North American Journal of Fisheries Management

ISSN: 0275-5947 (Print) 1548-8675 (Online) Journal homepage: http://www.tandfonline.com/loi/ujfm20

Bias and Precision of Estimates from an Age-Structured Stock Assessment Program in Relation to Stock and Data Characteristics

Yanshui Yin & David B. Sampson

To cite this article: Yanshui Yin & David B. Sampson (2004) Bias and Precision of Estimates from an Age-Structured Stock Assessment Program in Relation to Stock and Data Characteristics, North American Journal of Fisheries Management, 24:3, 865-879, DOI: 10.1577/M03-107.1

To link to this article: http://dx.doi.org/10.1577/M03-107.1

North American Journal of Fisheries Management 24:865–879, 2004

Bias and Precision of Estimates from an Age-Structured Stock Assessment Program in Relation to Stock and Data Characteristics

YANSHUI YIN AND DAVID B. SAMPSON*

Coastal Oregon Marine Experiment Station and Department of Fisheries and Wildlife, Hatfield Marine Science Center, Oregon State University, Newport, Oregon 97365, USA

Abstract.—Assessments for many U.S. Pacific coast groundfish stocks have been developed using the statistical catch-at-age method known as Stock Synthesis. This study used Monte Carlo simulation and a fractional factorial experiment to evaluate the effects of input data errors and stock characteristics on bias and precision in estimates of ending exploitable biomass, rate of fishing mortality, depletion, and other output variables. Nine factors were examined: length of the data series, rate of natural mortality, shape of the fishery selectivity curve, trend in fishing mortality, recruitment pattern, and level of sampling error in the data for catch, fishing effort, a survey biomass index, and sample size for fishery and survey age compositions. Length of the data series, age composition sample size, survey biomass variability, and fishing effort variability were the most influential factors for most of the output variables. The estimates of depletion had the least bias and the estimates of starting biomass the smallest variability; the estimates of ending recruitment had the greatest bias and largest variability. For all the output variables examined the estimates appeared to be median-unbiased. For the conditions considered in the experiment, it appears that the accuracy of assessment estimates for ending exploitable biomass and projected catch would be more readily improved by increased age composition sampling than by comparable (but much more expensive) improvements in survey estimates of stock biomass.

Fisheries managers and fishers generally appreciate that fish resources are not unlimited and that exploitation should be regulated. To efficiently manage an exploited fish stock, managers need to know the stock's status, particularly whether it is increasing or decreasing and why. This is one reason that stock assessments play a key role in fisheries management (Megrey 1989). Estimates of fish biomass, from which catch quotas are derived, may be inaccurate if they are based on models that are not robust to errors in the input data (National Research Council 1998). In the face of declining resources, reliable stock assessment information has become more crucial (Richards and Megrey 1994). Better understanding of the behavior of stock assessment models could lead to better estimates of stock status and thus better management of fisheries.

The Stock Synthesis program (Methot 1990, 2000) has been widely used on the West Coast of the United States for assessing groundfish stocks (Pacific Fishery Management Council 2000). The underlying population dynamics models and approach to fitting multiple data series are similar to those used in other statistical catch-at-age programs, including catch-at-age analysis (Deriso et al. 1985), integrated catch-at-age analysis (Patterson 1998), and Coleraine (Hilborn et al. 2003). Stock Synthesis can incorporate diverse types of information to reconstruct both the dynamics of an exploited fish population and the processes by which we observe the population and its fishery. A major strength of Stock Synthesis, which uses the maximum likelihood method for parameter estimation, is its ability to accommodate multiple input data sources having different degrees of uncertainty. For example, this program can simultaneously analyze data on catch biomass, age composition, stock abundance, and fishing effort, which are likely to be subject to different levels and types of error.

Given multiple data sources, it is unlikely that each will equally influence the estimates from Stock Synthesis. Also, the Stock Synthesis results are probably sensitive to biological traits of the fish (e.g., longevity and growth), characteristics of the fishery (e.g., exploitation rate and selection pattern), and the levels of observation error in the data series for total catch, fishery age composition, fishing effort, and survey indices of abundance. Determining the impact of these factors on the accuracy of Stock Synthesis estimates should help us understand how this type of assessment model

* Corresponding author: david.sampson@oregonstate.edu

Received June 11, 2003; accepted October 24, 2003

TABLE 1.—Low and high levels of the nine experimental factors in the factorial design to study the performance of a Stock Synthesis program. Coefficients of variation are defined as SD/mean.

Factor	Symbol	Description	Low level	High level
A	NYrs	Number of years of data	8	16
B	SSize	Sample size for age composition	100 fish	400 fish
C	EffCV	Fishing effort coefficient of variation	20%	80%
D	SurCV	Survey biomass coefficient of variation	20%	80%
E	NatM	Natural mortality coefficient	0.2/yr	0.4/yr
F	FInc	Fishing mortality increment	0.01/yr	0.03/yr
G	CatCV	Catch data coefficient of variation	10%	20%
H	FSel	Fishery selectivity	Domed	Asymptotic
J	RecV	Recruitment variability	Constant	Variable

performs under various circumstances and identify which input data are most in need of improvement so that sampling effort can be allocated more cost-effectively.

Methods

In this study we used Monte Carlo simulation (Rubinstein 1981) to determine the impacts of input data measurement errors and stock characteristics on the bias and precision of Stock Synthesis estimates. Our approach, which was similar to that used in Bence et al. (1993) and Punt et al. (2002), was to define a stock and its fishery with known characteristics, generate random data sets based on the defined fishery system, analyze the data sets with the Stock Synthesis program, and then compare the estimates from Stock Synthesis with the true values.

As the main tools for this exercise, we created a simulation package consisting of three C++ programs that (1) defined the stock and fishery characteristics, (2) generated random data sets for input to the Stock Synthesis program, and (3) tabulated summary statistics from the Stock Synthesis output files. A typical fishery system was composed of a fish stock, a fishery operating on the stock, a survey for monitoring the stock, and a series of sampling activities. Defining a fish stock involved specifying parameters for the stock's biological traits (average weight at age, maturity at age, natural mortality, and recruitment) and parameters for the processes of fishing and observing the stock (fishing mortality, catchability, fishery and survey selectivity, and frequency and sample size for age composition samples). The data simulator program calculated true demographic data based on the specified stock and fishery characteristics using the same deterministic equations that underlie Methot's Stock Synthesis program. In addition, it generated random data sets that were then analyzed directly by the Stock Synthesis program. The theoretical expected values for the random data sets were the same as those given by the deterministic equations. The data tabulation program scanned the output files from Stock Synthesis, calculated summary statistics for the Stock Synthesis estimates, compared these statistics with their corresponding true values, and generated measures of the relative accuracy and precision of the Stock Synthesis estimates. Details of the demographic equations, the random data generator, and the log-likelihood equations are provided in the Appendix.

Each simulated fishery in this study produced annual values for total catch, age composition, and fishing effort, and the simulated survey provided annual estimates of stock biomass and age composition. The simulated observations of total catch, fishing effort, and survey biomass followed lognormal distributions that were generated in such a manner that they would be unbiased. The simulated observations of fishery and survey age composition followed simple multinomial distributions and were generated without age-reading error. The Stock Synthesis program was configured to treat the observations in accordance with the way they were generated, that is, with the correct form and scale of random error.

