Survey Index and Mean Weight-at-age Calculations for North Sea Cod from NS-IBTS data

Casper W. Berg

January 27, 2015

1 ALKs

Spatially varying age length keys are estimated using the methodology described in [1]. Numbers-at-age are then calculated using the observed numbers-at-length and the estimated ALKs. This methodology was found to give internal consistency in survey indices for haddock when compared to current standard approach of estimating ALKs that are constant within "Roundfish" (RF) areas in [1]. It avoids ad-hoc borrowing of samples from neighbour RF areas, when certain age groups are missing, and it provides an objective fill-in procedure for missing length groups also. The methodology has been implemented in the DATRAS package with full source code available [3]

2 Survey Indices

Survey indices by age and area are calculated using the methodology described in [2], although we consider a broader class of equations describing the observed abundance in each haul. While [2] considered a time-invariant spatial effect and a data set consisting almost exclusively of 30 min hauls, the following model classes contains a space-time smoother, which allows for smooth changes in the spatial distribution of each age group over time, as well as haul duration effect. The space-time smoother is useful in the situation where survey indices need to be calculated by sub-area; a time-invariant spatial effect would lead to the same abundance trend in all subareas. Rather than dividing the data by subarea and estimating survey indices from the dis-aggregated data, the space-time smoother allows the data to be modelled jointly, and the usual problems with missing data (mostly missing ages) are avoided. In quarter 1, NS-IBTS consists almost only of hauls taken with the GOV gear, so gear effects need not be included. However, the data from Q3 consists of 4 gear types, but from 1992 and onwards only ABD and GOV, hence the first year (1991) is not included. The following equation describes the maximal model considered for both the presence-absence and the positive parts of the model:

$$
\begin{aligned}
g(\mu_i) =& \mathrm{Year(i)} + \mathrm{U(i)}_{\mathrm{ship}} + f_1(\mathrm{Year}_i, \mathrm{lon}_i, \mathrm{lat}_i) & (1)\\
&+ f_2(\mathrm{depth}_i) + f_3(\mathrm{time}_i) + \mathrm{Gear(i)} + \log(HaulDur_i) & (2)
\end{aligned}
$$

where $\mathrm{U(i)}_{\mathrm{ship}} \sim N(0, \sigma_u^2)$. where Gear(i) maps the ith haul to a categorical gear effect for each age group, and $\mathrm{U(i)}_{\mathrm{ship}} \sim N(0, \sigma_u^2)$ are random effects.

f_1 is a 3-dimensional tensor product spline (a 2D thin-plate spline for space $\times$ a 1D cubic spline for time), 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. The Gamma part of the delta-Gamma model is fitted with a log link. Each combination of quarter age group are estimated separately. The fitted models are then used to sum the expected catches over a fine grid by year,age, and subarea to obtain the survey index. Nuisance variable such as gear, time-of-day and haul duration are corrected for in this process.

The whole procedure consists of the following steps:

1. Apply spatial ALK
2. Fit model for catch-at-age by age and quarter
3. Select grid of haul positions
4. Predict abundance on grid by year (using reference vessel, time-of-day etc).
5. Sum of grid points = index
6. Approximate the full multivariate distribution of the index (optional)

The 6. step above was explored in for other species than cod in [2], where it was found that survey indices for neighbouring age-classes in the same year can be highly correlated, and that failing to account for these correlations led to significantly different assessments, and was much more important than accounting for time-varying CVs within age groups.

2.1 Model selection

Model selection can be based on likelihood ratio tests or information criteria such as AIC. This may not be the best way however because these are conditional on the model being correct, and the spatio-temporal smoother (which is not to be "standardized out") can be highly correlated with other variables that are clustered in space and/or time such as the vessel effects (which *are* to be "standardized out"). Overfitting may thus be an issue, and model selection should perhaps be based on other measures such as internal/external consistencies or even better, based on the prediction error from the stock assessment model.

Another way to look at this problem is as a bias-variance trade-off: The simpler models will have less variance at the expense of some bias, i.e. omitting the vessel effect will give a little bias since vessels are probably not truly equal, but the estimates of the vessel effects may be so uncertain and confounded with true changes in abundance that they are better left out of the model.

3 Mean Weight-at-age

Mean weight-at-age cannot be calculated from the CA-records alone, because these are length-stratified samples, and hence rare lengths are over-represented. Instead, both the observed length-distribution (HL-records) as well as the length-weight relationship found in the CA-records must be used.

For each year a length-weight relationship $W(L) = aL^b$ is estimated by fitting the log-transformed model as a linear model. Then the distribution of length given age $P(l_i|a)$ is calculated from an age-length key $P(a|l)$ (just the raw proportions of age per length group) and the length distribution $P(l)$ using Bayes rule:

$$P(L = l_i|A = a) = \frac{P(A = a|L = l_i)P(L = l_i)}{\sum_k P(A = a|L = k)P(L = k)}$$

and the mean weight at age is found by applying the formula for conditional expectation:

$$\mathbb{E}(W(L)|A = a) = \sum_k W(k)P(L = k|A = a)$$

The resulting estimates may be contain considerable random sampling noise from the ALK and the length-weight relationship. To avoid spurious fluctuation in SSB estimates derived from these weights, the mean weight-at-age is smoothed over time by applying a GAM for each age group:

$$w_{a,y} = s_a(y)$$

with gaussian noise.

References

[1] Casper W Berg and Kasper Kristensen. Spatial age-length key modelling using continuation ratio logits. *Fisheries Research*, 129:119–126, 2012.

[2] Casper W Berg, Anders Nielsen, and Kasper Kristensen. Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. *Fisheries Research*, 151:91–99, 2014.

[3] Kasper Kristensen and Casper W. Berg. Datras package for r. `http://rforge.net/DATRAS/`, 2012.