Fisheries Research 151 (2014) 91–99

Contents lists available at ScienceDirect

Fisheries Research

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

Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models

Casper W. Berg[a,b,*], Anders Nielsen[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 17 April 2013
Received in revised form 8 October 2013
Accepted 10 October 2013

Keywords:
Age-based stock assessment
Survey Indices
Generalized additive models

ABSTRACT

Indices of abundance from fishery-independent trawl surveys constitute an important source of information for many fish stock assessments. Indices are often calculated using area stratified sample means on age-disaggregated data, and finally treated in stock assessment models as independent observations. We evaluate a series of alternative methods for calculating indices of abundance from trawl survey data (delta-lognormal, delta-gamma, and Tweedie using Generalized Additive Models) as well as different error structures for these indices when used as input in an age-based stock assessment model (time-constant vs time-varying variance, and independent versus correlated age groups within years). The methods are applied to data on North Sea herring (*Clupea harengus*), sprat (*Sprattus sprattus*), and whiting (*Merlangius merlangus*), and the full stock assessments are carried out to evaluate the different indices produced. The stratified mean method is found much more imprecise than the alternatives based on GAMs, which are found to be similar. Having time-varying index variances is found to be of minor importance, whereas the independence assumption is not only violated but has significant impact on the assessments.

1. Introduction

Many fish stock assessments are based on two key sources of input data: (1) The removals from the population due to commercial fishing and (2) indices of abundance based either on catch and effort data from commercial or recreational fisheries, or from independent scientific surveys (Maunder and Punt, 2004). An index of abundance is a relative measure of e.g. the biomass or number of individuals in a population and most often proportionality is assumed:

$$I_y = qN_y$$

In this paper the focus will be on the analysis of fishery-independent survey data to create age-disaggregated indices of abundance, as well as on the subsequent use of these as input for a stock assessment model.

Several quite different approaches to the analysis of survey data exist depending on the design of the experiment, see Kimura and Somerton (2006) for a review. A popular method is based on stratified analysis, where the region of interest is divided into smaller strata and assuming that abundance is homogeneous within strata. The area weighted mean abundance is then calculated for each stratum and summed to give an index for the whole region. The probably simplest procedure uses the arithmetic mean within strata (e.g. ICES, 2012c). A slightly more refined alternative is the use of delta-distributions (e.g. Pennington, 1983), where zero values are modelled separately and the positive values are assumed to be log-normal (or Gamma) distributed. Although departures from the assumed delta-distributions can be hard to detect for small to moderate sample sizes, which may lead to biased estimates (Smith, 1990; Myers and Pepin, 1990), the mean in the delta-distribution is a more efficient estimator when the nonzero values are well approximated by a lognormal distribution, specifically it is less sensitive to the occasional huge catches that are often found in marine data sets (Pennington, 1996).

Discrete valued distributions such as the negative binomial (Kristensen et al., 2006; Cadigan, 2011) and the Log-Gaussian Cox Process (LGCP) (Lewy and Kristensen, 2009) have also been applied, but age-disaggregated indices are typically not discrete valued, so these will not be considered in this study. More recently the Tweedie distribution (Tweedie, 1984) has been suggested as an alternative to delta-distributions (Candy, 2004; Shono, 2008).

When external factors other than changes in abundance affect the catch rate, these need to be corrected for in order to obtain an unbiased index (Maunder and Punt, 2004). To this end, more advanced methods such as generalized linear models (GLMs),

* Corresponding author at: National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund Castle, 2920 Charlottenlund, Denmark. Tel.: +45 35883433.
E-mail address: cbe@aqua.dtu.dk (C.W. Berg).

0165-7836/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2013.10.005

generalized linear mixed models (GLMMs) and generalized additive models (GAMs) have previously been applied to correct for effects such as spatial position, depth, and time of day (Stefansson, 1996; Petrakis et al., 2001; Piet, 2002; Adlerstein and Ehrich, 2003; Beare et al., 2005). GAMs permit non-linear smooth relations between the response and the explanatory variables, so spatial stratification can conveniently be replaced by smooth functions of geographical coordinates (so-called splines). When stratification is used, there will be a trade-off between loss of spatial resolution due to the assumption of homogeneity within strata and problems with few or missing values when a too fine-grained stratification is used. When using GAMs, this trade-off problem is replaced with an easier problem of smoothness selection for the splines, which can be solved almost automatically using modern software packages (Wood, 2006a). Kernel density smoothing is an alternative but similar way of dealing changes in survey design (Chen et al., 2004). Sometimes the stratification is chosen such that it contains complex information about the seabed, under-water obstacles etc. (Cadigan and Tobin, 2010), which means a spatial smoother is less appropriate, but such situations will not be considered in this study.

Although useful on their own, one of the main uses of indices of abundance is to use them as input to an assessment model in combination with commercial catch data to obtain estimates of biomass and fishing mortality. The way that trawl survey data enters into many stock assessment models, can roughly be described as follows: (1) numbers-at-length data from individual hauls are preprocessed and reduced to a matrix of numbers-at-age (the index of abundance). (2) Each number in this matrix is taken as an observation of the relative abundance-at-age in the stock assessment model (often assumed independent and with constant CV through time). Although often separated for convenience, we will demonstrate that more information can be extracted from the data by combining the analyses: Instead of reducing the information from a survey to a matrix, we will use three matrices (by adding standard deviations and correlations), and by actually carrying out the stock assessments with different assumptions about the survey indices, we are provided with additional means to evaluate the impact of changes in the preprocessing step. This paper is therefore organized as follows: the first part of this paper deals with a comparison of the stratified mean method (SMM) with three variations of the Delta-GAM approach for calculating indices of abundance from trawl survey data, and second part of this paper deals with performing stock assessments using different assumptions on the error structure for the survey indices derived in the first part. Using bootstrap methodology we will show that there can be considerable positive correlations between abundance indices by age within the same year, and that including these correlations in a stock assessment model improves the model likelihood using the same number of parameters indicating improved precision in forecasts.