Experimental design.—Our study used a fractional factorial design to simultaneously examine the effects of nine factors (Table 1) on the performance of the Stock Synthesis program: (1) the number of years in the data series (NYrs); (2) the size of the annual age composition samples for both the fishery and the survey (SSize); (3) the coefficient of variation (CV = SD/mean) for the annual fishing effort data (EffCV); (4) the CV for the annual survey biomass data (SurCV); (5) the instantaneous rate of natural mortality (NatM); (6) the annual increment in the rate of fishing mortality (FInc); (7) the CV for the annual catch data (CatCV); (8) fishery selectivity (FSel); and (9) annual recruitment variation (RecV). The level of natural mortality determined the values for several stock parameters (Appendix). In choosing levels for the factors, we did not attempt to mimic the characteristics of any particular stock and fishery but instead selected values that spanned a realistic range of possible stock and data characteristics.

In all the simulated stocks, fishing began at the start of the first year of the simulated period with a step increase (FInc) in the instantaneous rate of fishing mortality at the start of each subsequent year. Thus, the fishing mortality coefficient during the last year of the simulated period was only 0.08/year for treatments with the short data series and low FInc but 0.48/year for the long data series and high FInc.

The experiment evaluated two types of fishery selectivity relationship: a ''domed'' curve in which intermediate age-classes experienced the full rate of fishing mortality and an ''asymptotic'' curve in which the oldest age-classes experienced the full rate of fishing mortality. The selectivity curve for the annual survey always had the asymptotic form.

The experiment considered two types of actual recruitment, constant and variable. The Stock Synthesis configurations differed between the two types of recruitment in terms of estimation of the initial age composition vector, but for both types Stock Synthesis estimated annual recruitment values for all years in the modeled period. For simulations with constant actual recruitment, the annual recruitment was 3.0 million fish, the initial age composition at the start of the first year was at equilibrium, and Stock Synthesis was configured to estimate the initial equilibrium age composition from the level of recruitment and the rate of natural mortality. For simulations with variable actual recruitment, the average annual recruitment was also 3.0 million fish, but the values varied according to the sequence 3.5, 4.0, 1.2, 4.2, 3.0, 3.2, 1.7, and 3.2 million (which were repeated as necessary). In this case, Stock Synthesis was configured to estimate the full set of values for the initial nonequilibrium age composition.

For each of the experimental treatments we applied the data simulator four times, each time generating a ''batch'' of 200 replicate data sets that were analyzed with Stock Synthesis. We used the batches to produce four independent replicate estimates of the median values and coefficients of variation from each treatment's data. Also, because averages of 200 values should be fairly normally distributed even though the individual values are not, the batch averages should conform more closely to the assumptions of the analysis of variance (ANOVA) method that we applied to the experimental data.

Random data sets were generated in accordance with a one-fourth fraction of the 2^9 factorial design, meaning that instead of examining the complete set of $2^9 = 512$ factor combinations, we only considered $2^{9-2} = 128$ combinations. Because this fractional design is incomplete, most of the higher-order interactions in the full model could not be estimated. However, one would expect high-order interactions to be small relative to low-order interactions (Box et al. 1978). The design generators were H = ACDFG and J = BCEFG, and the defining relation was I = ACDFGH = BCEFGJ = ABDEHJ. In these relationships each letter refers to one of the experimental factors (Table 1). Box et al. (1978) provide details and examples of how the design generators determine the coefficients in the design matrix for the last two experimental factors and how the defining relation determines the pattern of confounding among the factors. This design does not confound the main effects with any four-factor or lower-order interactions or two-factor interactions with each other or with any three-factor interactions. But the design does confound two-factor interactions with four-factor interactions and three-factor interactions with other three-factor interactions, and so on. For example, in our design the two-way interaction AB was aliased with the four-way interaction DEHJ, meaning that the value estimated for AB was actually the combined effect AB + DEHJ.

Dependent variables (Stock Synthesis outputs).—The Stock Synthesis program produces a wide variety of estimated outputs, such as the annual series of biomass, fishing mortality, catch, and recruitment. In this study we focused on seven categories (Table 2): the estimates for the last year for total biomass (EndBio), rate of fishing mortality (EndF), recruitment (EndRec), and exploitable biomass (EndXBio); the $F_{35\%}$ catch projected for the year following the last (F35Catch); the estimates for the first year for total biomass (StartBio); and the ratio of the total biomass for the last year to the total biomass for the first year (Depletion). The exploitable biomass, which represents the stock biomass susceptible to fishing, is calculated as the sum over all age-classes of the biomass at age times the age-specific fishery selection coefficient. The term $F_{35\%}$ is defined as the value of fishing mortality that would reduce the spawning stock biomass per recruit to 35% of the level that would exist with no fishing (Clark 1991). The $F_{35\%}$ catch is the predicted catch biomass that accu-

TABLE 2.—Output variables that were the focus of the analyses of bias and precision.

Variable	Description
EndBio	Stock biomass at the start of the last year in the time series.
EndF	Instantaneous rate of fishing mortality during the last year.
EndRec	Number of new recruits joining the stock at the start of the last year.
EndXBio	Exploitable stock biomass at the start of the last year.
F35Catch	Projected catch for the year after the last year given an $F_{35\%}$ rate of fishing mortality.[a]
StartBio	Stock biomass at the beginning of the first year in the time series.
Depletion	Ratio of stock biomass at the start of the last year to that at the start of the first year.

[a] $F_{35\%}$ is the fishing mortality that would reduce the spawning stock biomass per recruit to 35% of the level that would occur with no fishing.

mulates during a year given the $F_{35\%}$ fishing rate, the age-specific selection coefficients, and the exploitable biomass.

For each experimental treatment and output type, we calculated the relative bias and relative variability for each of the four batches, each batch containing 200 data sets. We measured the relative variability within each batch using the coefficient of variation and measured relative bias in two forms: the relative bias of the mean and the relative bias of the median, defined as

$$\frac{\text{mean(estimated } X) - \text{true } X}{\text{true } X} \quad \text{and}$$

$$\frac{\text{median(estimated } X) - \text{true } X}{\text{true } X}.$$

We calculated the median of each batch of 200 estimates as the average of the 100th and 101st ordered values.

For the sets of mean values, CVs, and median values from each of the seven focal variables, we conducted separate fractional ANOVAs using the Minitab statistics program (Minitab 1998). Because diagnostic plots from initial fits of the ANOVA models indicated a tendency for the residual variation to increase with the predicted values, we $\log_e$ transformed all the dependent variables for the final analyses.

Sensitivity to initial parameter values.—In the main experiment, we used the true values as the initial parameter values for starting the iterative search for the maximum likelihood estimates. However, the choice of initial parameter values may influence whether or not the search algorithm finds a local rather than the global maximum. To examine the influence of initial parameter values on the performance of Stock Synthesis, we conducted a randomization experiment on the two treatments that produced the results for the estimate of ending biomass that were the most extreme in terms of the absolute value of the relative bias. For this extra experiment we generated 100 random data sets for each of the two extreme treatments. For each random data set generated, we ran the Stock Synthesis program 100 times, each time using a different set of randomized initial parameter values, each parameter varying uniformly within 40% of its true value.

Results

The seven types of Stock Synthesis estimates varied greatly in relative bias and relative variability (Table 3). Across the 128 treatments the measurements of the average relative bias of the mean were positive for all seven estimates, ranging from 1.6% for Depletion to 10.1% for EndRec. The distributions of the relative bias of the mean were skewed to the right for all seven estimates. The estimates of EndF had the largest positively biased values (82.0%) and the widest range, and the estimates of F35Catch had the largest negatively biased values (−10.8%). The relative variability (CV) values ranged from 0.030 for the estimates of StartBio to 1.18 for the estimates of EndRec. The distributions of all the CV values were skewed to the right. For the measurements of the relative bias of the median, the largest negatively biased value (−33.2%) and the largest positively biased value (31.6%) both occurred within the estimates of F35Catch. The distributions of the relative bias of the median were fairly tightly and symmetrically centered about zero for all seven types of estimates, indicating that the Stock Synthesis estimates tended to be median-unbiased.

