# Stock reduction and virtual population analysis using catch at length data

Brett van Poorten and Carl Walters

June 2013

## Abstract

## Introduction

Modern stock assessments typically attempt to fit population dynamics models to catch at age and catch at length data, in hopes of extracting information from these data about age-size vulnerability and fishing mortality patterns (ref). In cases where age data are lacking, software like MULTIFAN attempt to obtain estimates only from size distribution data while assuming information about variability in length at age, while perhaps also attempting to recover information about changes in body growth patterns. Some older assessment methods attempt to put aside the length frequency data, by converting these data to age compositions using age-from-length tables, perhaps using iterative methods to estimate proportions of fish at age for each length interval (e.g. Shirippa and Goodyear, ref.). It is typical for assessment results from assessment models to show substantial deviations between predicted and observed length distributions of catches, reflecting both sampling variation in the length composition data and incorrect assumptions about stability of both growth and vulnerability patterns.

Here we suggest an alternative approach to assessment modeling that begins by assuming that the assessment model should exactly reproduce the observed catch at length distribution. This is similar to the classical assumption in virtual population analysis that reconstructed numbers at age should exactly match observed catch at age data, or the suggestion by Schnute (ref) that statistical catch at age models might best be run in a “conditioned on catch” format by subtracting observed catches at age from modeled numbers at age in estimation of numbers at age over time. The suggested approach may have two key advantages over statistical catch at age and/or catch at length models: (1) it does not require estimation of age or size vulnerability schedules, and (2) catch at length data are commonly available for every year, even when age composition sampling has not been conducted.

## Prediction and reconstruction of catch at length and age

Development of the method begins with the common approximation that harvesting occurs over a short, discrete season in each time step (year or shorter time period) and that natural survival rate is stable over time, so that numbers at age at the start of each time step propagate as

N(a+1,t+1)=N(a,t)S(a)[1-U(a,t)] (1)  
for ages a=1,…,na and times t=1,…,T. The next key approximation is that vulnerability to harvest depends on fish size rather than age, so that U(a,t) consists of a sum of exploitation rates for age-a fish that are in different length intervals L:

(2).

Here, length intervals L=1,…,nl represent fish with lengths in the intervals L1(L) to L2(L), where these L-specific length intervals need not be of equal width. P(L|a,t) is the proportion of fish of age a that are expected to be in length interval L at the time of harvest in step t (NB, P(L|a,t) is *not* the proportion of length L fish that are age a, or the proportion of age a fish in the catch that are in length interval L; it is the population proportion.). U(L,t) is the exploitation rate of length interval L fish in time t, given by:

U(L,t)=C(L,t)/N(L,t) (3)  
where C(L,t) is the estimated total catch of fish of length interval L in step t and N(L,t) is the estimated total number of fish vulnerable to harvest at the start of step t. Note that N(L,t) is a sum over ages of fish of numbers at age times (population!) proportions at age that are in length interval L, i.e.

(4).

In combination, eqs. (1)-(4) constitute a “complete”stock reduction (conditioned on catch) model for numbers at age, in the sense that a table of N(a,t) can be calculated forward over time given boundary conditions (as parameters to be estimated) for initial age structure N(a,1) a=1,…,na and recruitments N(1,t), t=2,…,T. For each t, we first calculate the N(L,t) using eq. (4), then estimate the U(L,t) using N(L,t) and observed C(L,t) in eq. (3), then calculate the U(a,t) using eq. (2) and apply these to calculate N(a+1,t+1) for t=2,…,T and a=2,…,na.

Interestingly, the same model can be used for virtual population analysis to back-calculate numbers at age over time, without assuming knowledge of catches at age. In this case, we set (as parameters to be estimated) the terminal boundary conditions N(a,T) for a=1,…,na for the last year, and post-harvest abundance N\*(na,t) for age na in years 1 to T-1. By post-harvest abundance we mean N\*(a,t)=N(a,t)\*[1-U(a,t)], i.e. the survivors after fishing but before natural survival S(a). Note that having calculated or set N(a,t) for any year (beginning at T), we can calculate post-harvest abundance for ages 1,…,na-1 the previous year from the simple inverse survival equation

N\*(a-1,t-1)=N(a,t)/S(a-1) (5).