2. Materials and methods

2.1. Data

The data sets consist of 20 years (1992–2011) of biannual (Q1 and Q3) trawl survey data from the International Bottom Trawl Survey (IBTS) in the North Sea, downloaded from the DATRAS database (www.datras.ices.dk, data downloaded 2012-04-30). The survey is based on a stratified random design in which the North Sea is divided into so-called statistical rectangles each of size 1° longitude × 0.5° latitude, where (ideally) two hauls should be taken each quarter (each by different countries), each separated by at least 10 n.m. (ICES, 2010). Trawling is mostly performed during daytime, although some hauls are also taken during the night. Trawling time is 30 min. for all but a few hauls. The last major revision of the IBTS sampling protocol was performed in 1991, so no major changes occured in the period considered. The commercial catch-at-age data, natural mortalities, proportion mature, and weight-at-age used in the stock assessments are taken from ICES (2012a) for sprat and herring and from ICES (2012b) for whiting, and the number of age-groups used in the analyses are also the same as in these two sources. Numbers-at-length from the trawl surveys are first converted to numbers-at-age using the method described Berg and Kristensen (2012) and implemented in Kristensen and Berg (2012), see online Supplemental Materials for details. We refer the reader to ICES (2010) for a complete description of the IBTS survey design.

2.2. Stratified mean method

The survey index is calculated using the stratified mean method (SMM) by taking the mean catch per rectangle, and then the mean over all rectangles in the North Sea. This method is similar to the current way that survey indices for use in assessment are calculated for stocks in the North Sea (ICES, 2012c), and is thus a relevant baseline to compare with.

2.3. Delta-GAM and Tweedie

The delta-models consist of two parts: one that describes the probability for a non-zero catch (binomial response), and another that describes the distribution of a catch given that it is non-zero (positive continuous). The two parts may be fitted independently which eases computation. We will consider both the lognormal distribution and the Gamma distribution for the positive values. We assume the following relationship in both parts of the model between the expected response (μ, which is numbers-at-age or 1/0 for positive/non-positive catch depending on the model) and external factors:

$$g(\mu_i) = \text{Year}(i) + U(i)_{\text{ship}} + f_1(\text{lon}_i, \text{lat}_i) + f_2(\text{depth}_i) + f_3(\text{time}_i) \quad (1)$$

where Year(i) maps the ith haul to a categorical effect for each year, $U(i)_{\text{ship}} \sim N(0, \sigma_u)$ is a random effect for the vessel collecting haul i, f_1 is a 2-dimensional thin plate regression spline on the geographical coordinates, f_2 is a 1-dimensional thin plate spline for the effect of bottom depth, and f_3 is a cyclic cubic regression spline on the time of day (i.e. with same start end end point). The function g is the link function, which is taken to be the logit function for the binomial model, and the logarithm for the strictly positive responses in the Gamma and Tweedie models. The lognormal part of the delta-lognormal model is fitted by log-transforming the response and using the Gaussian distribution with a unit link. Each combination of quarter age group are estimated separately.

The length to age conversion may produce numbers that are very close to zero, which poses problems for the log-normal distribution and Gamma distribution when the mean is far from zero (Myers and Pepin, 1990; Kimura and Somerton, 2006). This can be remedied by simply treating values below some small chosen constant k as zero, and thereby move these from the positive component of the delta-distribution to the zero component (Folmer and Pennington, 2000). A preliminary analysis using histograms of residuals from the positive part of the delta models indicated that $k = 0.01$ was a reasonable choice (not the often ad-hoc chosen value of $k = 1$, which resulted in clearly non-Gaussian residuals in positive part of the delta-lognormal model). A brief sensitivity analysis using $k = 0.05$ indicated that that exact choice of k is not important though. Both the Gamma and the lognormal distributions have quadratic variance functions, $Var[y_i] = \phi\mu_i^2$, which can be checked by plotting log(sample variance) versus log(sample mean) for homogeneous groups of data (see e.g. Brynjarsdóttir and Stefánsson, 2004).

The likelihood of the delta-distributions, can be found by fitting the model for the zero and positive observations independently (see online Supplemental Material). The Tweedie distribution has been proposed as a interesting alternative to delta-distributions (Tweedie, 1984; Candy, 2004; Shono, 2008) due to its nice interpretation as a compound Poisson-gamma distribution, and its ability to handle both zero and positive values simultaneously.

The Tweedie distribution has three parameters and is a member of the exponential family with variance $Var[y_i] = \phi\mu_i^p$. For $1 < p < 2$ this distribution has support on all non-negative real numbers, i.e. a continuous density on the positive reals with a point mass in zero. In this case it is also known as a compound Poisson distribution, because it is equivalent to the distribution of $Z = W_1 + \cdots + W_N$, where W_k are independent identically distributed Gamma variables, and N follows a Poisson distribution (Candy, 2004). The fitting of GAMs with a Tweedie distribution for $1 < p < 2$ can be accomplished with the mgcv-package in R, which uses the series evaluation by Dunn and Smyth (2005) for fixed values of p. We use the same strategy as Candy (2004) and Shono (2008) to fit p, which is by optimizing the profile likelihood for this parameter. The mean value specification for the Tweedie model is chosen to be identical to that of the delta-distributions. The thin plate splines are estimated with shrinkage smoothing (Wood, 2006a, p. 160) and smoothness selection is carried out with the marginal likelihood method (Wood, 2011). Smoothness and variance parameters are estimated indepedently by quarter and age as the other parameters.

2.4. Extracting the index of abundance from the models

The usual procedure for GLMs is to use the estimated year effects as the indices of abundance (Maunder and Punt, 2004). For the delta-models this is not possible, and instead we must integrate the fitted abundance surface to obtain the index (Stefansson, 1996). However, when covariates such as depth, that are measured as part of the sampling procedure, are included in the model, it is necessary to obtain the same covariates for each point in space that we would be integrating over. Although bathymetri maps could be used for depth, other variables might not be as easily obtainable. Another possible problem with a fine-grained integration of the abundance surface, is that we might be extrapolating to areas where the model is invalid. This could be extremely deep or shallow areas, where the estimated depth effect is inappropriate, or in-trawlable areas, where the abundance is unknown. To avoid these problems, we choose the following procedure to obtain the abundance estimates for each year: (1) Divide the survey area into small subareas of approximately equal size. (2) For each sub-area where at least one haul has been taken, choose one haul position to be representative of this sub-area, e.g. the one closest to the spatial centroid of all hauls in the given sub-area. Unsampled sub-areas are left out of the analysis. (3) Take the sum over all predicted abundances using the same reference vessel (or zero in case of a random vessel effects) in the chosen haul positions. This approach has the advantage, that all covariates are immediately available given that they were collected at the chosen haul positions. The resulting grid is shown in Supplemental Fig. S1. The number of grid cells chosen corresponds approximately to the number of hauls taken in the actual surveys, and experimentation suggested that using more grid points did not change the resulting index of abundance significantly.

