Performance of time-varying catchability estimators in statistical catch-at-age analysis

Michael J. Wilberg and James R. Bence

Abstract: We used Monte Carlo simulations to evaluate how different methods of estimating fishery catchability within statistical catch-at-age analysis (SCA) performed when fishery catchability changed over time. Data-generating models included cases where catchability changed abruptly or gradually over time and where catchability was explicitly a function of population abundance, and we considered corresponding estimation models. In many cases, including fishery effort data in the estimation model and allowing catchability to follow a random walk provided the best (or nearly best) estimates of biomass in the last year as measured by the median of the absolute value of the relative error. Exceptions were cases where fishing mortality was low and catchability trended over time. The estimation model that ignored fishery effort data performed well in cases with a good survey, but performance degraded as survey precision decreased. The white noise estimation model performed poorly in situations where catchability trended over time. No estimation model was best for all underlying models of catchability, but the random walk estimation model performed well under most circumstances and should be used as a starting point for SCAs.

Résumé : Des simulations de Monte Carlo nous ont servi à évaluer comment fonctionnent les différentes méthodes d'estimation de la capturabilité de pêche dans des modèles statistiques de captures en fonction de l'âge (SCA), lorsque la capturabilité de pêche varie dans le temps. Les modèles de génération de données englobent des situations dans lesquelles la capturabilité change abruptement ou graduellement dans le temps et dans lesquelles la capturabilité est explicitement une fonction de l'abondance de la population et nous avons considéré les modèles d'estimation correspondants. Dans plusieurs cas, inclure les données d'effort de pêche dans le modèle d'estimation et laisser la capturabilité suivre un trajet aléatoire fournissent les meilleures (ou presque les meilleures) estimations de la biomasse de l'année précédente, d'après la mesure de la médiane de la valeur absolue de l'erreur relative. Les exceptions incluent les cas où la mortalité due à la pêche est basse et la capturabilité suit une tendance temporelle. Le modèle d'estimation qui ne tient pas compte des données d'effort de pêche fonctionne bien dans les cas qui comportent un bon inventaire, mais sa performance se dégrade à mesure que la précision des inventaires diminue. Le modèle d'estimation de bruit blanc fonctionne mal dans les situations où la capturabilté suit une tendance temporelle. Aucun des modèles d'estimation ne s'avère être le meilleur pour tous les modèles sous-jacents de capturabilité, mais le modèle d'estimation de trajet aléatoire fonctionne bien dans la plupart des circonstances et devrait servir de point de départ pour les SCA.

[Traduit par la Rédaction]

Introduction

Statistical catch-at-age analysis (SCA) is used to estimate abundance, recruitment, and fishing mortality for many exploited fish stocks throughout the US and the rest of the world (National Research Council (NRC) 1998; Quinn and Deriso 1999). In contrast to virtual population analysis and its variants, SCAs generally assume that fishing mortality rate at age can be modeled as a function of a year effect and an age effect (selectivity). This approach allows statistical estimation in which fishery catch-at-age data are assumed to be measured with error (Fournier and Archibald 1982; Megrey 1989). These models require catch-at-age data, as well as an index of abundance; other data sources can also be included in the model (Deriso et al. 1985). Under many conditions, SCA provides more accurate estimates of stock size and other important management quantities than other stock assessment techniques (NRC 1998; Punt et al. 2002; Radomski et al. 2005).

Many SCAs use fishery catch-per-unit-effort (CPUE) as an index of relative abundance and thus assume that fishery CPUE is proportional to abundance (Quinn and Deriso 1999). However, violations of this assumption can cause SCA models (and other stock assessment models) to produce

Received 8 September 2005. Accepted 18 May 2005. Published on the NRC Research Press Web site at http://cjfas.nrc.ca on 29 September 2006.
J18878

M.J. Wilberg[1,2] and J.R. Bence. Department of Fisheries and Wildlife, Michigan State University, 13 Natural Resources Building, East Lansing, MI 48824, USA.

