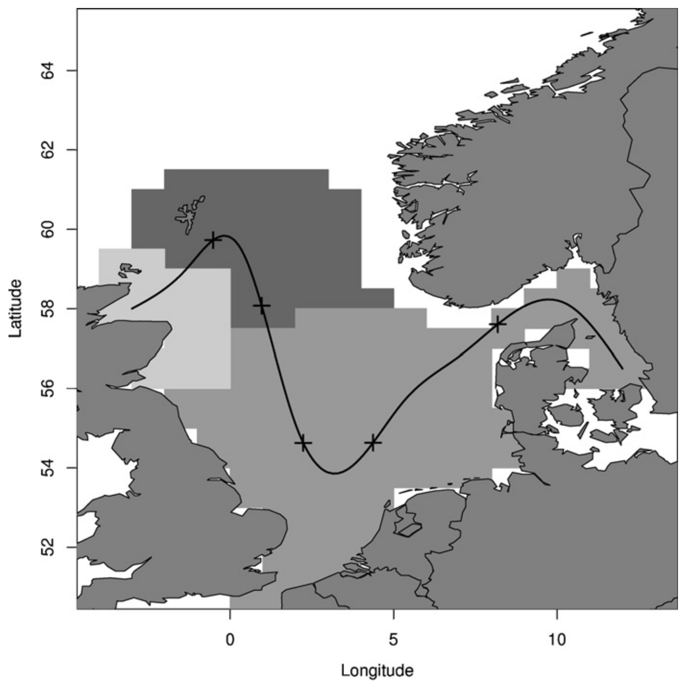

Fig. 3. A selected route through the North Sea and some selected points marked by '+'.

online supplemental materials. The models based on GAMs (3–6) all show a steep increase in probability for observing younger individuals on last part of the route around the Skagerrak region. Although there is considerable variation between years and quarters in the estimated probabilities, the spatial pattern seems to be relatively consistent. This is also supported by the fact, that model 6, which has the same spatial effect over all the years, was chosen as the best model by the BIC criterion for Q3. Models 4 and 5 display very similar results, while model 3 in some years estimates some more wiggly curves in comparison, due to the AIC criterion being less restrictive than BIC in terms of the amount of smoothing.

To illustrate the implications of using the different models for our simple index of abundance we have plotted $\log(I_{2yq})$ and $\log(I_{3yq})$ in Fig. 5. There seems to be very high consistencies between the series, both internally and externally, for all ALKs. This implies, that even though significant differences were found between the ALKs, the resulting indices of abundance turned out to be quite similar. The uncertainties on the indices of abundance were further investigated using bootstrapping (not shown), and these analyses confirmed that the difference between the calculated indices were generally not statistically different.

The internal and external consistencies are shown in Table 2, which confirms the apparent high correlation observed in Fig. 5, which implies a very strong signal in data. On average, models 4–6 have higher consistencies than the rest, which validates our conclusion that there is a spatial effect and that the GAM framework outperforms the stratified approach.

Table 2
Haddock: internal and external consistencies for models 1–6. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistency is shown in bold face.

Type\Model	1	2	3	4	5	6
IC(Q1, 1)	0.961	0.955	0.956	0.954	0.954	0.961
IC(Q1, 2)	0.910	0.919	0.918	0.917	0.916	0.918
IC(Q3, 1)	0.951	0.950	0.949	0.968	0.973	0.966
IC(Q3, 2)	0.970	0.976	0.944	0.963	0.968	0.969
EC(1)	0.972	0.973	0.969	0.968	0.969	0.955
EC(2)	0.985	0.993	0.980	0.992	0.992	0.994
EC(3)	0.921	0.948	0.954	0.956	0.945	0.963
Avg	0.953	0.959	0.953	0.960	0.959	**0.961**

Table 3
Haddock: internal and external consistencies for models 1 and 4. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistencies are shown in bold face.