2.5. Estimating the statistical distribution of the indices

Often only the point estimates of the indices are used as input to stock assessment models, and additional knowledge about the distribution of the indices is thus discarded. However, various methods exist for obtaining approximations of the probability distribution of the indices from the Delta-GAM and Tweedie models, e.g. Wood, 2006a,b, as well as for the SMM ICES (2012c). Bootstrapping is a widely used technique when analytic methods are infeasible, and has previously been applied to fisheries survey data to deal specifically with the variability due to sub-sampling of age and length (e.g. Cervino and Saborido-Rey, 2006). We therefore apply bootstrapping to estimate the distribution of the indices. In the following we assume that the length distribution is known without error, and that all variability is due to sampling variability from the hauls and the age-sampling procedure. This is reasonable since the number of length samples is much greater than the number of age samples. If we sample entire hauls and thereby also the age samples in the bootstrap procedure, we will automatically incorporate the uncertainty due to the age sampling and possible correlations due to clustering of similar age groups. We can also obtain estimates of the correlation between the different age groups in our estimated indices of abundance from the bootstrap samples. The bootstrapping is carried out as follows:

Let n_y denote the number of hauls in a given year, and let a haul consist of both its associated length distribution as well as the age samples taken from that particular haul.

1. Create a bootstrap sample of the hauls, i.e. for each year sample n_y hauls with replacement from the data from the given year.
2. For each year, use the bootstrapped age data to estimate an ALK, and convert from length to age for each haul.
3. Estimate indices of abundance by age from the bootstrapped data set.

Alternatively one could resample each stratum such that the experimental design guidelines are met for each bootstrapped data set as in e.g. Cervino and Saborido-Rey (2006). In practice these guidelines are not always met however, so even though we might overestimate the variance we consider this the more conservative choice.

All parameters, including smoothing parameters, are re-estimated for each bootstrap sample. We choose to use 400 bootstrap replicates per survey to estimate standard deviations and correlation matrices for every vector of log-indices by age for a given year, $\log I_y$. Although it would be possible to use the estimated standard deviations on $\log I_y$ directly in the stock assessment model, these typically underestimates variability between I_y and the estimated values of qN_y from the stock assessment model (Maunder and Punt, 2004). This can be due to violations of some of the assumptions in the catch standardization model, the stock assessment model, or both, e.g. that that catchability q is not constant. We will discuss further the choice error structure in Section 2.6.

2.6. Stock assessment model

While internal and external consistencies provide some means to evaluate survey indices, they rely on assumptions of constant total mortality over time (Berg and Kristensen, 2012). Hence, a more appropriate way to evaluate the survey indices when mortality varies over time is to carry out full stock assessments using a statistical model, such that we utilize the extra information we have from the commercial catches and thereby estimate a time-varying mortality. State-space models allow separation of process and observation errors, which leads to an objective way of weighting each data source in the estimation process. We will therefore use a state-space stock assessment model to compare the uncertainty on the final estimates of spawning stock biomass and fishing mortality using differently derived survey indices as well as to evaluate assumptions of time-constant variance and independence among ages for the indices.

In the following we show how the observation equation for the survey indices is changed to facilitate the analysis. For a more complete description of the stock assessment model we refer to the online Supplemental Materials. Assuming independent observations, the observation equation is:

$$\log I_{a,y}^{(s)} = \log\left(Q_a^{(s)} e^{-Z_{a,y}\frac{D^{(s)}}{365}} N_{a,y}\right) + \varepsilon_{a,y}^{(s)} \tag{2}$$

where $Z_{a,y} = M_a + F_{a,y}$ is the total mortality rate, $D^{(s)}$ is the number of days into the year where the survey s is conducted, and $Q_a^{(s)}$ are catchability parameters, and $\varepsilon_{a,y}^{(s)} \sim N(0, \sigma_s^2)$.

Now, to accommodate correlated observations of $\log I_y$ we change (2) to

$$\log I_y^{(s)} = \log\left(Q^{(s)} \circ e^{-Z_y\frac{D^{(s)}}{365}} N_y\right) + \varepsilon_y^{(s)} \tag{3}$$

where $\varepsilon_y^{(s)} \sim N\left(\mathbf{0}, \mathbf{\Sigma}_y\right)$ (note again, that the scalars in Eq. (2) are replaced with vectors containing all age groups at once, and "$\circ$" denotes element-wise multiplication).

The covariance matrices $\mathbf{\Sigma}_y$ are the empirical covariance matrices from the bootstrapped log-indices described in Section 2.5. The effect of using a multivariate normal distribution rather than the independent normal distribution for each age group can be examined by inspection of the scaled residuals (e.g., Myers and Cadigan, 1995) of the survey indices:

$$\mathbf{\Sigma}_y^{-\frac{1}{2}}\left(I_y^{(s)} - \hat{I}_y^{(s)}\right)$$

Rather than working directly with the estimated covariance matrices, we choose instead to parameterize them in terms of correlation matrices and vectors of standard deviations $\mathbf{\Sigma} = \text{diag}(\sigma)\mathbf{R}\text{diag}(\sigma)$. In order to account for additional uncertainty on the survey indices other than that which is accounted for in the bootstrapping procedure, we examine the following error structures:

$$\sigma_t = \sigma, \quad \mathbf{\Sigma}_t = \sigma_t \mathbf{I} \tag{4}$$

$$\sigma_t = w_t\sigma, \quad \mathbf{\Sigma}_t = \sigma_t \mathbf{I} \tag{5}$$

$$\sigma_t = \sqrt{\sigma^2 + w_t^2}, \quad \mathbf{\Sigma}_t = \sigma_t \mathbf{I} \tag{6}$$

$$\sigma_t = \sigma, \quad \mathbf{\Sigma}_t = \text{diag}(\sigma_t)\mathbf{R}_t\text{diag}(\sigma_t) \tag{7}$$

$$\sigma_t = w_t\sigma, \quad \mathbf{\Sigma}_t = \text{diag}(\sigma_t)\mathbf{R}_t\text{diag}(\sigma_t) \tag{8}$$

$$\sigma_t = \sqrt{\sigma^2 + w_t^2}, \quad \mathbf{\Sigma}_t = \text{diag}(\sigma_t)\mathbf{R}_t\text{diag}(\sigma_t) \tag{9}$$

where the observation variance at time t, σ_t is either constant through time (σ), time-varying and proportional to the estimated standard deviation from the bootstrapping procedure, w_t, or time-varying with a total variance given by the sum of w_t and an additional constant variance component. These are the three common parameterizations of the variance found in the litterature (Maunder and Punt, 2004). In all cases σ is a parameter to be estimated, while w_t and the correlation matrices $\mathbf{R}_t$ are assumed to be known without error from the bootstrap procedure, and $\mathbf{I}$ is the identity matrix.