Across the 128 treatments, the batch-level values for the relative bias of the mean and CV were all significantly correlated with one another (Table 4), with almost perfect correlation for some combinations of the seven focal variables. For example, the correlation coefficients among the values for EndBio, EndRec, and EndXBio were all greater than 0.95 for the estimates of the bias of the mean and CV. In contrast, the bias and CV values for

TABLE 3.—Summary statistics for the seven output variables (Table 2) tabulated across the 128 treatments with four replicate batches per treatment ($N = 512$). Relative bias is defined as the ratio of (1) the mean estimated value less the true value and (2) the true value; the coefficient of variation is defined as the ratio of the SD and the mean.

Statistic	EndBio	EndF	EndRec	EndXBio	F35Catch	StartBio	Depletion
Relative bias of the mean							
Mean	0.0816	0.0586	0.1014	0.0787	0.0791	0.0413	0.0159
Maximum	0.5579	0.8198	0.7709	0.5151	0.6723	0.3493	0.1624
95th percentile	0.2462	0.2477	0.3256	0.2464	0.2642	0.1385	0.0731
75th percentile	0.1105	0.0711	0.1327	0.1081	0.1116	0.0584	0.0249
Median	0.0561	0.0274	0.0663	0.0542	0.0503	0.0232	0.0122
25th percentile	0.0192	0.0044	0.0260	0.0187	0.0143	0.0074	−0.0001
5th percentile	−0.0018	−0.0208	−0.0038	−0.0061	−0.0132	−0.0018	−0.0231
Minimum	−0.0465	−0.0671	−0.0698	−0.0529	−0.1078	−0.0155	−0.0844
Range	0.6044	0.8869	0.8408	0.5680	0.7801	0.3649	0.2468
Coefficient of variation							
Mean	0.3468	0.3567	0.4486	0.3509	0.4042	0.1873	0.1879
Maximum	0.9469	1.1547	1.1843	0.9431	1.1210	0.6774	0.4467
95th percentile	0.7638	0.6652	0.8882	0.7733	0.8417	0.4589	0.3617
75th percentile	0.4562	0.4291	0.5619	0.4613	0.5293	0.2607	0.2103
Median	0.3039	0.3218	0.4057	0.3107	0.3521	0.1389	0.1637
25th percentile	0.1905	0.2417	0.2900	0.1979	0.2304	0.0754	0.1369
5th percentile	0.1362	0.1751	0.1852	0.1343	0.1573	0.0447	0.1092
Minimum	0.0933	0.1345	0.1435	0.0945	0.1226	0.0296	0.0920
Range	0.8536	1.0202	1.0408	0.8485	0.9985	0.6478	0.3547
Relative bias of the median							
Mean	−0.0016	−0.0050	−0.0123	−0.0044	−0.0094	0.0004	−0.0024
Maximum	0.2757	0.3000	0.2653	0.2648	0.3158	0.1834	0.0731
95th percentile	0.0565	0.0680	0.0533	0.0488	0.0617	0.0308	0.0389
75th percentile	0.0178	0.0179	0.0120	0.0147	0.0156	0.0101	0.0116
Median	−0.0007	−0.0082	−0.0091	−0.0019	−0.0049	0.0007	−0.0015
25th percentile	−0.0206	−0.0299	−0.0366	−0.0215	−0.0273	−0.0076	−0.0142
5th percentile	−0.0681	−0.0730	−0.0885	−0.0729	−0.0953	−0.0371	−0.0393
Minimum	−0.1997	−0.2071	−0.2728	−0.2195	−0.3316	−0.1667	−0.1728
Range	0.4754	0.5071	0.5381	0.4843	0.6474	0.3501	0.2459

EndF and Depletion were much less strongly correlated with any of the other variables or with each other. Also, these two variables produced the only negative correlation among the seven variables, for the estimates of bias of the mean.

The ANOVA results given here are limited to the Stock Synthesis estimates of the bias of the mean and CV for the variables EndXBio, EndF, and Depletion. The results for EndBio, EndRec, F35Catch, and StartBio were all very similar to the results for EndXBio. In the ANOVAs the main effects, two-way interactions, three-way interactions, and four-way interactions were all highly significant ($P < 0.01$; Table 5). However, the main effects and interactions differed in relative importance, the main effects accounting for most of the variability in the dependent variables. For example, in the ANOVA with the transformed values

TABLE 4.—Correlations among the estimates of relative bias (above the diagonal) and the estimated coefficients of variation (below the diagonal) for the seven output variables (diagonal). Each correlation coefficient was based on 512 paired observations. All were significant at $P < 0.001$ except for that between the estimates of the bias of the mean for variables EndF and F35Catch, which was significant at $P < 0.02$. See the caption to Table 3 for definitions of the relative bias and coefficient of variation.

Bias of the mean or coefficient of variation						
EndBio	0.245	0.981	0.996	0.915	0.913	0.514
0.900	EndF	0.267	0.237	0.105	0.231	−0.308
0.965	0.871	EndRec	0.969	0.890	0.855	0.545
0.998	0.896	0.962	EndXBio	0.915	0.924	0.499
0.976	0.898	0.950	0.975	F35Catch	0.831	0.515
0.901	0.778	0.843	0.915	0.852	StartBio	0.201
0.633	0.719	0.649	0.611	0.678	0.283	Depletion

TABLE 5.—Analysis of variance tables for the transformed estimates of the relative bias of the mean and the coefficient of variation (CV) for the output variables EndXBio, EndF, and Depletion. Abbreviations are as follows: SS = sum of squares; MS = mean square. See the caption to Table 3 for definitions of the relative bias and CV.

Source	df	SS	MS	*F*	*P*
$\log_e$(bias of EndXBio +0.1)					
Main effects	9	50.23	5.58	180.9	<0.001
Two-way interactions	36	13.47	0.37	12.1	<0.001
Three-way interactions	55	9.36	0.17	5.5	<0.001
Four-way interactions	27	1.92	0.07	2.3	<0.001
Residual (pure) error	384	11.85	0.03		
$\log_e$(bias of EndF +0.1)					
Main effects	9	45.62	5.07	93.2	<0.001
Two-way interactions	36	28.71	0.80	14.7	<0.001
Three-way interactions	55	12.67	0.23	4.2	<0.001
Four-way interactions	27	1.55	0.06	1.1	0.394
Residual (pure) error	384	20.88	0.05		
$\log_e$(bias of Depletion +0.1)					
Main effects	9	4.66	0.52	18.7	<0.001
Two-way interactions	36	11.92	0.33	11.9	<0.001
Three-way interactions	55	7.07	0.13	4.6	<0.001
Four-way interactions	27	1.08	0.04	1.4	0.073
Residual (pure) error	384	10.65	0.03		
$\log_e$(CV of EndXBio)					
Main effects	9	131.08	14.56	2,000.0	<0.001
Two-way interactions	36	12.33	0.34	38.9	<0.001
Three-way interactions	55	3.66	0.07	7.6	<0.001
Four-way interactions	27	0.72	0.03	3.0	<0.001
Residual (pure) error	384	3.38	0.01		
$\log_e$(CV of EndF)					
Main effects	9	71.56	7.95	2,000.0	<0.001
Two-way interactions	36	7.99	0.22	42.3	<0.001
Three-way interactions	55	2.02	0.04	7.0	<0.001
Four-way interactions	27	0.26	0.01	1.9	0.007
Residual (pure) error	384	2.01	0.01		
$\log_e$(CV of Depletion)					
Main effects	9	49.09	5.45	2,000.0	<0.001
Two-way interactions	36	11.28	0.31	94.1	<0.001
Three-way interactions	55	2.46	0.04	13.4	<0.001
Four-way interactions	27	0.20	0.01	2.2	0.001
Residual (pure) error	384	1.28	0.00		

of the bias of the mean for EndXBio, the mean square (MS) for the main effects was almost 15, 33, and 79 times as large as the MS values for the estimable two-, three-, and four-way interactions, respectively.