Eq. (5) along with the boundary (to be estimated) N\*(na,t) values give total numbers at age just after harvest in year t, given abundance at age before harvest the next year. Next, note that we can calculate pre-harvest total abundance at length for step t, N(L,t) from the catches at length C(L,t) and post harvest abundances, as

(6).

That is, N(L,t) is taken to be observed catch C(L,t) plus a sum of post-harvest numbers contributions to the length L population from fish of different ages. Next, we simply use eq. (3) to calculate U(L,t) as the ratio of C(L,t) to N(L,t), use eq. (2) to calculate U(a,t) from the resulting U(L,t), and estimate N(a,t) by expanding from N\*(a,t) using the estimated U(a,t):

N(a,t)=N\*(a,t)/[1-U(a,t)] (7).

One way to think about eqs. (6)-(7) is that they apportion the observed catches C(L,t) to ages of fish so that N(a,t) consists of N\*(a,t) calculated from the next age and time plus a catch contribution dependent on C(L,t) and P(L|a,t) as well as on the abundances of other cohorts that contributed to C(L,t).

In order to apply either the SRA equations (1)-(5) or the VPA equations (5)-(7), we need to solve two simple problems and one very difficult one. First, we need to assess the size distribution probability matrices P(L|a,t); the simplest estimate would be to assume that sizes at age are normally distributed around mean lengths at age given by some growth equation, with standard deviations at age proportional to length (constant CV for length). In this case, P(L|a,t) is given by the difference between the cumulative normal distribution evaluated at L2(L) and the cumulative normal evaluated at L2(L). Second, we need to estimate the total catch at length C(L,t), typically from data on total weight yield Y(t) and observed proportions p(L,t) of fish by length interval. This can be done in two steps: (1) estimate the total number of fish caught, C(t) as

w(L,t) (8)  
where w(L,t) is the average weight of a size L fish at time t; and (2) estimate C(L,t) as C(L,t)=C(t)p(L,t).

Third and most difficult, we need to estimate the boundary numbers N(a,1) and N(1,t) for SRA or N(a,T) and N(na,t) for VPA. This problem needs to be approached through the various time series fitting procedures of stock assessment, discussed in the following section.

## Fitting the SRA and VPA models to time series data

Typically stock assessment models are fitted to time series data on (1) relative or absolute abundances, hopefully from surveys rather than fishery cpues; (2) length-age composition data; and/or (3) point assessments of recent biomass or exploitation rate. Typically assessment fitting results related to harvest management reference points (unfished biomass and MSY biomass, Fmsy, etc.) are most sensitive to the relative abundance data, with composition data providing only relatively modest improvements in estimation of fishing mortality rates and recent abundances in the SRA case (and then only when vulnerability at age follows a simple, temporally stable logistic pattern; see Walters, Can. J. Fish. Aquat. Sci. 61: 1061–1065 (2004)).

For the VPA case, estimation of the {N(a,T),N(na,t)} set by fitting to relative abundance data, i.e. VPA “tuning”, is straightforward and can be done using the same likelihood or sums of squares criteria used in software like ADAPT (gavaris, ref), except for one wrinkle. Unlike standard VPA formulations, the N(na,t) oldest age abundances cannot be assumed to be zero or to be estimable from terminal age catches and exploitation rates, so there is not the standard VPA “convergence” as the model is propagated backward over time. This means that it is necessary to either provide relative abundance estimates for the whole historical period over which the model is fitted, or to assume stable or smoothly changing values of N(na,t) for earlier times t. Unfortunately, assuming stable values of N(na,t) for earlier years, based on estimates from fitting to data from later years, is likely to result in underestimates of early or unfished abundance.