2.7. Model evaluation

2.7.1. AIC/BIC

To evaluate which distribution (log-normal, Gamma, or Tweedie) that provides the best fit to the individual haul data we compare the AIC and BIC values for each distribution. Since we are using GAMs, we replace the number of observations with the effective degrees of freedom (edf, see Wood, 2006a). Since age groups are estimated independently, we can simply add the log-likelihoods and edfs for each age group to obtain one the AIC/BIC for all age groups combined. Note, that we must add one to the number of parameters used for the Tweedie model to account for the estimation of p by profile likelihood.

2.7.2. Internal and external consistency

We define the internal consistency (IC) as the correlation between $\log I_{y,q,a}$ and $\log I_{y+1,q,a+1}$, and external consistency as the correlation between $\log I_{y,q_1,a}$ and $\log I_{y,q_3,a}$, where y, q, and a denotes year, quarter and age respectively. Positive consistencies implies that we can "follow the cohorts" within (IC) and between (EC) surveys, (see e.g. Berg and Kristensen, 2012 for details).

2.7.3. Areas of confidence ellipses

Since we cannot use any standard tests for comparing different data sets or in this case survey indices, we will compare the precision with which we can estimate the spawning stock biomass (SSB) and average fishing mortality in the stock assessment model. It is intuitively clear that consistent data sources with a low amount of noise will lead to smaller confidence ellipses and hence higher precision than inconsistent data sources with a larger amount of noise given time-series of equal length. Confidence ellipses for pairs of parameters are constructed from the corresponding marginals of the estimated covariance matrix for the parameters, and from pairs of derived quantities using the delta method (Oehlert, 1992). The area of a confidence ellipsis is proportional to $\sqrt{e_1 e_2}$, where e_1 and e_2 are the eigenvalues of the corresponding covariance matrix.

2.7.4. Likelihood comparison

For selection between error structures, the above criteria are not appropriate, since change to a more appropriate model for the observations will not necessarily lead to smaller confidence ellipses. Instead, likelihood based criteria such as AIC or BIC are appropriate, but since all the error structures under consideration has the same number of parameters in the stock assessment model, the choice of penalty due to the number of parameters does not matter and direct comparison of likelihoods are possible.

2.8. Software

The trawl data was handled in R (R Development Core Team, 2012) using the DATRAS package (Kristensen and Berg, 2012), and the GAM models were fitted using the `mgcv` package (Wood, 2006a). The source code for the GAM models can be obtained by contacting the corresponding author. The stock assessment models were fitted using AD model builder (Fournier et al., 2012), and the entire source code as well as data for the models are available online at www.stockassessment.org (look for the assessments named "NS-Sprat-MV","NS-Herring-MV", and "NS-Whiting-MV").

3. Results

3.1. Calculation of indices

The delta-lognormal model (DLN) gives the best overall fit to the haul data when compared to the delta-gamma and the Tweedie models using both AIC as well as BIC as the criterion (Supplemental Tables 1,5, and 9), although the resulting indices are quite similar (Fig. 1 and Supplemental Fig. S2), so the choice between proposed GAM methods does not seem crucial here. In contrast, the SMM produces quite different indices than the GAM based approaches. Based on the consistency criteria the SMM is only best in 1 out of 9 cases (Table 1), whereas the DLN model has the best consistencies overall (age specific consistencies are provided in Supplemental Tables 2–4, 6–8, and 10–12).

The improvement in consistency seems to be linked whether the index should be closer to the sample mean or the median,

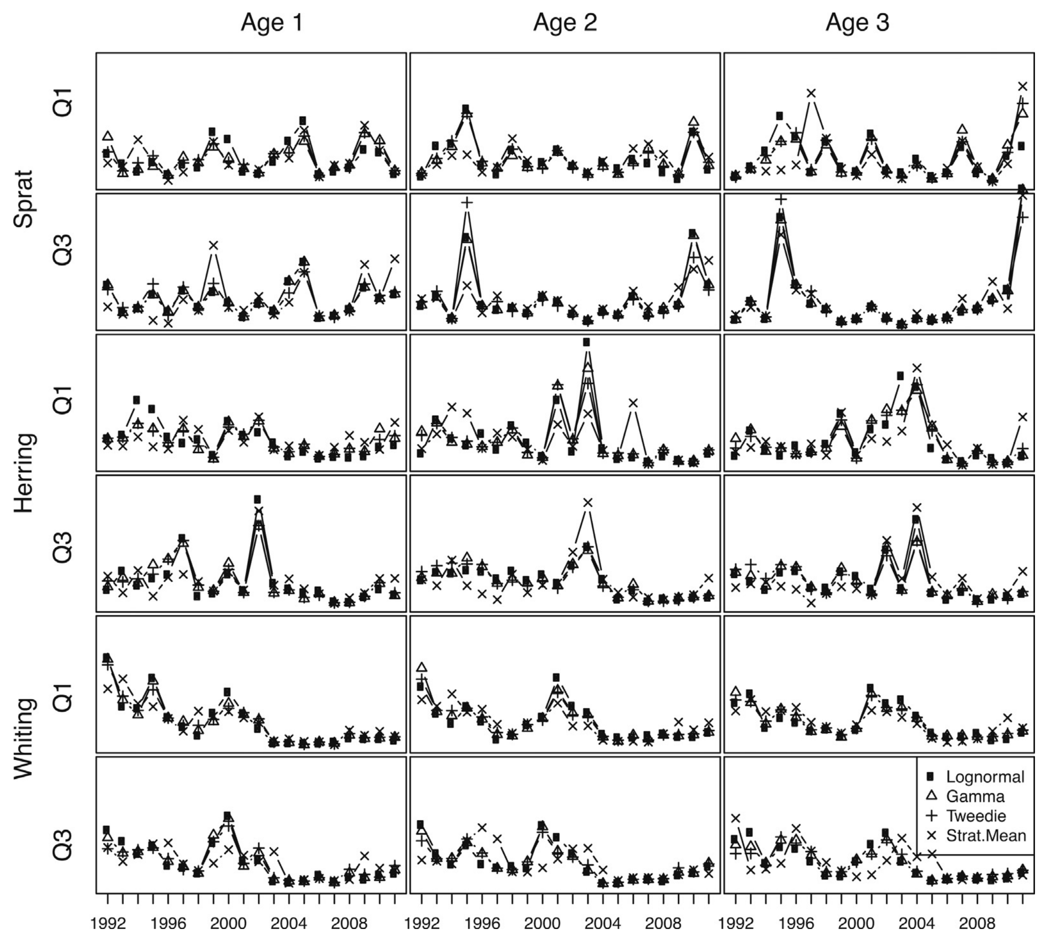

Fig. 1. Estimated relative survey indices by species, quarter, age and method. Scaling on y-axis is irrelevant since indices are relative.

rather than being due to the inclusion of covariate effects which only slightly improves the consistencies (Table 1). The largest index discrepancies between the SMM and GAM approaches were found in years where the sample median was low and the mean was high compared to the whole time-series or the reverse, with the SMM being closer to the mean while the GAM approaches tended towards the median (see species secific results in Supp. Mat.). Based on the consistency criterion it seems that the sample mean is too sensitive to large catches. Plots of log(sample variance) versus log(sample mean) produces scatters around a line with a slope of 2 (Supplemental Fig. S1), which supports the use of the lognormal and Gamma models over the SMM, which assumes a constant variance. We should however not put too much emphasis on the consistency criterion, since it is based on assumptions of constant total mortality, which is why the full stock assessments give better criteria to evaluate the precision of the indices.