Influential Factors

Across the 128 treatments, the grand mean values from the ANOVAs of the transformed bias estimates indicated small but statistically significant ($P < 0.01$) positive bias, ranging from a low of 1.2% (= exp[−2.19] − 0.1) for Depletion to a high of 6.3% for EndXBio (Table 6). For the transformed values of EndXBio, the two most influential factors were the number of years in the data series (NYrs) and the size of the age composition samples (SSize). The coefficients for these two factors were negative, indicating that longer data series and larger samples produced less biased estimates, but their interaction was positive and quite strong (fifth in absolute value), indicating that the joint effect was less than additive on the logarithmic scale. The main effects of the survey data variability (SurCV) and the fishing effort data variability (EffCV) were both strong (ranking third and sixth in absolute value) and positive, which indicates, as expected, that noisier data sets produced more biased estimates. For bias in EndF, the factors EffCV and SurCV as main effects were the first and second most influential terms, and the main effects NYrs and SSize ranked third and fourth. In contrast, for bias in Depletion the factors EffCV, SurCV, NYrs, and SSize were quite inconsequential as main effects, ranking 43rd, 36th,

TABLE 6.—Top 15 terms from fractional factorial analyses of variance for log-transformed values of the relative bias of the mean and the coefficient of variation (CV); see Table 3 for definitions of EndXBio, EndF, and Depletion.

Rank	Term	$\log_e$(bias + 0.1) Coefficient	$\log_e$(bias + 0.1) t-value	Term	$\log_e$(CV) Coefficient	$\log_e$(CV) t-value
			EndXBio			
	Grand mean	−1.8144	−233.8	Grand mean	−1.1933	−287.9
1	NYrs	−0.2467	−31.8	NYrs	−0.3676	−88.68
2	SSize	−0.1160	−14.9	SSize	−0.1935	−46.68
3	SurCV	0.0894	11.5	SurCV	0.1836	44.3
4	FInc	−0.0843	−10.9	FInc	−0.1521	−36.68
5	NYrs × SSize	0.0701	9.0	EffCV	0.1208	29.16
6	EffCV	0.0652	8.4	NatM	0.0964	23.26
7	NYrs × NatM	0.0557	7.2	EffCV × SurCV	0.0636	15.35
8	NYrs × SurCV × FInc	−0.0534	−6.9	NatM × FSel	−0.0633	−15.27
9	SurCV × FSel	0.0494	6.4	SurCV × FSel	0.0542	13.08
10	NYrs × EffCV × FInc	−0.0458	−5.9	SSize × FSel	0.0501	12.08
11	CatCV	0.0408	5.3	EffCV × CatCV	−0.0461	−11.11
12	EffCV × CatCV	−0.0380	−4.9	CatCV	0.0432	10.43
13	NatM × RecV	−0.0378	−4.9	NYrs × EffCV	0.0322	7.78
14	NYrs × FSel × RecV	0.0367	4.7	EffCV × FSel	0.0320	7.73
15	NYrs × SSize × FSel	−0.0366	−4.7	NYrs × SurCV × FInc	−0.0311	−7.51
	Standard deviation	0.0078		Standard deviation	0.0041	
			EndF			
	Grand mean	−1.9677	−190.9	Grand mean	−1.1162	−348.7
1	EffCV	0.1595	15.5	NYrs	−0.2190	−68.42
2	SurCV	0.1457	14.1	EffCV	0.1745	54.51
3	NYrs	−0.1233	−12.0	SSize	−0.1472	−45.98
4	SSize	−0.1205	−11.7	SurCV	0.1422	44.42
5	NYrs × NatM	−0.1002	−9.7	FInc	−0.0912	−28.48
6	NatM	0.0877	8.5	CatCV	0.0745	23.27
7	EffCV × SurCV	0.0855	8.3	NatM	0.0716	22.36
8	NYrs × FSel	−0.0742	−7.2	EffCV × SurCV	0.0683	21.33
9	NatM × FSel	−0.0695	−6.7	NatM × FSel	−0.0468	−14.63
10	NYrs × SSize × RecV	−0.0631	−6.1	SurCV × FSel	0.0401	12.53
11	NatM × FInc	−0.0619	−6.0	SSize × SurCV	−0.0359	−11.2
12	Finc	−0.0558	−5.4	SSize × EffCV	−0.0326	−10.17
13	SSize × EffCV	−0.0556	−5.4	SSize × FSel	0.0299	9.35
14	NYrs × NatM × FSel	0.0532	5.2	EffCV × FSel	0.0296	9.26
15	NYrs × SurCV × FSel	−0.0512	−5.0	SSize × EffCV × SurCV	−0.0266	−8.31
	Standard deviation	0.0103		Standard deviation	0.0032	
			Depletion			
	Grand mean	−2.1870	−297.1	Grand mean	−1.7398	−682.3
1	FSel	0.0763	10.4	SurCV	0.2062	80.88
2	NYrs × FInc	−0.0629	−8.6	SSize	−0.1345	−52.76
3	NYrs × NatM	0.0535	7.3	EffCV	0.1254	49.19
4	SSize × FSel	−0.0477	−6.5	FSel	0.1241	48.66
5	NYrs × SurCV × FInc	−0.0430	−5.8	EffCV × SurCV	0.0699	27.4
6	NYrs × FSel	0.0421	5.7	SurCV × FSel	0.0508	19.92
7	SSize × NatM	0.0384	5.2	NatM × FSel	−0.0503	−19.72
8	NatM	−0.0371	−5.0	EffCV × CatCV	−0.0440	−17.27
9	SurCV × FInc	0.0365	5.0	CatCV	0.0411	16.1
10	SurCV × FSel	0.0364	5.0	SSize × FSel	0.0397	15.58
11	NYrs × EffCV × FInc	−0.0359	−4.9	NYrs × SurCV	0.0380	14.89
12	Finc × CatCV	0.0348	4.7	SSize × SurCV	−0.0372	−14.6
13	NYrs × SSize	0.0347	4.7	SSize × EffCV	−0.0371	−14.53
14	NYrs × SurCV × FSel	0.0339	4.6	NatM	0.0356	13.95
15	NYrs × EffCV × FSel	0.0325	4.4	Nyrs × FSel	−0.0311	−12.2
	Standard deviation	0.0074		Standard deviation	0.0026	

76th, and 108th, respectively. The factor for the shape of the fishery selection curve (FSel) was the most influential factor for Depletion, with asymptotic fishery selection producing the more positively biased estimates. There was a very strong interaction between SSize and FSel (ranking fourth), indicating that the effect of FSel was amplified when coupled with small age composition samples. In general, the most influential terms for bias in Depletion were two- and three-way interactions between NYrs and other factors, even though NYrs was unimportant as a main effect. The factor for variation in recruitment (RecV) was the only factor that was not one of the top 15 influential terms as a main effect for EndXBio, EndF, or Depletion, although it was in several interactions that were among the top 15. The factor for catch data variability (CatCV) was also relatively unimportant, never ranking higher than 11th, either as a main effect or in an interaction.

