SEDAR15-AW01

SEDAR

Southeast Data, Assessment, and Review

SEDAR 15
Stock Assessment Model

DRAFT

October, 2007

Paul Conn, Kyle Shertzer, and Erik Williams
Center for Coastal Fisheries and Habitat Research
Beaufort, North Carolina

Statistical Catch-at-age Model

General Modeling Approach

SEDAR 15 assessments of greater amberjack and red snapper applied two different assessment models: a statistical catch-at-age model and a surplus production model. Application of the production model is described elsewhere; this document describes the general structure of the catch-at-age model. Application of this general structure to each species differed slightly to accommodate stock-specific characteristics.

The essence of statistical catch-at-age models is to simulate a population forward in time while including aspects of the fishing process (i.e., gear selectivity). Quantities to be estimated are systematically varied from starting values until characteristics of the simulated populations match available data on the real population as closely as possible. Such data include total catch by fishery and year; observed age and length composition by gear and year; and observed indices of abundance.

The method of forward projection has a long history in fishery models. It was introduced by Pella and Tomlinson (1969) for fitting production models and then used by Fournier and Archibald (1982), Deriso et al. (1985) in their CAGEAN model, and Methot (1989) in his stock-synthesis model. The model developed for this assessment is an elaboration of the CAGEAN and stock-synthesis models and is similar in structure to SEDAR assessments of red porgy, black sea bass, tilefish, snowy grouper, and gag grouper. Statistical catch-at-age models share many attributes with ADAPT-style tuned and untuned VPAs.

Methods

A general description of the assessment model follows.

Properties of age-structured model
The statistical catch-at-age model for this assessment was implemented in the AD Model Builder (ADMB) software (Otter Research 2001). The model is detailed in Table 1. It's major characteristics can be summarized as follows:

Natural mortality rate - The natural mortality rate was assumed constant over time, but age-specific. A vector of age-specific *M*, based on Lorenzen (1996), was used as a starting estimate. The age-specific *M* vector was then re-scaled based on a fraction of survivors at the oldest age consistent with the findings of Hoenig (1983).

Stock dynamics – In the assessment model, new biomass was acquired through growth and recruitment, and poplulion size experienced exponential decay from fishing and natural mortality. The population was assumed closed (no net migration to or from the study area). The oldes age class allowed for the accumulation of fish (i.e., plus group). Because landings were estimated back to the onset of the fishery, initial stock was considered to be at the unfished (virgin) level.

Growth/Maturity – Size at age was modeled with a von Bertalanffy equation. Maturity of females was modeled with a logistic equation.

Spawning biomass – Spawning biomass was considered to be the mature female biomass, assuming a 50:50 sex ratio. It was computed from number at age at the time of year when spawning peaks.

Recruitment – Recruitment was predicted from mature female biomass, assuming a 50:50 sex ratio, using a Beverton–Holt spawner-recruit model. In years when composition data could provide information on year-class strength, estimated recruitments were loosely conditioned on the Beverton-Holt model with autocorrelated residuals. In years prior to when composition data were available, recruitment followed the Beverton-Holt model precisely (similar to an age-structured production model).

Fishing - Four fisheries were modeled individually: commercial handline, commercial diving, headboat, and general recreational (MRFSS). Separate fishing mortality rates were estimated for each of these fisheries. Selectivity at age was allowed to vary by regulation period.

Selectivity functions - Selectivity was fit internally using a logistic or double-logistic (dome-shaped) model for each fishery, as appropriate. This parameteric approach reduces the number of estimated parameters and imposes theoretical structure on the estimates. Selectivity of each fishery was estimated separately for different periods of size-limit regulations.

Landings – Landings were estimated via the standard Baranov catch equation.

Discards – Discards were available for the headboat, MRFSS, and commercial handline fisheries. Diving discards were believed to be negligible and were not included. Discard mortalities were estimated via the Baranov catch equation.

Indices of abundance -The model was fit to three fishery dependent indices of abundance: headboat, MRFSS, and commercial handline. Predicted indices were computed from number at age at the midpoint of the year.

Biological benchmarks - Biological benchmarks were calculated based on maximum sustainable yield (MSY) estimates from the Beverton–Holt recruitment model with bias correction. Computed benchmarks included MSY, fishing mortality rate at MSY (Fmsy), and total mature biomass at MSY (SSBmsy). These benchmarks are conditional on the estimated selectivity functions. The selectivity pattern used here was the effort-weighted selectivities at age, with effort from each fishery averaged over the last three years of the assessment.

Estimated Parameters
The model estimates annual fishing mortality rates of each fishery, selectivity parameters of each fishery in each period of fishing regulations, Beverton–Holt parameters, annual recruitment deviations, catchability coefficients associated with abundance indices, and CV of size at age.