Table 1
Average consistencies (average over ages 1–5). The columns “Lognormal”, “Gamma”, and “Tweedie” correspond to indices calculated using Eq. (1) with the respective distributions, whereas “Lognormal 2” is a delta-lognormal model with year effects only in Eq. (1). Best consistencies are shown in bold face.

Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean	Quarter	Species
0.27	0.18	0.21	0.13	0.08	IC(Q1)	Sprat
0.60	0.50	0.32	**0.61**	0.20	IC(Q3)	Sprat
0.52	**0.62**	0.59	0.47	0.46	EC	Sprat
0.61	0.56	0.48	0.60	0.06	IC(Q1)	Herring
0.68	0.63	**0.71**	0.62	0.66	IC(Q3)	Herring
0.59	0.54	0.53	0.50	0.46	EC	Herring
0.73	0.76	0.82	0.80	**0.88**	IC(Q1)	Whiting
0.87	0.86	0.85	0.83	0.85	IC(Q3)	Whiting
0.83	0.82	0.82	**0.84**	0.78	EC	Whiting

3.2. Stock assessments

Among the GAM-based approaches we restrict our attention to the DLN model in the stock assesments. The stratified mean method differs quite drastically from the DLN model in the last years (e.g. Sprat in Q3, Herring in Q1, cf. Fig. 1). Since the stratified mean estimates are much larger than the corresponding DLN estimates, the former yields higher estimates of SSB and lower estimates of F compared to the latter (change in [SSB,$\bar{F}$] in final year: [+37%,−11%] for sprat, [+49%,−73%] for herring, and [+29%,−43%] for whiting when using SMM, see Fig. 2). The confidence ellipses of the the SSB and $\bar{F}$ estimates in the final year (Fig. 3) also illustrate the substantial differences between the SMM and the DLN model. For all three stocks the areas of these confidence ellipses are smaller when using input from DLN model compared to the SMM under the assumption of independent observations in each age group for both methods. This implies, that there is more consistency among the data sources based on the DLN approach, since highly inconsistent and noisy data sources will result in more uncertain estimates about the state of the stock. Including correlations between age groups within years gives slightly larger confidence ellipses for all stocks (see Fig. 3). This is not too surprising, since having correlated data reduces the effective number of observations and hence should give larger uncertainties. It is worth also to notice that the confidence ellipses from the model with correlations included are still smaller than the ones from the SMM.

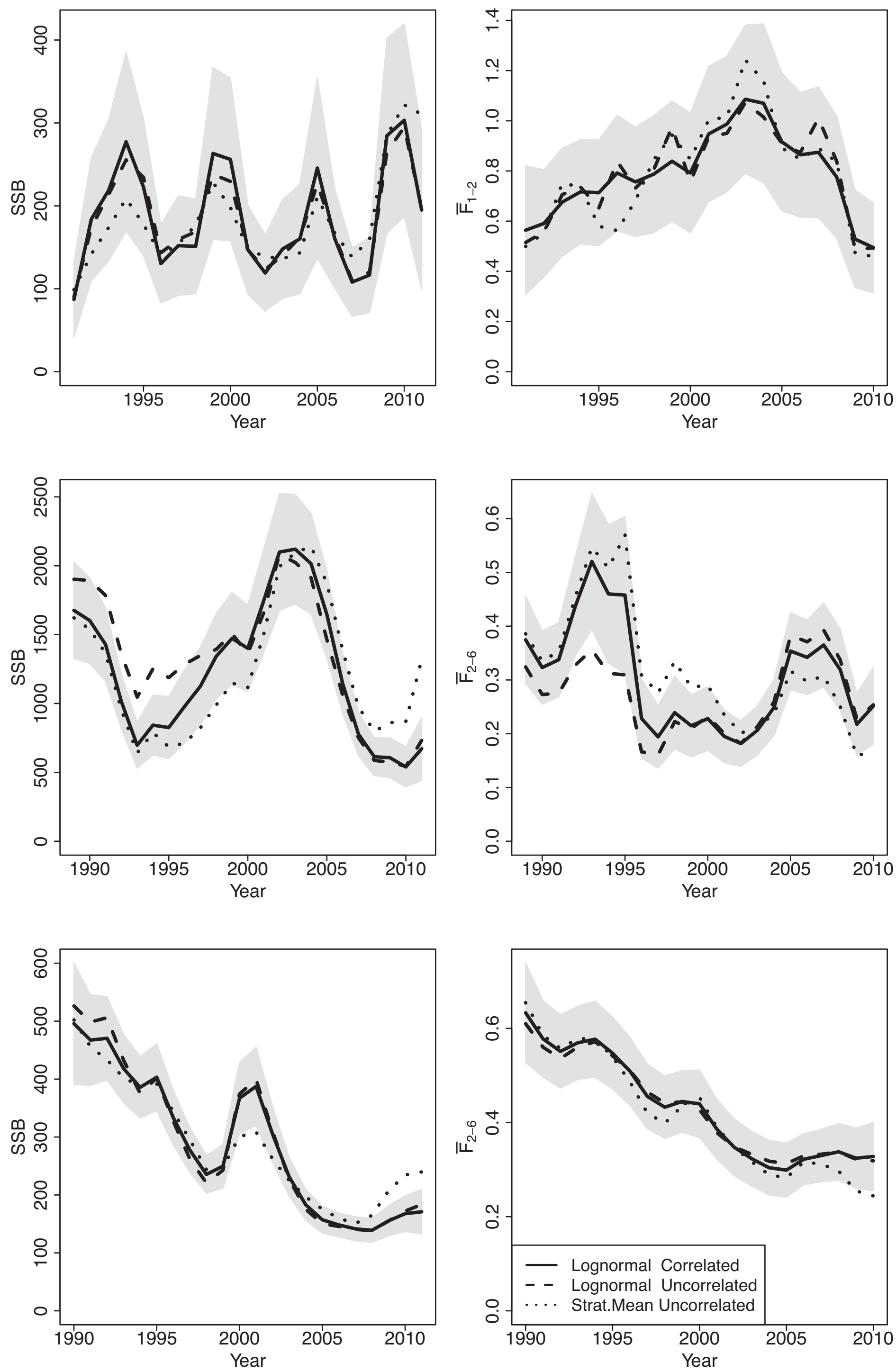

Fig. 2. Estimated spawning stock biomasses (left column) and average fishing mortalities (right column) for sprat (top row), herring (middle row), and whiting (bottom row). The shaded areas represent 95% marginal confidence intervals calculated using the lognormal indices with correlations.