For the ANOVAs of the transformed CV values, the overall grand mean values were significantly larger than zero ($P < 0.01$), ranging from a low of 17.6% (= exp[−1.740]) for Depletion to a high of 32.0% for EndF (Table 6). For the transformed values of both EndXBio and EndF, the five most influential factors were NYrs, SSize, SurCV, EffCV, and the trend in F (FInc), NYrs being the most significant factor for both dependent variables. As expected, the coefficients for NYrs and SSize were negative, indicating that longer data series and larger samples produced less variable estimates, and the coefficients for SurCV and EffCV were positive, indicating that the noisier data sets produced more variable estimates. The coefficients for FInc were negative, indicating that fisheries with more rapidly increasing rates of fishing produced less variable estimates. For the transformed values of Depletion, the factors SurCV, SSize, and EffCV were the three most influential factors, larger samples producing less variable estimates of Depletion and noisier data sets producing more variable estimates. The factor NYrs was surprisingly inconsequential for the CV of Depletion, appearing in the top 15 influential terms only in interactions with SurCV and FSel. Only the factor RecV was not one of the top 15 influential terms for EndXBio, EndF, or Depletion, either as a main effect or in an interaction.

Sensitivity to Initial Parameter Values

To evaluate the effect of initial parameter values, we conducted a small experiment with the two treatments that produced the most extreme results. The treatment ABcdeFghj (where a lowercase letter denotes the low level of that factor and an uppercase letter denotes the high level; Table 1) produced the best results for the estimate of EndBio, with an overall average relative bias of 0.2%. The treatment abCDeFGhJ, in contrast, produced the worst results for the estimate of EndBio, with an average relative bias of 44.0%. The two treatments are opposites for the factors NYrs, SSize, EffCV, SurCV, CatCV, and RecV, and, as one would expect, the best results were from the treatment that had the greatest amounts of data and the least variability in those data. Box-and-whisker plots for the 800 estimates from each of these treatments showed that the Stock Synthesis estimates of the relative bias of the mean for the seven focal variables were fairly symmetrically distributed for the best treatment and were highly right-skewed for the worst treatment (Figure 1). The estimates from the best treatment were considerably less variable than those for the worst treatment. For both the best and worst treatments, the estimates for all seven focal variables appeared to be more or less median-unbiased. Scatter plots of the individual

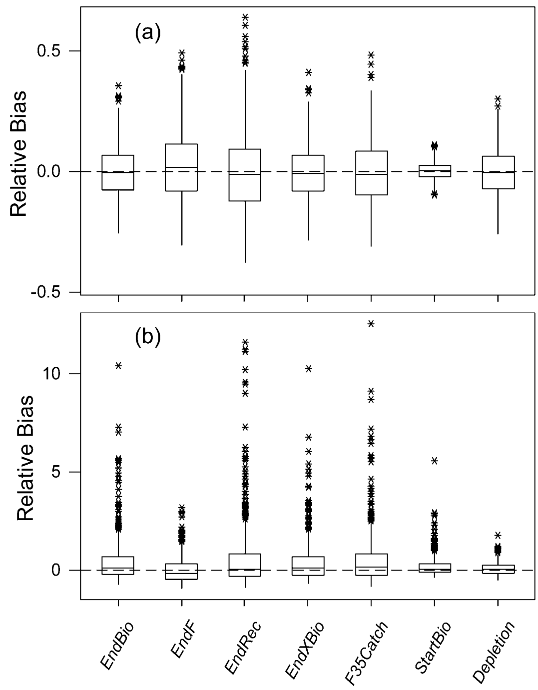

FIGURE 1.—Box-and-whisker plots of the estimates of relative bias for the seven focal variables (see Table 2) from **(a)** the treatment that produced the least biased estimate of ending biomass and **(b)** the treatment that produced the most biased estimate of ending biomass. Relative bias is defined as the ratio of (1) the mean estimated value less the true value and (2) the true value. The lower edge of each box corresponds to the first quartile of the data, the upper edge to the third quartile, and the horizontal line inside to the median value. The vertical lines (whiskers) extend to the extreme data values that are within 1.5 times the difference between the third and first quartiles. The asterisks indicate individual data values lying beyond the limits denoted by the whiskers. Note that the graphs have very different scales.

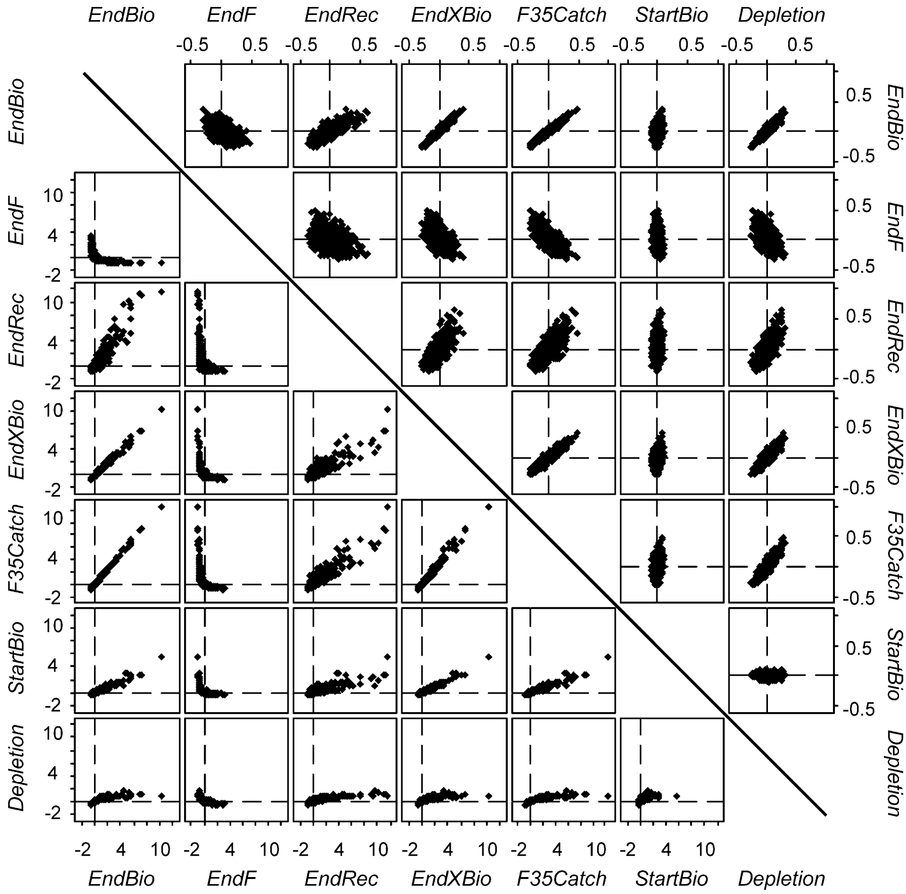

FIGURE 2.—Scatter plots of the estimates of relative bias for the seven focal variables (Table 2). Relative bias is defined as in Figure 1. The set of plots above the diagonal are from the treatment that produced the least biased estimate of ending biomass, and the set below the diagonal are from the treatment that produced the most biased estimate of ending biomass. Note that the graphs above and below the diagonal have very different scales.

estimates against one another for these two extreme treatments indicated fairly linear relationships among the estimates for EndBio, EndRec, EndXBio, F35Catch, StartBio, and Depletion and nonlinear relationships between these variables and the estimates for EndF (Figure 2). The curvature in the scatterplots involving EndF was particularly pronounced for the worst treatment.

For the treatment that produced the minimum relative bias in the EndBio estimates (treatment ABcdeFghj), using randomized initial parameter values (instead of the true parameter values) had essentially no effect on the final estimates. Averaged across the 100 replicates, the relative bias was 0.3% for both the randomized and the nonrandomized runs. For the treatment that produced the maximum relative bias in EndBio (treatment abCDeFGhJ), using randomized initial parameter values produced slightly bigger relative bias (52% versus 37%), although a two-sample *t*-test showed no strong evidence of a difference ($P = 0.238$). Thus, the initial parameter values appear to have little or no effect in scenarios where Stock Synthesis works well but may have some effect in scenarios where Stock Synthesis performs poorly.

