ICES CM 2009/N:21 Ref no 1008. Incorporating survey variance in sequential population analysis

Noel Cadigan, Fisheries and Oceans

1. Abstract

We derive some basic statistics that describe the variability of a survey index derived from stratified random sampling for several Northwest Atlantic fish stocks. The variability is expressed as a function of population abundance and is based on a Negative Binomial (NB) distribution assumption for trawl catches. Diagnostics that support this assumption are presented. However, maximum likelihood estimates of the NB over-dispersion parameter based on a stratum-effects model can have severe bias, and an alternative estimator is shown to give much better results. We also show how the survey variance component can be incorporated into stock assessment models like ADAPT or XSA. Interestingly, this results in an estimation procedure that is more similar to the implicit and intuitive weighting that many fisheries scientists use to track cohorts in survey data by focusing on the ages that tend to be caught well, whereas ADAPT or XSA tend to give higher weight to ages not well caught.

2. Statistical distribution of survey catches

•We assume that survey trawl catches (R_i, $i=1,\ldots,n_h$) within strata ($h=1,\ldots,H$) are independent and identically distributed NB random variables, with mean $\lambda_h = E(R_{i\epsilon h})$.

•This is a stratum-effects model.

•The NB distribution is often suggested to be appropriate for modelling the variability in catches (see Gunderson, 1993).

•The NB model (ξ) variability is $Var_\xi(R_{i\epsilon h}) = \lambda_h + \lambda_h^2/k$.

2.1. NB model (ξ) diagnostics

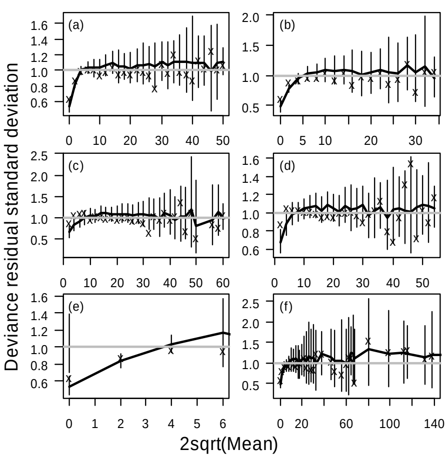

Figure 1. Standard deviations (SD) of binned NB deviance residuals (x's) versus $2\hat{\mu}^{1/2}$. The vertical lines denote parametric bootstrap percentile confidence intervals for the SD's, while the solid black lines show the expected values. The grey lines at one denotes the nominal expected SD. The stocks are a) 2J+3KL cod, b) 3Ps cod, 3LNO American plaice c) fall, d) spring, e) 3Ps Striped wolffish, f) 3LNO Shrimp (fall).

Conclusion: The NB is appropriate for these stocks.

3. Statistical Moments of Survey Indices

•Survey index is an estimate of the average catch at all (i.e. N) tow stations in the survey region, $\bar{R} = N^{-1}\sum_{i=1}^{N} R_i$

•A common (and unbiased) estimator of $\bar{R}$ is $\hat{\bar{R}} = \sum_{h=1}^{H} W_h \bar{r}_h$

•where $\bar{r}_h$ is the sample mean in stratum h and W_h is the fraction of total area in strata h.

•Can show that $E(\hat{\bar{R}}) = E_\xi E_D(\hat{\bar{R}}) = \sum_{h=1}^{H} W_h\lambda_h \equiv \lambda$

•The expectation E_ξ is with respect to (wrt) the NB model, and E_D is wrt the survey sampling probability.

•A good approximation of the total variance (wrt to ξ and D), especially for proportional allocation, is

$$Var(\hat{\bar{R}}) \doteq n^{-1}\left(\lambda + \frac{\lambda^2}{k_p}\right)$$

•where n is the total sample size and k_p is an over-dispersion parameter ($k_p < k$ in Section 2).

•Use a moment-type estimator for k_p, based on the adjusted extended quasi-likelihood estimate of NB k.

The mle is very biased for highly stratified survey data

4. Additional variation due to catchability

•A time series of age-based survey indices may have additional variability due to inter-annual variations in the fraction of the stock available to the survey (Q).

•Survey abundance (λ_{ay}) is related to total stock abundance (N_{ay}) through Q, $\lambda_{ay} = Q_{ay}N_{ay}$.

•Conditional survey moments: $E(R_{ay}|Q_{ay}) = Q_{ay}N_{ay}$

$$Var(R_{ay}|Q_{ay}) = n_y^{-1}\left(Q_{ay}N_{ay} + \frac{Q_{ay}^2N_{ay}^2}{k_y}\right)$$

•If the survey covers a fraction q_a of the range of age a fish and if fish are distributed completely at random then $E(Q_{ay}) = q_a$ and $Var(Q_{ay}) = \frac{q_a(1-q_a)}{N_{ay}}$

•A first approximation to account for the additional variability caused by schooling is $Var(Q_{ay}) = \frac{\phi q_a(1-q_a)}{N_{ay}}$

•Unconditional moments:

$$E(R_{ay}) = E_Q\{E(R_{ay}|Q_{ay})\} = q_aN_{ay} = \mu_{ay}$$

$$Var(R_{ay}) = \frac{\mu_{ay}}{n_yk_y}\{\upsilon_y(\phi) + \mu_{ay}\}$$

$$\upsilon_y(\phi) = \{k_y + \phi(1-q_a)(1+n_yk_y)\}$$

•k_y is estimated from surveys, but need a stock model to estimate q_a, φ and N_{ay}'s.