Table 2
Negative log likelihood from the stock assessment models using different survey data observation error structures.

Model	(4)	(5)	(6)	(7)	(8)	(9)
Correlated	No	No	No	Yes	Yes	Yes
σ_t	σ	$w_t\sigma$	$\sqrt{\sigma^2+w_t^2}$	σ	$w_t\sigma$	$\sqrt{\sigma^2+w_t^2}$
Sprat	267.79	266.05	266.25	255.18	**253.63**	253.73
Herring	257.66	255.44	255.07	218.44	217.63	**216.34**
Whiting	163.17	165.44	163.24	131.87	134.04	**131.76**

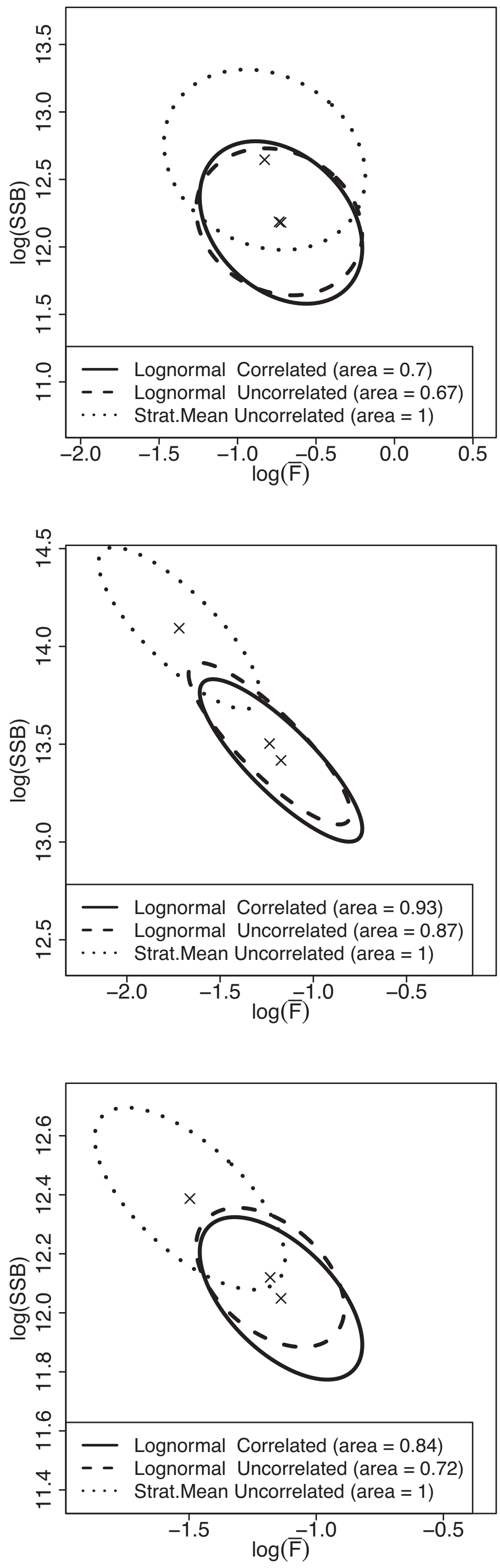

Fig. 3. 95% contour ellipses for the joint distribution of $\log \overline{F}$ and log(SSB) in the last data year (2011) for sprat (top row), herring (middle row), and whiting (bottom row).

The joint negative log-likelihoods for the different error structures (eqns 4–9) are shown in Table 2. Since smaller values are to be preferred, the errors on the logarithm of the survey abundance indices are more likely to be distributed with the assumed multivariate normal structure than to be independent among age groups within years. The significant drop in negative log-likelihood when using the correlated errors suggests that this model has significant smaller prediction errors. On the other hand, the change in likelihood by changing from constant variance (Eq. (4)) to time-varying (Eqs. (5) or (6)) is minor compared to the differences between uncorrelated and correlated errors, though there is a slight improvement for sprat and herring, but for whiting the equal variance assumption has the slightly better likelihood. The standardized residuals for the survey data obtained from the stock assessment for whiting using the SMM, DLN model and Eq. (4), and DLN model and Eq. (9) are shown in Fig. 4. The SMM residuals are clearly problematic, there seems to be problems with correlations across ages as well as time, and maybe even cohorts for the Q3 index. The DLN residuals are somewhat better, although the Q1 index seems to have an overweight of negative values in the beginning of the time-series for age 2+, and consequently an overweight of positive residuals in the end, but vice versa for age group 1. The differences between the correlated and uncorrelated DLN residuals are more subtle than between the SMM and DLN, but there are some small improvements. The corresponding residuals from the sprat and herring assessments can be found in the Supplemental Materials. For these two species there are generally less differences between the patterns in the residuals between the three models than for whiting. The Q3 herring index has a series of positive residuals in the beginning of the time-series followed by negative ones for age 2+regardless of the model for the indices. The SMM indices seem to suffer from variance inhomogeneity with larger residuals occurring in the last five years in Q1 and the first six years in Q3. The DLN indices look better with respect to homogeneity of variance. The residuals from the sprat assessments look very similar between methods, and there are no immediately apparent problems. The similarity between using uncorrelated versus correlated observations is expected here, since the correlations found for sprat were minor compared to those for herring and whiting (Supp. mat.).

The distribution maps and internal/external consistencies suggested that the Q3 indices described the sprat and herring stocks better than the Q1 indices (Supplemental Figs. S3, S4, S14, and S15). This can be investigated further by comparing the estimated precisions (inverse variances) from the stock assessment model (Supp. Table 13), and especially for herring we must conclude that the Q3 index is the more informative of the two.

4. Discussion

We have considered statistical aspects in the joint process of calculating indices of abundance by age and using them as input for a stock assessment model.