Discussion

For all seven types of output variable examined in this study the estimates tended to be right-skewed and positively biased, whereas the median values appeared to be unbiased. If multiple estimates of an output variable were available, the median of those estimates would provide a better overall estimate than the mean of the estimates. However, in most applications of Stock Synthesis the program is applied to a single data set to produce a single set of output variables. Hence, for any given output variable there is only one data point with which to derive a median value and no practical advantage can be derived from the lack of bias of the median of multiple estimates. However, that the Stock Synthesis estimates tended to be skewed implies that one might obtain less biased and less variable estimates by using alternative parameterizations of the nonlinear population dynamics equations that underlie the Stock Synthesis program (Ratkowsky 1986).

In a previous investigation of the performance of Stock Synthesis (Sampson and Yin 1998), we conducted a similar but smaller experiment based on a one-eighth fraction of a 2^8 factorial design with 200 random data sets for each experimental treatment. The extra factor in our new experiment was the catch data variability (CatCV). The results of Sampson and Yin (1998) were generally in accord with the results from the larger experiment reported here, but the earlier experiment detected considerably fewer significant factors. For example, in the 1998 study there were 32 terms in the fractional factorial models, 16 of which were significant ($P < 0.05$) in the model for bias in EndXBio, whereas in the current study, which had 128 terms in the factorial model, there were 94 significant terms (73%). For the factor RecV the 1998 study did not find any significant ($P < 0.05$) effects on the bias of Stock Synthesis estimates, but the current study found significant effects on relative bias for six of the seven types of estimates. The new experiment had much greater power to detect small differences because of the increased number of replicates and the more complete factorial design.

Compared with the results of Sampson and Yin (1998), the results for the relative importance of the main effects were very similar for the measurements of relative variability but differed for some of the measurements of relative bias. For example, in the ANOVA of the relative variability in EndXBio, the absolute values of the coefficients for NYrs and SSize ranked first and second for both experiments. However, in the ANOVA of the relative bias in EndXBio, the absolute values of the coefficients for NYrs ranked first in the new experiment, whereas it ranked 14th in the 1998 experiment. We think such differences between the two studies were at least partially due to differences in experimental design. In the current study the main effects in the fractional ANOVAs were not confounded with any four-factor or lower-order interactions, whereas in the 1998 study each of the main effects was aliased with several three- and four-factor interactions.

Some results from our new experiment were counter to our expectations. The ratio between ending biomass and starting biomass reflects the degree of stock depletion, and this depletion ratio has become an important measure of stock status for the management of many fisheries (e.g., Restrepo et al. 1998). Stock Synthesis performed much better when estimating this depletion ratio than it did when estimating either the starting or ending biomass (Table 3), suggesting that estimates of depletion from Stock Synthesis may be fairly reliable even though estimates of absolute stock size are not. However, our results almost certainly overstate the program's true ability to estimate depletion. Real fisheries are not usually monitored until after several years of exploitation, whereas complete data series were available in all our simulations here. Also, our simulated stocks did not include any tendency for recruitment to decline with stock depletion. In real assessments the estimates of pristine stock size strongly depend on the assumed or estimated steepness parameter of the stock−recruit relationship.

Another surprising result was the relative unimportance of the level of sampling error in the catch data. Only very minor gains in accuracy resulted from reducing CatCV from 20% to 10%, even for treatments with the long data series, low level of natural mortality, and the large increment in fishing mortality, for which the effects of catch variability should be the most pronounced. Our study did not explore the effects of having biased estimates of catch, as would occur if there were unreported landings or no accounting for discards at sea. If unbiased estimates of total catch and age composition are available, as in our experiments, then obtaining highly precise estimates of total removals may not be that important. Rather than implementing a large-scale observer program to measure discards with great accuracy, a small-scale program might suffice. Accounting for the magnitude and characteristics of discards can be very important, however. Williams (2002) demonstrated that biased estimates of catch age composition due to unaccounted size-specific discards can produce very biased estimates of $F_{35\%}$.

While it is fairly standard practice to conduct sensitivity analyses that evaluate how model assumptions influence stock assessment results, systematic evaluation of assessment estimates against exact results are rare. Such comparisons can only be done by means of computer simulations with generated data for which exact results are known. A U.S. National Research Council committee (NRC 1998) conducted a limited evaluation of a suite of stock assessment models using a fixed set of simulated data (five treatments, each with one replicate) that they distributed to stock assessment scientists for independent analysis. The focus of the NRC study was not on evaluating the effects of data variability on assessment results (as in our studies) but rather on the effects of applying different assessment models to the same data and the analysts' subjective decisions about how to model the stock dynamics and data.

Punt et al. (2002) conducted a simulation study that evaluated the performance of six different stock assessment programs, including one (described as integrated analysis) that takes the same general approach as Stock Synthesis. The study developed operating models based on the characteristics of four Australian fish stocks and their associated fisheries. The model for each stock was used to generate 100 replicate random data sets that were then analyzed using each of the six stock assessment programs. The generated data sets had numerous sources of variability, including measurement error in the observations of catch, effort, length composition of landed and discarded fish, age-at-length keys, and survey estimates of abundance as well as process errors in recruitment, natural mortality, catchability, and selectivity. The study did not measure the effects of individual sources of data variability on the assessment results. Instead, the focus was on the comparison of the different stock assessment methods, and the authors concluded that the integrated analysis generally performed better than the other methods. As in our study, Punt et al. (2002) found that biomass depletion was estimated more accurately than ending biomass.

The annual catch quotas for many groundfish stocks on the U.S. Pacific coast are based on applying a target fishing rate such as $F_{35\%}$ to estimates of ending exploitable biomass (PFMC 2000). Overestimates of a catch quota are costly if they result in stock depletion and subsequent loss in yield; underestimates result in lost fishing opportunities. For the suite of scenarios examined in this study, Stock Synthesis under good conditions (long data series with large age composition samples and low variability in the fishing effort, survey biomass, and catch biomass series) produced reasonably accurate estimates of ending exploitable biomass and the $F_{35\%}$ catch. On average, the relative bias in the estimates of EndXBio and F35Catch were both less than 1% and the relative root mean square errors were 13% and 15%. However, under bad conditions (short data series with small age composition samples and high variability in the fishing effort, survey biomass, and catch biomass series), Stock Synthesis produced estimates that most people would consider to be unacceptable. The relative bias in the estimates of both EndXBio and F35Catch averaged 27%, and the relative root mean square errors were 84% and 98%. Catch quotas derived from these assessments would generally be much larger than they should be, which would probably result in the stocks being gradually overfished.

Collecting the input data for an assessment is a costly process that requires an appropriate balance of sampling resources. By adjusting the allocation of sampling effort between collecting catch data, age composition samples, and survey data, it may be possible to achieve the same level of assessment accuracy but at reduced cost. In our experiment the least accurate estimates of EndXBio and F35Catch resulted when there were short data series (NYrs = 8), small age composition samples (SSize = 100), and variable survey biomass estimates (SurCV = 80%). For these treatments, the average values for the relative root mean square error (RRMSE) were 72% for the EndXBio estimates and 83% for the F35Catch estimates. A fourfold increase in the number of age composition samples (SSize = 400) resulted in average RRMSE values of 43% for the EndXBio estimates and 49% for the F35Catch estimates. In contrast, a four-fold reduction in the variability of the survey biomass estimates (SurCV = 20% but SSize = 100) only reduced the average RRMSE values to 48% for the EndXBio estimates and to 54% for the F35Catch estimates. Thus, increased age composition sampling results in greater improvements in assessment accuracy than similar improvements in survey precision and almost certainly at much lower cost. A fourfold reduction in the survey biomass CV would require a sixteen-fold increase in the number of stations sampled. In terms of sampling efficiency, our results imply that the same overall assessment accuracy could be achieved at reduced cost from an increase in age composition sampling and a decrease in the amount of survey sampling.