Fitting criterion -The fitting criterion was a total likelihood approach in which fishery catch, observed age and length compositions, and the abundance index patterns were fit to the degree that they are compatible. Landings data and abundance index data were fit using a lognormal likelihood. Age and length composition data were fit using a multinomial likelihood. Relative statistical weightings of likelihood components for an initial model run were chosen after examining candidate model runs. The criteria for choice were a balance of reasonable fit to all available data and a good degree of biological realism in estimated population trajectory.

Likelihood Component Weights
The selection of likelihood component weights for the initial run model involved an iterative process of model fitting, examination of the fit, and adjustment of the weights. The performance of an individual model run was evaluated based on its fit to the observed datasets. These datasets include four time series of landings, three time series of discards, three abundance indices, and age and length compositions from both fishery and survey sources. The influence of each dataset on the overall model fit was determined by the specification of the error terms in each likelihood component. In the case of lognormal likelihoods, error was quantified by the annual coefficient of variation, and for the multinomial components, by the annual sample sizes. These terms determine the influence of each year of data relative to other years of the same data source. However, the relative influence of different components can only be treated by re-weighting each likelihood. An objective determination of these weights is an unsolved problem in statistical modeling. In this case, the weights were determined by examination of overdispersion, model mis-specification (e.g. runs of residuals), and the general reliability (i.e. our understanding of information content) of the data source.

We reduced the number of weights to be examined by grouping likelihood components based on their type, scale, and method of collection. For example, the four fisheries landings data were grouped, so that a single weight was applied to all four components. Similarly the discard components were grouped, the index components were grouped, the age composition components were grouped, and the length composition components were grouped. The model also contained a likelihood component for the annual recruitment deviation parameters, which were constrained to follow a Beverton–Holt stock-recruit curve.

Model Testing

To ensure that the assessment model produces viable estimates (i.e., that all model parameters are identifiable), we analyzed test data that were generated with known parameter values. For simplicity, we considered a stripped down version of the model in Table 1 that nevertheless retained all essential components. In particular, we used a simulation model to generate data from one fishery and included likelihood contributions of landings, CPUE, and age composition. We assumed that the selectivity curve remained the same over time, and set all likelihood weights (λ_i) equal to one. The simulation model was programmed independently (written in R) of the assessment model (written in ADMB)

Parameter identification was determined using the "analytical-numeric" approach of Burnham et al. (1987). Expected value data were generated deterministically from input parameter values (Table 2), without any process or sampling error. These data were then analyzed via the assessment model in attempt to obtain the exact parameters that generated the data.

In this test, all model parameters were estimated exactly. This result provides evidence that all parameters could be properly identified. It further suggests that the assessment model is implemented correctly.

Literature Cited

Burnham, K. P., Anderson, D. R., White, G. C., Brownie, C., and K. H. Pollock. 1987. Design and analysis methods for fish survival experiments based on release-recapture. American Fisheries Society Monograph 5, 437 pp.

Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences 42:815–824.

Fournier, D. and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39:1195–1207.

Hoenig, J. M. 1983. Empirical use of longevity data to estimate mortality rates. Fishery Bulletin 81:898–903.

Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49:627–642.

Methot, R. M. 1989. Synthetic estimates of historical abundance and mortality for northern anchovy. American Fisheries Society Symposium 6:66–82.

Otter Research, Ltd. 2001. AD Model Builder version 7.7.1. Otter Research, Sidney, B.C, Canada.

Pella, J. J. and P. K. Tomlinson. 1969. A generalized stock production model. Bulletin of the Inter-American Tropical Tuna Commission 13:419–496.

Table 1: General definitions, input data, population model, and negative log-likelihood components of the statistical catch-at-age model applied to red snapper. The model applied to greater amberjack was identical in structure, but differed in some details particular to the stock.