The first problem was how to derive the indices from individual hauls from trawl survey data. To this end we compared the method of stratified sample means with three GAM-based alternatives: a delta-lognormal, delta-gamma, and Tweedie model. The delta-lognormal was found to provide the overall best fit, but the largest contrast was found between the SMM and the alternatives based on GAMs, whereas the latter produced rather similar indices. The concistency criteria showed that the SMM performed rather poorly compared to the GAM approaches.

The advantage of using the proposed DLN model over the SMM was further confirmed by carrying out stock assessments using the indices – higher precision on estimates of SSB and $\overline{F}$ was obtained

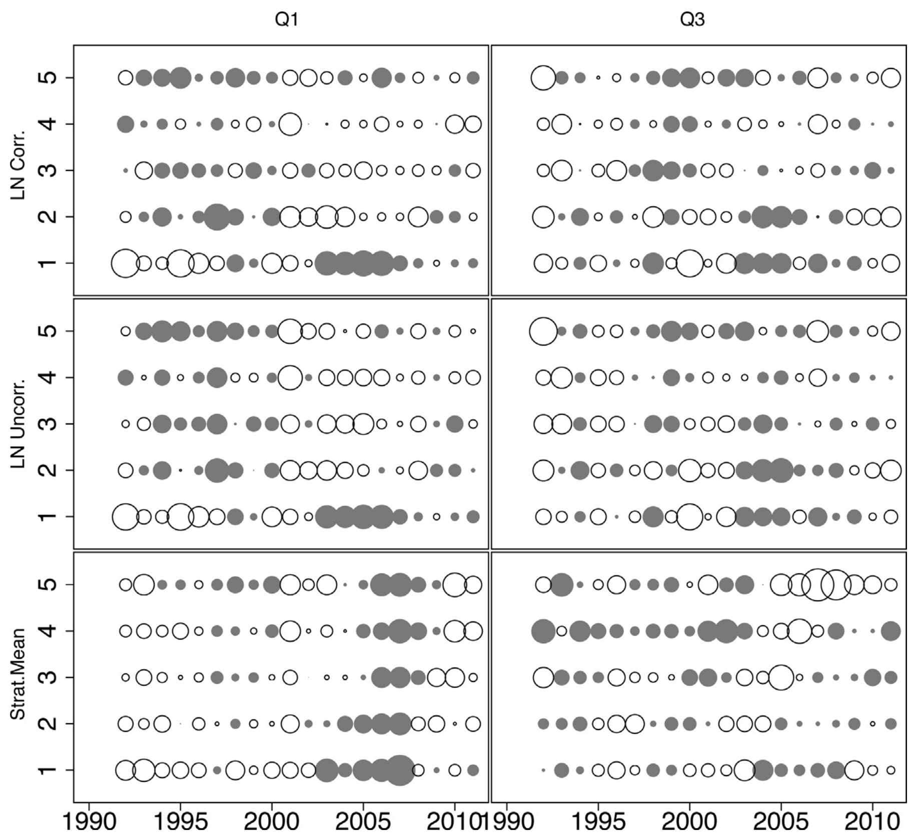

Fig. 4. Whiting: survey standardized residual plots from the fitted assessment models.

with the DLN indicating that it provided the more accurate index. The SMM method examined in this study is similar to the one used as basis for management advice in ICES for most species, including the three species considered, which makes this finding extremely relevant.

Although indices from the SMM are simple to compute and do not need to be updated back in time once new data are available, it has certain shortcomings compared to a proper statistical model such as the DLN model. One problem, which has long been known, is the sensitivity to extraordinary large catches (Pennington, 1996). When such huge catches occur at the end of the time-series, as they did in our examples, it will lead to overly optimistic biomass estimates. Another problem is the inability of the SMM to control for effects such as depth, gear, environmental factors etc. on the catch rate, although these effects were, although statistically significant, minor compared to the effect of large catches for the cases in this study. These problems are all well described in the literature (Pennington, 1996; Stefansson, 1996; Maunder and Punt, 2004), although the majority have applied area stratification and/or GLM methodology in place of GAMs to deal with spatial heterogeneity and uneven sampling effort. Recent studies (Berg and Kristensen, 2012; Maxwell et al., 2012) have advocated the use of GAMs and 2D-splines to replace area stratification, which, in addition to providing more accurate estimates using fewer parameters, alleviates the modeller of problems with selecting the strata and possibly the subsequent problems of having few or missing data points for some combinations of years and strata. The second problem addressed in this paper was whether it was reasonable to assume uncorrelated observations within years and/or time-constant variances for the indices. Utilizing the estimated standard deviations on the log-indices from the DLN model in the assessment model did not change the results much regardless of the variance parameterization. One obvious reason could be, that the estimated SDs were rather constant through time. Also, there is a fundamentally different interpretation of the SDs from the DLN models and those from the assessment model (Maunder and Starr, 2003), i.e. the main source of the assessment SD is the discrepancy between the measured survey abundance and the true population, and not the measuring uncertainty from the experiment itself. In contrast to the estimated SDs, including the estimated correlations improved the likelihood of the data substantially, and did prove to have some impact on the assessment results. Although our motivation is the same as in Walters and Punt (1994) and Myers and Cadigan (1995), our method differs in that we are using bootstrap estimates from DLN model to estimate the correlations as opposed to estimating the correlation structure within the assessment model. When estimated within the assessment model using indices only, the information from the individual hauls is not included, and therefore the only option is to assume a simple correlation structure with few extra parameters due to the limited number of data points. An important reason why including correlations among age groups can improve assessments is given in Myers and Cadigan (1995): Say we observe higher abundances than expected in the final year of the older age groups and these are known to be positively correlated with the youngest age group, then we can utilize this information to heighten the estimates of the youngest age group even though this cohort has only been observed once. However, the estimated correlations in this study were found to be greatest between the oldest age groups whereas the estimates of the youngest age group were often nearly uncorrelated with those of the older age groups. Assuming a common correlation among age groups as in Walters and Punt (1994) and Myers and Cadigan (1995) in this situation would be problematic – it could improve the estimates of older age groups at expense of the youngest, which is undesirable for forecasting purposes.

We do not advocate that Eq. (1) is appropriate for all survey time series – e.g. sufficient spatio-temporal overlap in sampling between different vessels must be present to ensure identifiability of the parameters, and the stock spatial distribution must be relatively stable in the period considered. Although the precise formulation of Eq. (1) should depend on the stock under examination, the framework is general and may be easily extended to account for e.g. shifts in spatial distribution over time. 3D smoothing in spatial coordinates and depth may also be considered as in Cadigan and Chen (2010). We did also not consider alternatives to the ML-method recommended by Wood (2011) for the GAM smoothness selection, but recommend that future research should address this issue.