Our study greatly simplified many of the problems encountered in actual stock assessments, and in this case Stock Synthesis almost certainly performed much better than it would in applications with data from actual fisheries. The simulated fisheries had simple linear trends in fishing mortality, the age composition data were not contaminated by age-reading error, and the Stock Synthesis models were always configured to have the same structural equations as the data simulator (e.g., the correct type of selection curves, initial conditions, and value for the natural mortality coefficient). Also, recruitment variability was much less than one might expect in many stocks, and the experiment considered only two patterns of recruitment variability. Furthermore, in our experiment Stock Synthesis was told the true form and levels of error associated with each data source (e.g., lognormal error with 20% CV for the survey estimates of biomass). In any real stock assessment application, these aspects of the problem would be hidden from the analyst and would thus contribute to uncertainty in the assessment results.

Some forms of structural uncertainty in stock assessment models can be resolved from the fit of the stock assessment model to the available data. For example, Helu et al. (2000) conducted a series of experiments in which they evaluated whether two particular statistical model selection criteria provide a reliable basis for selecting the correct structural equations in applications with Stock Synthesis. For the three equation forms and alternatives that they considered, they found that use of either the Akaike information criterion or Schwarz's Bayesian information criterion would usually lead to the correct choice of model form. Compared with using the wrong model, the accuracy of ending stock biomass estimates improved by about 25%.

One very important area of uncertainty that we did not investigate is the choice of weights for the different data components that contribute to the overall log-likelihood (NRC 1998). Applications of Stock Synthesis or other statistical catch-at-age methods to data from real fisheries are often left with a good model fit to one data source (e.g., survey estimates of stock biomass) at the expense of a poor model fit to a different data source (e.g., fishery age composition data). In this situation the maximum likelihood parameter estimates are largely determined by the relative weights assigned to the data sources. At present, there are no standard procedures for assigning weights to different data sources. Theory indicates that the weights should be inversely related to the variability associated with each data source (Quinn and Deriso 1999), but in practice the analyst is usually forced to subjectively evaluate the relative reliability of the alternative data sources. A challenge for the next generation of stock assessment procedures is to try to remove this subjectivity so that stock assessment can become less of an art and more of a science.

Acknowledgments

We are grateful to Richard Methot for answering our numerous questions about his Stock Synthesis program; our study would not have been possible without his assistance. We are also grateful for the helpful suggestions provided by two anonymous reviewers. This research was partially supported by grant NA36RG0451 (project R/OPF-50) from the National Oceanic and Atmospheric Administration (NOAA) to the Oregon Sea Grant College Program and by appropriations made by the Oregon state legislature. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies.

References

Bence, J. E., A. Gordoa, and J. E. Hightower. 1993. Influence of age-selective surveys on the reliability of stock synthesis assessments. Canadian Journal of Fisheries and Aquatic Sciences 50:827–840.

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. Statistics for experimenters: an introduction to design, data analysis, and model building. Wiley, New York.

Clark, W. G. 1991. Groundfish exploitation rates based on life history parameters. Canadian Journal of Fisheries and Aquatic Sciences 48:734–750.

Deriso, R. B., T. J. Quinn, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences 42:815–824. CAGEAN software available: http://www.fisheries.org/cus/. (June 2003).

Helu, S. L., D. B. Sampson, and Y. Yin. 2000. Application of statistical model selection criteria to the Stock Synthesis assessment program. Canadian Journal of Fisheries and Aquatic Sciences 57:1784–1793.

Hilborn, R., M. Maunder, A. Parma, W. Ernst, J. Payne, and P. Starr. 2003. COLERAINE: a generalized age-structured stock assessment model. University of Washington, Seattle. Available: http://www.fish.washington.edu/research/coleraine/. (June 2003).

Megrey, B. A. 1989. Review and comparison of age-structured stock assessment models from theoretical and applied points of view. Pages 8−48 *in* E. F. Edwards and B. A. Megrey, editors. Mathematical analysis of fish stock dynamics. American Fisheries Society, Symposium 6, Bethesda, Maryland.

Methot, R. D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. International North Pacific Fisheries Commission Bulletin 50:259–277.

Methot, R. D. 2000. Technical description of the Stock

Synthesis assessment program. NOAA Technical Memorandum NMFS-NWFSC-43.
Minitab. 1998. Minitab release 12 for Windows. Minitab, Inc., State College, Pennsylvania.
National Research Council. 1998. Improving fish stock assessments. National Academy Press, Washington, D.C.
Pacific Fishery Management Council. 2000. Status of the Pacific Coast groundfish fishery through 2000 and recommended acceptable biological catches for 2001. Pacific Fishery Management Council, Portland, Oregon.
Patterson, K. R. 1998. Integrated catch at age analysis, version 1.4. International Council for the Exploration of the Sea, Copenhagen. Available: http://www.ices.dk/datacentre/software.asp. (June 2003).
Punt, A. E., A. D. M. Smith, and G. Cui. 2002. Evaluation of management tools for Australia's southeast fishery, 2. How well can management quantities be estimated? Marine and Freshwater Research 53: 631–644.
Quinn, T. J., II, and R. B. Deriso. 1999. Quantitative fish dynamics. Oxford University Press, New York.
Ratkowsky, D. A. 1986. Statistical properties of alternative parameterizations of the von Bertalanffy growth curve. Canadian Journal of Fisheries and Aquatic Sciences 43:742–747.
Restrepo, V. R., G. G. Thompson, P. M. Mace, W. L. Gabriel, L. L. Low, A. D. MacCall, R. D. Methot, J. E. Powers, B. L. Taylor, P. R. Wade, and J. F. Witzig. 1998. Technical guidance on the use of precautionary approaches to implementing National Standard 1 of the Magnuson—Stevens Fishery Conservation and Management Act. NOAA Technical Memorandum NMFS-F/SPO-31.
Richards, L. J., and B. A. Megrey. 1994. Recent developments in the quantitative analysis of fisheries data. Canadian Journal of Fisheries and Aquatic Sciences 51:2640–2641.
Rubinstein, R. Y. 1981. Simulation and the Monte Carlo method. Wiley, New York.
Sampson, D. B., and Y. Yin. 1998. A Monte Carlo evaluation of the Stock Synthesis assessment program. Pages 315–338 *in* F. Funk, T. J. Quinn, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, and P. J. Sullivan, editors. Fishery stock assessment models. University of Alaska, Alaska Sea Grant College Program, Report AK-SG-98-01, Fairbanks.
Williams, E. H. 2002. The effects of unaccounted discards and misspecified natural mortality on harvest policies based on estimates of spawners per recruit. North American Journal of Fisheries Management 22:311–325.

Appendix: Detailed Stock Synthesis Approach

Stock Synthesis and Data Simulator Equations

The Stock Synthesis program and our data simulator program used the following deterministic equations to mimic the behavior of an age-structured fish stock. The variables and parameters are described in Table A.1; the specific parameter values used in the data simulations are given in Table A.2.