For the SRA case, estimation of the {N(a,1),N(1,t)} set is considerably more complex, as indicated by the simulated estimation example in Fig. 1, where we pretended to have only (precise) biomass trend data to use for parameter estimation. For this example, we generated simulated data using eqs. (1)-(5) (M=0.3, von Bertalanffy K=0.3, L∞=10, 1 cm length intervals, Beverton-Holt stock-recruitment with Goodyear compensation ratio 10.0, unfished recruitment Ro=1.0) and a logistic size-vulnerability relationship, then tried to fit the resulting biomass time trend “data” while allowing only annual relative recruitment multipliers and the Beverton-Holt stock recruitment parameters to vary. That is, we forced structure onto the N(1,t) pattern by predicting it in the estimation model for each year using the correct stock-recruitment form, but not assuming to know the parameters and while trying to estimate the recruitment anomaly pattern over time. The estimation criterion was the sum of squared deviations between observed and predicted survey biomasses, plus the sum of squared recruitment anomalies (of log recruits/spawner) divided by an assumed variance for these anomalies. The results of this simple estimation approach are clearly not satisfactory; the recruitment anomaly penalty in the estimation criterion leads to underestimation of the recruitment anomalies and the recruitment compensation ratio, and overestimation of biomasses over time.

An important point is that the recruitment estimation errors seen in this simple fitting exercise lead to a striking diagnostic pattern in the estimated U(L,t) exploitation rate pattern (Fig. 1c). For each year when the simulated population had relatively high recruitment, there then followed a peak in U(L,t) that moved to increasing lengths over time. Likewise, the low simulated recruitment in year 11 led to a trough in U(L,t) values over subsequent years. Were we examining these patterns without knowing that they were due to estimation failure (failure of the estimation to “see” the high and low recruitments), we might conclude that the fishery was somehow able to target large cohorts over time, and avoid weak cohorts. Such cohort-specific targeting and avoidance might actually be possible in some fisheries, where cohorts can be spatially separated along seasonal and/or ontogenetic migration trajectories due to size-related differences in movement speeds, but such apparent targeting is much more likely to be indicative of errors in estimation of recruitments. An obvious initial test for possible recruitment estimation errors is to do the VPA as well as the SRA; then if the VPA fails to show cohort-specific ridges in the U(L,t) pattern, and/or considerably higher recruitment variation, then it is very likely that the SRA has given poor recruitment anomaly (and stock-recruitment parameter) estimates.

One might wonder at this point whether it is even worthwhile to do the SRA, i.e. why bother if it has to be cross-validated or corrected using VPA estimates? The answer is simple: the SRA can be used directly for simulation of future stock sizes and policy options, just by running the equations forward over time from time T with various assumptions about future exploitation rates and recruitment anomalies. Such forward simulations cannot be carried out with the VPA until a stock-recruitment relationship is fitted to the N(1,t) results. This fitting can be included in the VPA tuning, by including a sum of squared deviations from a stock-recruitment model in the fitting criterion (and stock-recruit parameters in the set of parameters to be estimated), but simulation tests of this combined estimation procedure indicate that it will result in the same bias pattern (underestimation of recruitment anomalies, patterned U(L,t) variation as the SRA fitting.

Fortunately, it looks like there is a very simple strategy for “rescuing” the SRA from providing poor estimates of recruitment variation, at least for cases where it can be reasonably assumed that fishers cannot target or avoid specific cohorts as these pass through the population. This strategy involves adding a penalty term to the fitting criterion for departures of the estimated length-vulnerability schedule from its long-term average. That is, first construct a table of length-specific apparent vulnerabilities V(L,t) for each year, as

(9).

Then calculate the arithmetic average of these V(L,t) over time, Calculate the sum of squared deviations

2 (10).

Add this to the fitting criterion, multiplied by an inverse variance weight 1/σV2 on order 0.1, and fit the model while estimating the {N(a,T),N(na,t), and stock recruit parameter} set. For the example shown in Figure 1, this strategy immediately results in almost exactly the correct parameter values. Perhaps the best way to think about SSV (or more precisely SSV/σV2) is as the exponent of a normal prior probability distribution for variation in V(L,t), with low values of σ2 representing a belief that V(L,t) has not varied much over time.