5. Model estimation (via Quasi-likelihood)

•Can use $Var(R_{ay})$ to develop a quasi-likelihood (e.g. McCullagh and Nelder, 1989) fit function to estimate model parameters.

•Annual survey variability is incorporated in $Var(R_{ay})$ via k_y. Annual survey sample sizes are also incorporated.

•The deviance (D) for an age a index, r_{ay}, is

$$D(r_{ay},\mu_{ay},\phi) = \frac{2n_yk_y}{\upsilon_y(\phi)}\left[r_{ay}\log\left\{\frac{r_{ay}}{\mu_{ay}}\right\} + \{r_{ay}+\upsilon_y(\phi)\}\log\left\{\frac{r_{ay}+\upsilon_y(\phi)}{\mu_{ay}+\upsilon_y(\phi)}\right\}\right]$$

•For fitting model parameters θ (i.e. q_a's and parameters for N_{ay}'s) and φ we use the extended quasi-likelihood function:

$$\Lambda(\theta,\phi) = -\tfrac{1}{2}\sum_{a,y}[\log\{2\pi V(r_{ay})\} + D(r_{ay},\mu_{ay},\phi)]$$

•where $V(r_{ay}) = Var(R_{ay})$ evaluated at $\mu_{ay} = r_{ay}$.

•An adjusted to $V(r_{ay})$ is required to accommodate indices with values of zero.

•The fit function can be extended for multiple surveys.

6. Separable total mortality survey-only model (~SURBA) for 3Ps cod.

•Use a simple model for surveys in which total mortality (Z) is modelled as age×year effects, $Z_{ay} = f_y \times s_a$.

•Cohort abundance at age is estimated as recruitment × cumulative mortality.

•Model parameters: N_{oy}'s (for every cohort), f_y's, s_a's, q_a's, and φ.

•Can show that φ and q_a's are confounded so we fix φ at trial values.

•A highly parameterized model. We used some shrinkage to smooth variations in f_y's, s_a's, q_a's.

•Compare with common log error sums of squares (LN) estimation, in which q_a's are usually specified (i.e. fixed).

7. Results

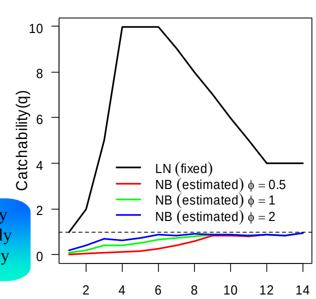

Figure 2. NB estimates of q_a's for three values of φ. Fixed values of q_a's (LN) used in a recent stock assessment are also shown.

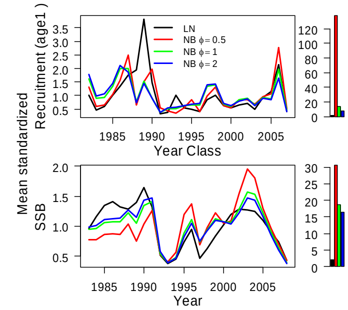

Figure 3. Estimates of stock size, standardized (i.e. divided by) the mean of each series. These means are shown in the right-hand panels.

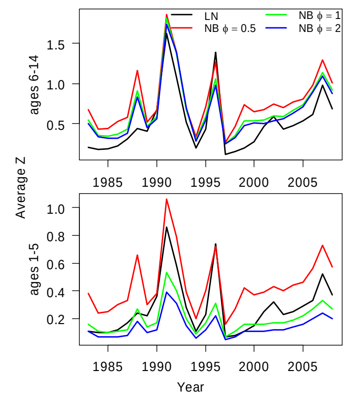

Figure 4. Average total mortality (Z) for two age groups.

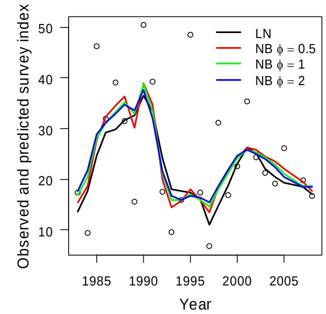

Figure 5. Fits to survey indices (aggregated over ages).

7. Sensitivity to ages used.

Survey catches are highest at age 3, and lowest at age 14

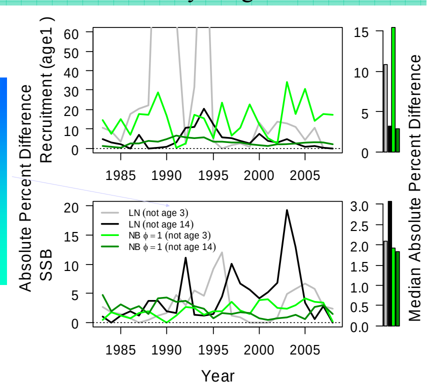

Figure 6. Change in estimates when an index-age is not used for estimation.

Conclusions: 1) NB estimates of recruitment are more sensitive to deleting age 3 than LN estimates. 2) LN estimates of SSB are more sensitive to deleting age 14 than NB estimates.

8. Some summary points

•The NB fit function we propose includes annual survey variability and samples sizes.

•The $Var(Q)$ model adds "information" about q_a's. It seems possible to estimate the q pattern even for a SURBA-type model.

•The NB fit function accommodates zero indices.

•NB estimates seem less sensitive to poorly samples ages.