Number at age in year y for the terminal age-class (age $= T$):

$$N_{y,T} = N_{y-1,T} \cdot \exp(-M - F_{y-1} \cdot S_T) + N_{y-1,T-1} \cdot \exp(-M - F_{y-1} \cdot S_{T-1}) \quad \text{(A1)}$$

Number at age for nonterminal age-classes (age $< T$):

$$N_{y,a} = N_{y-1,a-1} \cdot \exp(-M - F_{y-1} \cdot S_{a-1}) \quad \text{(A2)}$$

Fishing mortality coefficient in year y:

$$F_y = y \cdot \text{FInc} \quad \text{(A3)}$$

Fishery selection at age:

$$S_a = S_a'/\max(S_a') \quad \text{(A4)}$$

where

$$S_a' = \{1 + \exp[-b_1 \cdot (a - a_1)]\}^{-1} \times \{1 + \exp[b_2 \cdot (a - a_2)]\}^{-1} \quad \text{(A5)}$$

Total stock biomass:

$$B_y = \sum_a N_{y,a} \cdot W_a \quad \text{(A6)}$$

where

$$W_a = 10 \cdot [1 - \exp(-0.2 \cdot a)]^3 \quad \text{(A7)}$$

Total exploitable biomass:

$$\text{XB}_y = \sum_a N_{y,a} \cdot W_a \cdot S_a \quad \text{(A8)}$$

Catch at age in numbers:

$$C_{y,a} = N_{y,a} \cdot \frac{F_y \cdot S_a}{M + F_y \cdot S_a} \cdot [1 - \exp(-M - F_y \cdot S_a)] \quad \text{(A9)}$$

Catch at age in weight (yield):

$$Y_{y,a} = C_{y,a} \cdot W_a \quad \text{(A10)}$$

Total catch biomass:

$$Y_y = \sum_a Y_{y,a} \quad \text{(A11)}$$

TABLE A.1.—Variables and parameters used in Stock Synthesis equations.

Symbol	Description
$N_{y,a}$	Number of fish at the start of year y that are age a
M	Instantaneous rate of natural mortality
F_y	Instantaneous rate of fishing mortality for fully selected age-classes in year y
S_a	Fishery selection coefficient for age-a fish
FInc	Annual increment in F
b_1, a_1	Parameters controlling the ascending portion of the double-logistic fishery selection curve
b_2, a_2	Parameters controlling the descending portion of the fishery selection curve
B_y	Stock biomass at the start of year y
W_a	Weight of age-a fish at the start of year
XB_y	Exploitable biomass at the start of year y
$C_{y,a}$	Number of age-a fish caught during year y
$Y_{y,a}$	Weight of age-a fish caught during year y
Y_y	Total catch biomass in year y
Mat_a	Proportion of age-a fish that are mature
b_{mat}, a_{mat}	Parameters controlling the logistic maturity-at-age curve
$P_{y,a}$	Proportion at age a at the start of year y
SurS_a	Survey selection coefficient for age-a fish
b_{sur}, a_{sur}	Parameters controlling the logistic survey selection curve
SurB_y	Survey biomass at the start of year y
SurQ	Survey catchability coefficient
f_y	Fishing effort during year y
Q	Fishery catchability coefficient
$L_{\mathrm{fishery\ age\ comp}}$	Log-likelihood component for observed fishery age composition
$L_{\mathrm{survey\ age\ comp}}$	Log-likelihood component for observed survey age composition
J_y	Number of fish sampled for age composition during year y
L_{survey}	Log-likelihood component for observed survey biomass
σS	Log-scale standard deviation of the lognormally distributed survey biomass observations
σF	Log-scale standard deviation of the lognormally distributed fishing effort observations
SurCV	Arithmetic-scale coefficient of variation of the lognormally distributed survey biomass observations
L_{effort}	Log-likelihood component for observed fishing effort
EffCV	Arithmetic-scale coefficient of variation of the lognormally distributed fishing effort observations
L_{total}	Total log-likelihood

TABLE A.2.—Data simulator input parameters grouped according to whether or not they vary with the instantaneous rate of natural mortality (M). Two values are given for parameters that vary with M; the first assumes that $M = 0.2$, the second (in parentheses) that $M = 0.4$. See Tables 1 and A.1 for symbol definitions.

Parameters	Value(s)
Parameters that do not vary with *M*	
Q	0.003
SurQ	0.1
F during first year	0.07
b_1	1
b_2	1 if FSel is domed, 0 otherwise
b_{sur}	1.5
Recruitment	3.0 million if RecV is constant, otherwise the sequence 3.5, 4.0, 1.2, 4.2, 3.0, 3.2, 1.7, and 3.2 million (repeated as necessary)
Parameters that vary with *M*	
Initial age (years)	4 (2)
Terminal age (years)	20 (10)
a_1	6 (4)
a_2	16 (8)
a_{mat}	5 (3)
b_{mat}	1 (2)
a_{sur}	5 (3)

Maturity at age:

$$\mathrm{Mat}_a = \{1 + \exp[-b_{\mathrm{mat}} \cdot (a - a_{\mathrm{mat}})]\}^{-1} \tag{A12}$$

Parameter Estimation

The Stock Synthesis program selects estimates for unknown parameters using the method of maximum likelihood. For each type of observed data there is a separate log-likelihood component that measures the goodness of fit between the observed values and the values predicted by the Stock Synthesis equations given the current set of parameter values. The maximum likelihood estimates are the set of parameter values that maximize the sum of all the log-likelihood components.

Predicted fishery catch proportion at age:

$$E(pC_{y,a}) = C_{y,a} \Big/ \sum_a C_{y,a} \tag{A13}$$

Predicted survey proportion at age:

$$E(pS_{y,a}) = N_{y,a} \cdot \mathrm{SurS}_a \Big/ \sum_a N_{y,a} \cdot \mathrm{SurS}_a \tag{A14}$$

$$\mathrm{SurS}_a = \mathrm{SurS}'_a / \max(\mathrm{SurS}'_a) \tag{A15}$$

$$\mathrm{SurS}'_a = \{1 + \exp[-b_{\mathrm{sur}} \cdot (a - a_{\mathrm{sur}})]\}^{-1} \tag{A16}$$

Predicted survey biomass:

$$E[\mathrm{SurB}_y] = \mathrm{SurQ} \sum_a N_{y,a} \cdot W_a \cdot \mathrm{SurS}_a \tag{A17}$$

Predicted fishing effort:

$$E[f_y] = F_y / Q \tag{A18}$$

Log-likelihood components for observed age composition data (multinomial error):

$$L_{\text{fishery age comp}}, L_{\text{survey age comp}} = \sum_y J_y \sum_a \{p \log_e[E(p)] - p \log_e(p)\} \tag{A19}$$

where p denotes $pC_{y,a}$ for the catch age compositions and $pS_{y,a}$ for the survey age compositions.

Log-likelihood component for observed survey biomass data (lognormal error):

$$L_{\text{survey}} = -\log_e(\sigma S) - \sum_y \frac{1}{2 \cdot \sigma S^2} \cdot \left\{ \log_e \left[\frac{\mathrm{SurB}_y}{E(\mathrm{SurB}_y)} \right] \right\}^2 \tag{A20}$$

$$\sigma S = \sqrt{\log_e(1 + \mathrm{SurCV}^2)} \tag{A21}$$

Log-likelihood component for observed fishing effort data (lognormal error):

$$L_{\text{effort}} = -\log_e(\sigma F) - \sum_y \frac{1}{2 \cdot \sigma F^2} \cdot \left\{ \log \left[\frac{f_y}{E(f_y)} \right] \right\}^2 \tag{A22}$$

$$\sigma F = \sqrt{\log_e(1 + \mathrm{EffCV}^2)} \tag{A23}$$

Total log-likelihood:

$$L_{\text{total}} = L_{\text{fisher age comp}} + L_{\text{survey age comp}} + L_{\text{survey}} + L_{\text{effort}} \tag{A24}$$

The total log-likelihood does not include a component for error in the catch observations. Stock Synthesis assumes that the catch data are known without error and selects values for F that generate an exact correspondence between the observed and predicted values of catch biomass.

Stock Synthesis allows the user to specify weights to differentially emphasize individual log-likelihood components. For our study, we gave all components equal weight.