	IC(Q1, x)		IC(Q3, x)		EC(x)	
x\Model	1	4	1	4	1	4
1	0.96	0.95	0.93	0.97	0.97	0.97
2	0.91	0.92	0.95	0.97	0.99	0.99
3	0.95	0.95	0.97	0.96	0.92	0.96
4	0.93	0.95	0.97	0.96	0.93	0.99
5	0.94	0.97	0.98	0.99	0.95	0.96
6	0.88	0.95	0.92	0.93	0.91	0.95
7	0.73	0.68	0.94	0.92	0.88	0.90
Avg	0.90	**0.91**	0.95	**0.96**	0.94	**0.96**

Table 4
Cod: internal and external consistencies for models 1 and 4. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistencies are shown in bold face.

	IC(Q1, x)		IC(Q3, x)		EC(x)	
x\Model	1	4	1	4	1	4
1	0.56	0.68	0.86	0.85	0.91	0.93
2	0.71	0.88	0.19	0.31	0.77	0.75
3	0.87	0.83	0.36	0.57	0.43	0.49
4	0.66	0.63	0.30	0.27	0.42	0.48
5	0.37	0.33	0.40	0.37	0.58	0.64
Avg	0.63	**0.67**	0.42	**0.47**	0.62	**0.66**

3.2. Models 1 and 4 on more species

Tables 3–6 show internal and external consistencies for models 1 and 4 for cod, haddock, whiting and herring in the North Sea. The choice of model 4 among the different GAM formulations was rather arbitrary, although it can be considered the more conservative choice with respect to the amount of spatial variation in the ALKs, as it uses the fewest number of effective parameters of the GAMs. Since we do not consider model 2, we can include a higher number of age groups without worrying about years with no observations of older age groups. For all species except herring, model 4 is consistently better than model 1 with respect to average consistency over age groups. While haddock has very high consistencies even in older age classes, herring has appalling consistencies for Q1 (some are even negative). Whiting and Cod have fairly good consistencies, perhaps with the exception of IC(Q3) for cod (4). These

Table 5
Whiting: internal and external consistencies for models 1 and 4. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistencies are shown in bold face.

	IC(Q1, x)		IC(Q3, x)		EC(x)	
x\Model	1	4	1	4	1	4
1	0.79	0.76	0.70	0.72	0.84	0.86
2	0.96	0.98	0.83	0.82	0.85	0.84
3	0.86	0.87	0.76	0.78	0.88	0.90
4	0.63	0.65	0.85	0.85	0.67	0.67
5	0.37	0.47	0.85	0.84	0.57	0.57
Avg	0.72	**0.75**	0.80	0.80	0.76	**0.77**