General Definitions	**Symbol**	**Description/Definition**
Year index	y	$y = \{1901,..,2006\}$
Age index	a	$a = \{1,...,A\}$, where $A = 25+$
Length bin (mm)	l'	$l' = \{190,220,\ldots,1000\}$, bin size = 30 mm
Fishery index	f	$f = \{$1 handline, 2 diving, 3 headboat, 4 MRFSS$\}$
CPUE index	u	$u = \{$1 headboat CPUE, 2 MRFSS CPUE, 3 handline CPUE$\}$
Input Data	**Symbol**	**Description/Definition**
Mean length-at-age	l_a	$l_a = L_\infty\left(1-\exp\left[-K\left(a-t_0\right)\right]\right)$ where parameters L_∞, K, and t_0 are fixed
Age-length conversion matrix	$\psi_{a,l'}$	$\psi_{a,l'} = \dfrac{\exp\left[-\left(\dfrac{l'-l_a}{2c^l l_a}\right)\right]}{\sqrt{2\pi}\left(c^l l_a\right)^2}$ where c^l is a fixed value for the coefficient of variation in length at age and the matrix is re-scaled to sum to 1 across ages
Population weight-at-age	w_a	Computed from size at age at the midpoint of the year $w_a = \gamma l_a^\beta$, where γ and β are fixed
Maturity-at-age	m_a	Logistic function of age
Observed CPUE indices	$U_{u,y}$	u=1, headboat ($y = 1976,\ldots,2006$), based on number of fish captured per hook-hour. u=2, MRFSS CPUE ($y = 1983,\ldots,2006$), based on number of fish captured per angler-trip. u=3, handline CPUE($y = 1993,\ldots,2006$), based on pounds (whole weight) caught per hook-hour.
Coefficient of variation for U ‘s	$c_{u,y}$	$u = \{1, 2, 3\}$ (see above), annual values from GLM model or sampling error.
Observed age compositions	$p_{f,a,y}$	Computed as percent age composition at age (a) for each year (y) and fishery (f)
Age composition sample sizes	$n_{f,y}$	Number of age samples collected in each year (y) from each fishery (f)

Observed length compositions	$p'_{f,l,y}$	Computed as percent length composition at length (*l*) for each year (*y*) and fishery (*f*)
Length composition sample sizes	$n'_{f,y}$	Number of length samples collected in each year (*y*) from each fishery (*f*)
Observed fishery landings	$L_{f,y}$	Reported landings in weight for each year (*y*) from each fishery (*f*)
Coefficient of variation for L_f	$c_{L_{f},y}$	Annual values fixed based on understanding of historical accuracy of estimates
Age-specific natural mortality	M_a	Fixed across years from Lorenzen (1996), re-scaled based on Hoenig (1983)

Population Model	**Symbol**	**Description/Definition**
Fishery selectivity	$s_{f,a}$	Estimated separately for each period of size regulations: $$s_{f,a} = \begin{cases} \left[\frac{1}{1+\exp\left(-\eta_{1,f}\left[a-\alpha_{1,f}\right]\right)}\right] & \text{for } f=1 \\ \left[\frac{1}{1+\exp\left(-\eta_{1,f}\left[a-\alpha_{1,f}\right]\right)}\right]\left[1-\frac{1}{1+\exp\left(-\eta_{2,f}\left[a-\left(\alpha_{1,f}+\alpha_{2,f}\right)\right]\right)}\right]\left[\frac{1}{\max\left(s_{f,a}\right)}\right] & \text{for f=\{2,3,4\}} \end{cases}$$ where $\eta_{1,f}$, $\eta_{2,f}$, $\alpha_{1,f}$ and $\alpha_{2,f}$ are estimated parameters.
Index selectivity	$s'_{u,a}$	Assumed equal to the corresponding fishery selectivity in year *y*.
Fishing mortality	$F_{f,a,y}$	$F_{f,a,y} = s_{f,a} F_{f,y}$ where $F_{f,y}$'s are fully selected estimated parameters
Total mortality	$Z_{a,y}$	$$Z_{a,y} = M_a + \sum_{f=1}^{4} F_{f,a,y}$$
Mature biomass per recruit at $F = 0$	ϕ_y	$$\phi_y = \sum_{a=1}^{A} 0.5 N_{a,y} m_a w_a \Big/ N_{1,y}$$ where $N_{a+1,y} = N_{a,y} \exp\left(-Z_{a,y}\right)$ and $N_{A,y} = N_{A-1,y} \exp\left(-Z_{A-1,y}\right) / \left[1-\exp\left(-Z_{A,y}\right)\right]$