We should also note, that several alternative methods exist for dealing with the problems discussed in this study. In so-called integrated assessment models (Maunder and Punt, 2004) the raw data is handled within the assessment model, i.e. age-length relationships and index calculations are computed as part of the assessment instead of as a preprocessing step like in the present study. The advantage of this is that the uncertainty from the preprocessing step can be easily propagated onto the final assessment output (at least when this is formulated as a maximum likelihood problem). However, assessment models are often, as in this study, quite complex and computer intensive, so reducing the dimensionality of the data in advance by preprocessing can be necessary. This study shows, that in order to avoid too much loss of information, it is important to consider not only point estimates from the preprocessing step, but to include standard deviations and correlations as well, and to continuously re-evaluate such preprocessing steps. Also, we must recommend that the SMM is no longer used to create indices of abundance for the highly aggregated stocks considered in this study.

Acknowledgements

The authors wish to thank all staff involved in the IBTS data collection, ICES for providing easy access to the data, and finally Anna Rindorf, Peter Lewy and two anonymous reviewers 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.2013.10.005.

References

Adlerstein, S., Ehrich, S., 2003. Patterns in diel variation of cod catches in north sea bottom trawl surveys. Fisheries Research 63, 169–178.

Beare, D., Needle, C., Burns, F., Reid, D., 2005. Using survey data independently from commercial data in stock assessment: an example using haddock in ICES Division VIa. ICES Journal of Marine Science: Journal du Conseil 62, 996–1005 http://icesjms.oxfordjournals.org/content/62/5/996.full.pdf+html

Berg, C.W., Kristensen, K., 2012. Spatial age-length key modelling using continuation ratio logits. Fisheries Research 129-130, 119–126.

Brynjarsdóttir, J., Stefánsson, G., 2004. Analysis of cod catch data from Icelandic groundfish surveys using generalized linear models. Fisheries Research 70, 195–208.

Cadigan, N., Chen, J., 2010. Kernel regression estimators for nonparametric model calibration in survey sampling. Journal of Statistical Theory and Practice 4, 1–25.

Cadigan, N., Tobin, J., 2010. Estimating the negative binomial dispersion parameter with highly stratified surveys. Journal of Statistical Planning and Inference 140, 2138–2147.

Cadigan, N.G., 2011. Confidence intervals for trawlable abundance from stratified-random bottom trawl surveys. Canadian Journal of Fisheries and Aquatic Sciences 68, 781–794, http://www.nrcresearchpress.com/doi/pdf/10.1139/f2011-026.

Candy, S., 2004. Modelling catch and effort data using generalised linear models, the tweedie distribution, random vessel effects and random stratum-by-year effects. CCAMLR Science 11, 59–80.

Cervino, S., Saborido-Rey, F., 2006. Using the bootstrap to investigate the effects of varying tow lengths and catch sampling schemes in fish survey. Fisheries Research 79, 294–302.

Chen, J., Thompson, M.E., Wu, C., 2004. Estimation of fish abundance indices based on scientific research trawl surveys. Biometrics 60, 116–123.

Dunn, P.K., Smyth, G.K., 2005. Series evaluation of Tweedie exponential dispersion model densities. Statistics and Computing 15, 267–280.

Folmer, O., Pennington, M., 2000. A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. Fisheries Research 49, 165–178.

Fournier, D.A., Skaug, H.J., Ancheta, J., Sibert, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., 2012. AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 27, 233–249.

ICES, 2010. Manual for the International Bottom Trawl Surveys Revision VIII.

ICES, 2012. Report of the Herring Assessment Working Group for the Area South of 62 deg N (HAWG). ICES Document CM 2012/ACOM:06.

ICES, 2012. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). ICES Document CM 2012/ACOM:13.

ICES, 2012. Report of the Workshop on Implementation in DATRAS of Confidence Limits Estimation of, 10–12 May 2006. ICES Headquarters, Copenhagen.

Kimura, D.K., Somerton, D.A., 2006. Review of statistical aspects of survey sampling for marine fisheries. Reviews in Fisheries Science 14, 245–283.

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. Canadian Journal of Fisheries and Aquatic Sciences 63, 2531–2542, http://www.nrcresearchpress.com/doi/pdf/10.1139/f06-135.

Lewy, P., Kristensen, K., 2009. Modelling the distribution of fish accounting for spatial correlation and overdispersion. Canadian Journal of Fisheries and Aquatic Sciences 66, 1809–1820.

Maunder, M., Starr, P., 2003. Fitting fisheries models to standardised CPUE abundance indices. Fisheries Research 63, 43–50.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research 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. Fisheries Research 117-118, 146–155, Egg Production Methods in Marine Fisheries.

Myers, R.A., Cadigan, N.G., 1995. Statistical analysis of catch-at-age data with correlated errors. Canadian Journal of Fisheries and Aquatic Sciences 52, 1265–1273.

Myers, R.A., Pepin, P., 1990. The robustness of lognormal-based estimators of abundance. Biometrics 46, 1185-1192+.

Oehlert, G.W., 1992. A Note on the Delta Method. American Statistician 46, 27–29.

Pennington, M., 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics, 39.

Pennington, M., 1996. Estimating the mean and variance from highly skewed marine data. Fishery Bulletin (Washington D C) 94, 498–505.

Petrakis, G., MacLennan, D.N., Newton, A.W., 2001. Day-night and depth effects on catch rates during trawl surveys in the North Sea. ICES Journal of Marine Science 58, 50–60.

Piet, G.J., 2002. Using external information and GAMs to improve catch-at-age indices for North Sea plaice and sole. ICES Journal of Marine Science: Journal du Conseil 59, 624–632 http://icesjms.oxfordjournals.org/content/59/3/624.full.pdf+html

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.

Shono, H., 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research 93, 154–162.

Smith, S.J., 1990. Use of statistical models for the estimation of abundance from groundfish trawl survey data. Canadian Journal of Fisheries and Aquatic Sciences 47, 894–903.

Stefansson, G., 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Science 53, 577–588.

Tweedie, M.C.K., 1984. An index which distinguishes between some important exponential families. In: Ghosh, J.K., Roy, J. (Eds.), Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. Indian Statistical Institute, Calcutta, pp. 579–604.

Walters, C., Punt, A., 1994. Placing odds on sustainable catch using virtual population analysis and survey data. Canadian Journal of Fisheries and Aquatic Sciences 51, 946–958.

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

Wood, S.N., 2006b. On confidence intervals for generalized additive models based on penalized regression splines. Australian & New Zealand Journal of Statistics 48, 445–464.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 3–36.