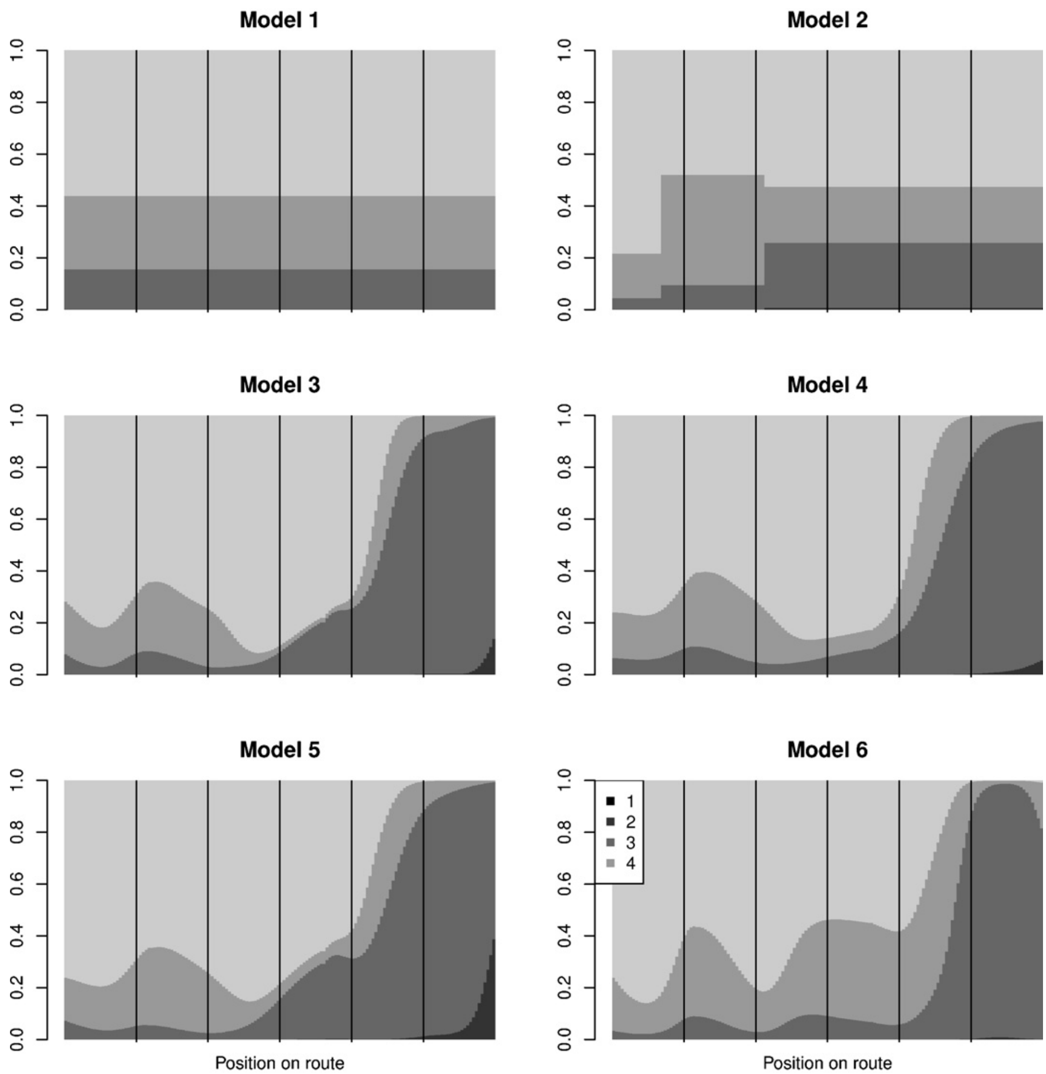

Fig. 4. Estimated age probabilities for a 30-cm haddock along the route shown in Fig. 3 from models 1 to 6 for year 2001 Q3. The x-axis corresponds to the position on the route from west to east, and the vertical lines indicate the positions marked with a '+' on the map.

results emphasize the results found for haddock, namely that there generally is spatial variation in length-at-age, and that improved precision in indices of abundance can be obtained by including this variation in the ALKs.

4. Discussion

Several studies have suggested using continuation ratio logits for modelling the age distribution in catch data from length-stratified subsamples of age in place of raw proportions of age given length. Two studies have also shown regional as well as other effects using CRLs for North Sea haddock (Gerritsen et al., 2006; Stari et al., 2010), a result that is confirmed in this study. While these studies used a number of parameters proportional to the number of boxed areas using GLM methodology, we propose to use GAM methodology to model spatial effects as a smooth surface and thereby be able to predict numbers-at-age at the haul level, whenever the required information is available. This removes the problem of having to select appropriate boxes for the data, and the problem of missing data whenever a too fine-grained stratification is chosen. This effect is comparable to the result found in Maxwell et al. (2012), who compared GAMs with a stratified mean method for modelling egg production in fishes. Also, the ALKs based on GAMs provided a much better fit to data than the GLM based methods examined in this study, and they were also superior in terms of both AIC and BIC. Our proposed model allows for a higher number of age groups than usual to be considered when an age based index of abundance is to be created, and, although there were only small differences in the survey indices between ALK methods, our results indicated that including spatial variation in ALKs seemed to improve the precision of the indices. It is straightforward to expand the number of covariates used in this study, using the same

Table 6
Herring: internal and external consistencies for models 1 and 4. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistencies are shown in bold face.