Population numbers Population mature biomass	$N_{a,y}$ S_y	$N_{1,1901} = R_0$ $N_{a+1,1901} = N_{a,1901}\exp(-M_a)$ $N_{A,1901} = N_{A-1,1901}\exp(-M_{A-1})/[1-\exp(-M_A)]$ $N_{1,y} = \dfrac{0.8R_0 h\varepsilon_y}{0.2\phi_y R_0(1-h)+(h-0.2)\varepsilon_y} + R_y$ $N_{a+1,y+1} = N_{a,y}\exp(-Z_{a,y})$ $N_{A,y} = N_{A-1,y-1}\exp(-Z_{A-1,y-1}) + N_{A,y-1}\exp(-Z_{A,y-1})$ $S_y = \sum_{a=1}^{A} 0.5 N_{a,y} m_a w_a$ where R_0 (virgin recruitment) and h (steepness) are parameters of the spawner-recruit curve and R_y are annual recruitment deviation parameters.
Population biomass	B_y	$B_y = \sum_{a=1}^{A} N_{a,y} w_a$
Predicted catch-at-age	$\hat{C}_{f,a,y}$	$\hat{C}_{f,a,y} = \dfrac{F_{f,a,y}}{Z_{a,y}} N_{a,y}[1-\exp(-Z_{a,y})]$
Predicted landings	$\hat{L}_{f,y}$	$\hat{L}_{f,y} = \sum_{a=1}^{A} \hat{C}_{f,a,y} w_a$
Predicted age composition	$\hat{p}_{\{f,u\},a,y}$	$\hat{p}_{\{f,u\},a,y} = \hat{C}_{\{f,u\},a,y} \Big/ \sum_{a=1}^{A} \hat{C}_{\{f,u\},a,y}$

Predicted CPUE indices	$\hat{U}_{u,y}$	$\hat{U}_{u,y} = \begin{cases} \sum_{a=1}^{A} N_{a,y} s'_{1,a} q_1 & \text{for } u = 1 \\ \sum_{a=1}^{A} N_{a,y} s'_{2,a} q_2 & \text{for } u = 2 \\ \sum_{a=1}^{A} N_{a,y} s_{3,a} q_3 & \text{for } u = 3 \end{cases}$ where q_1, q_2, and q_3 are catchability parameters
Negative Log-Likelihood	**Symbol**	**Description/Definition**
Multinomial age composition	Λ_1	$\Lambda_1 = -\lambda_1 n_{\{f,u\},y} \sum_{a=1}^{A} (p_{\{f,u\},a,y}+x)\log(\hat{p}_{\{f,u\},a,y}+x) - (p_{\{f,u\},a,y}+x)\log(p_{\{f,u\},a,y}+x)$ where λ_1 is a preset weighting factor and x is fixed at an arbitrary value of 0.00001

Multinomial length composition	Λ_2	$\Lambda_2 = -\lambda_2 n'_{\{f,u\},y} \sum_{l'} \left(p'_{\{f,u\},l',y} + x\right) \log\left(\hat{p}'_{\{f,u\},l',y} + x\right) - \left(p'_{\{f,u\},l',y} + x\right) \log\left(p'_{\{f,u\},l',y} + x\right)$ where λ_2 is a preset weighting factor and x is fixed at an arbitrary value of 0.00001
Lognormal indices	Λ_3	$\Lambda_3 = \lambda_3 \sum_y \frac{\left[\log\left(U_{u,y} + x\right) - \log\left(\hat{U}_{u,y} + x\right)\right]^2}{2c_{u,y}^2}$ where λ_3 is a preset weighting factor and x is fixed at an arbitrary value of 0.00001
Lognormal landings	Λ_4	$\Lambda_4 = \lambda_4 \sum_y \frac{\left[\log\left(L_{f,y} + x\right) - \log\left(\hat{L}_{f,y} + x\right)\right]^2}{2c_{L_f,y}^2}$ where λ_4 is a preset weighting factor and x is fixed at an arbitrary value of 0.00001
Recruitment constraint	Λ_5	$\Lambda_5 = \lambda_5 \sum_y R_y^2$

Table 2. Parameter values used to simulate expected value data.

Parameter	Definition	Value
M	Instantaneous natural mortality	0.25
A	Number of age classes	15
Y	Years of data	31
$m_{.5}$	Age at 50% maturity	5
β_m	Slope of maturity-at-age logistic function	1.0
l_∞	Asymptotic length (von Bertalanffy)	1000
K	von Bertalanffy growth rate	0.35
t_0	Age at size 0 (von Bertalanffy)	-0.5
a_{wt}	Length-weight conversion parameter ($W = a_{wt} L^{b_{wt}}$)	1.0 E-6
b_{wt}	Length-weight conversion parameter ($W = a_{wt} L^{b_{wt}}$)	3.0
$a_{.5}$	Age at 50% selectivity	4.0
β_s	Slope of selectivity logistic function	1.0
R_0	Recruitment at F=0	1.0 E+5
h	Steepness of Beverton-Holt recruitment model	0.7
q	Catchability for CPUE index	1.0 E-5
n	Annual sample size for age compositions	200
F_1/F_{MSY}	Ratio of F in year 1 to F_{MSY}	0.01
F_A/F_{MSY}	Ratio of F in year A to F_{MSY}	2.0
F_0/F_{MSY}	Ratio of F to F_{MSY} in years prior to landings	0.01