	$IC(\mathrm{Q1}, x)$		$IC(\mathrm{Q3}, x)$		$EC(x)$	
x\Model	1	4	1	4	1	4
1	0.22	0.36	0.58	0.56	0.58	0.57
2	−0.14	−0.09	0.78	0.77	0.28	0.34
3	0.07	−0.07	0.71	0.69	0.44	0.55
4	0.23	0.05	0.76	0.75	0.61	0.49
5	0.24	0.35	0.80	0.83	0.63	0.54
Avg	0.12	0.12	**0.73**	0.72	**0.51**	0.50

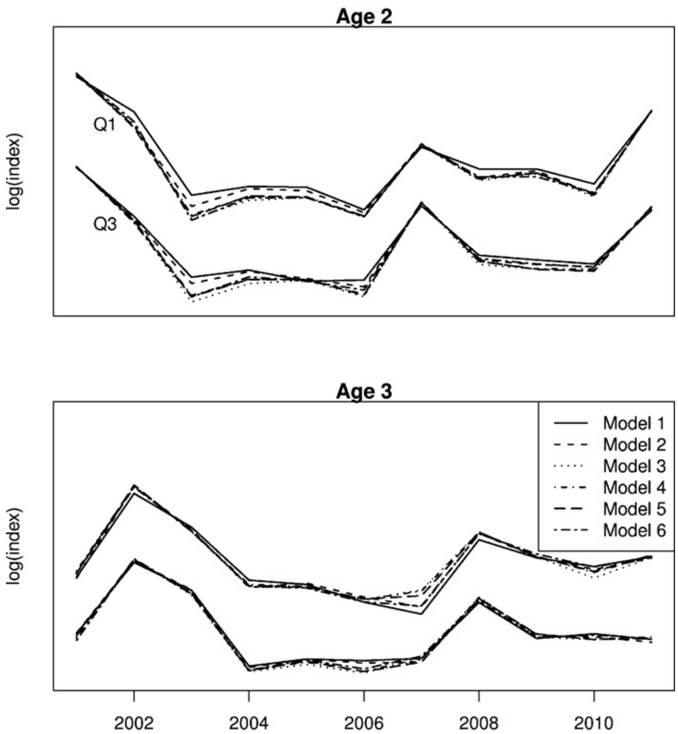

Fig. 5. Index of abundances for age groups 2 and 3 in Q1 and Q3. The series have been rescaled to prevent overlap between Q1 and Q3 for better overview, so only relative comparisons of the time-series are meaningful.

technique. While a spatial smoother is a convenient way of modelling the observed differences in ALKs between areas, it does not offer us an explanation for the observed effects. Possible explanations could be regional differences in growth, but also local variation in relative abundance of age-classes, which can occur due to migration, local differences in natural mortality, or even effects due to the data collection such as different laboratories used for ageing. In other words, the observed differences might be equally well explained by other covariates not included in our models, but given adequate spatial overlap between for instance different ageing labs, it will be possible to test for such an effect within our model framework while still accounting for residual unexplained spatial correlations by including a thin plate regression spline. Very high internal and external consistencies were found for all the age classes examined for haddock. While lack of consistency points to problems with some of the usual assumptions made for survey indices, strong consistencies are not proof of an excellent survey index, e.g., a constant index, which could hardly be informative, would yield perfect consistencies. We found fairly good consistencies for whiting and cod, but poorer consistencies for herring.

To ensure that changes in catch rates are due to changes in the population size rather than changes in survey design or other factors, survey indices should be standardized in some way to make them representative for the stock and comparable between years, a process which is sometimes called catch-rate standardization (e.g., Maunder and Punt, 2004). We should however keep in mind, that we did not perform a proper catch-rate standardization, but instead used a very simple index based on average numbers per haul. Also, catch rates for herring are generally much more variable than for the other species considered in this study, which can explain why our simple index performs so poorly for herring.

We should note, that even though many stock assessment models use age-structured indices of abundance as input, alternatives exist such as purely length-based models (e.g., Kristensen et al., 2006) or integrated stock assessments (e.g., Fournier et al., 1998) in which the separation into age-classes is performed within the stock assessment model, such that the associated uncertainty is included in the estimation. For stock assessments it should certainly be preferred to include the uncertainties due to the ALK estimation, either by integrating the ALK estimation within the stock assessment model, or to estimate the uncertainties on the derived indices of abundance by age outside the model, and provide these uncertainties as input to the stock assessment model along with the indices. The latter approach could be accomplished by bootstrapping, and is possible to carry out using the DATRAS-package.

Another useful aspect of ALKs is to combine them with the distributions of length and apply Bayes formula to get the probability of length given age (as opposed to age given length in ALKs), which for instance can be used to examine growth or differences in length distributions between regions. This idea was pursued in Rindorf and Lewy (2001) where CRLs were used for both the ALKs as well as the length distributions. The idea is, that since length distributions suffer from the same problems as age distributions, namely being patchy when small areas or individual hauls are considered, CRLs can be used to obtain smooth length distributions. Rindorf and Lewy (2001) used a seventh degree polynomial to obtain the length distributions on different locations, but noted that other types of smooth functions could be considered. GAMs could be considered in this respect, and this could be an interesting area for future research.

Fisheries data can be very complex, and the data sets available from DATRAS are certainly no exception to this rule. Producing an age-based survey index, which includes the application of an ALK, is therefore often a challenging task, and reproducing them by other people even more so. We have provided a software package for R that allows for manipulation of data from the DATRAS database, and easy generation and application of robust ALKs without the need for area stratification. The software package and all its source code is publicly available (Kristensen and Berg, 2012), which allows for adaptation to other data sets than those from the DATRAS database, including samples from commercial fisheries. Example code showing how to reproduce the models found in this paper is included in the online supplemental material. We have shown, that our approach is superior to the stratified approach with respect to AIC and BIC, and that it generally leads to better internal and external consistencies for age based survey indices.

Acknowledgements

The authors wish to thank all staff involved in the IBTS data collection, ICES for providing easy access to the data and answering our questions about how to interpret them, two anonymous reviewers for their constructive comments, and finally we wish to thank Anna Rindorf and Peter Lewy for valuable input to this manuscript.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fishres.2012.06.016.

References

Agresti, A., 2010. Analysis of Ordinal Categorical Data. Wiley Series in Probability and Statistics. Wiley.

Fournier, D.A., Hampton, J., Sibert, J.R., 1998. Multifan-cl: a length-based, age-structured model for fisheries stock assessment, with application to sout Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55, 2105–2116.

Gerritsen, H.D., McGrath, D., Lordan, C., 2006. A simple method for comparing age-length keys reveals significant regional differences within a single stock of haddock (*Melanogrammus aeglefinus*). ICES J. Mar. Sci. 63, 1096–1100.

Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. J. Roy. Stat. Soc. B 55, 757–796.

ICES, 2012. Datras: Survey Descriptions. http://datras.ices.dk/Home/Descriptions.aspx.

Kristensen, K., Berg, C.W., 2012. Datras Package for R. http://rforge.net/DATRAS/.

Kristensen, K., Lewy, P., Beyer, J.E., 2006. How to validate a length-based model of single-species fish stock dynamics. Can. J. Fish. Aquat. Sci. 63, 2531–2542.

Kvist, T., Gislason, H., Thyregod, P., 2000. Using continuation-ratio logits to analyze the variation of the age composition of fish catches. J. Appl. Stat. 27, 303–319.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70, 141–159.

Maxwell, D.L., Armstrong, M.J., Beggs, S., Aldridge, J.N., 2012. Annual egg production estimates of cod (*Gadus morhua*), plaice (*Pleuronectes platessa*) and haddock (*Melanogrammus aeglefinus*) in the Irish sea: the effects of modelling choices and assumptions. Fish. Res. 117–118, 146–155. Egg Production Methods in Marine Fisheries.

Payne, M.R., Clausen, L.W., Mosegaard, H., 2009. Finding the signal in the noise: objective data-selection criteria improve the assessment of western Baltic spring-spawning herring. ICES J. Mar. Sci. 66, 1673–1680.

R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.

Rindorf, A., Lewy, P., 2001. Analyses of length and age distributions using continuation-ratio logits. Can. J. Fish. Aquat. Sci. 58, 1141–1152.

Stari, T., Preedy, K.F., McKenzie, E., Gurney, W.S., Heath, M.R., Kunzlik, P.A., Speirs, D.C., 2010. Smooth age length keys: observations and implications for data collection on north sea haddock. Fish. Res. 105, 2–12.

Toscas, P.J., Vance, D.J., Burridge, C.Y., Dichmont, C.M., Zhou, S., Venables, W.N., Pendrey, R.C., Donovan, A., 2009. Spatio-temporal modelling of prawns in albatross bay, karumba and mornington island. Fish. Res. 96, 173–187.

Wood, S., 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.