[1]Corresponding author (e-mail: wilberg@cbl.umces.edu).
[2]Present address: Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, P.O. Box 38, Solomons, MD 20688, USA.

 doi:10.1139/F06-111

biased estimates (NRC 1998). Time-varying catchability has been documented in a wide range of fisheries, spanning commercial and recreational fisheries and freshwater and marine systems. In some cases, catchability may change with abundance or the area inhabited by a stock (e.g., Peterman and Steer 1981; Winters and Wheeler 1985; Harley et al. 2001), environmental effects (Ziegler et al. 2003), or changes in fisherman behavior or gear (Hilborn and Walters 1992). Interactions between population size, stock area, and fisherman behavior can lead to hyperstable fishery CPUE, where CPUE remains high despite decreases in abundance (Hilborn and Walters 1992; Harley et al. 2001). Hyperstable CPUE in combination with a stock assessment model that does not account for this can lead to overestimated stock size and catch limits (NRC 1998).

Methods have been developed to account for time-varying fishery catchability (e.g., Fournier and Archibald 1982; Fournier 1983; Methot 1990), but there is little consensus about best practices in this area. Generally, fishery effort or CPUE data are ignored if an independent survey is available for a stock (NRC 1998). However, in many fisheries, survey data are not available and ignoring fishery effort data is not an option (NRC 1998). Likewise, ignoring fishery effort data may decrease the accuracy and precision of SCA estimates in some cases because fishery CPUE may be informative about changes in relative population size in comparison to survey data. Our objective was to determine how well different methods of estimating time-varying catchability performed within an SCA framework. Specifically, we tested four estimation models to determine how well they performed in scenarios where catchability changed over time.

Materials and methods

We used Monte Carlo simulations to compare how four different methods of estimating fishery catchability within an SCA model performed when models were confronted with different data-generating scenarios. Our data-generating models included five cases where catchability changed abruptly or gradually over time and where catchability was explicitly a function of population abundance. Our data-generating models also contained three levels of fishing mortality and three levels of survey measurement error. Although the general influence of fishing mortality level and survey measurement error on the performance of SCA methods is well understood (e.g., Bence et al. 1993), we included these factors to determine whether they act to change the relative performance of different approaches of modeling time-varying catchability. We generated 1000 data sets for each scenario (45 total scenarios). For data sets that included survey data, we fitted each data set with four different models that made different assumptions regarding fishery catchability: catchability was modeled as white noise, a random walk, or density-dependent or catchability was effectively estimated as a free parameter for each year. This last method ignores any information contained in fishery CPUE or effort because the fishing mortality rate for fully selected age classes is estimated as a parameter during the model fitting procedure (rather than calculated from estimated catchability and observed effort). This method can only be used if an independent index of abundance is available. For data sets that did not include survey data, we used the first three of these estimation methods. Each scenario used the same sets of random numbers.

All models contained 15 years of data and eight age classes, with the last age class representing all fish that age and older. Data-generating models were based on commercial fisheries for lake whitefish (*Coregonus clupeaformis*) in the upper Great Lakes. Symbols and equations defining the data-generating models and estimation models are presented in Tables 1, 2, and 3. Equations are referred to in the text as eq. Tx.y, where x is the table number and y is the equation number within Table x. To avoid redundancy, equivalent quantities and parameters in estimation and data-generating models are not differentiated except when they both appear in the same equation, in which case, estimated quantities are denoted with a caret above the symbol.

Data-generating model

The data-generating model described the population dynamics and created data sets of total fishery catch, the age composition of the fishery catch, and in some scenarios, total survey CPUE and the age composition of the survey. For the population dynamics, we used an age-structured model that followed cohorts over time. Recruitment (abundance at age 1) was generated from a lognormal distribution with a log-scale standard deviation (SD) of 100%. The SD is approximately equal to the coefficient of variation on the arithmetic scale. Numbers-at-age in the first year were calculated assuming a stable age distribution with lognormal errors, where recruitment and mortality rates before the first year of the simulation were on average the same as in the first year (eq. T2.1). Cohorts were tracked over time by applying a simple exponential mortality model (eq. T2.2*a*); the last age class was treated as representing all fish age 8 and older (eq. T2.2*b*). Biomass each year was the sum of age-specific abundance multiplied by mean weight-at-age (eq. T2.3).

We used a separable model to generate fishing mortality rates. Total mortality rates were determined by the natural mortality rate and age-specific fishing mortality rates (eq. T2.4). Instantaneous natural mortality rate (M) was held constant across ages and years at 0.25. Instantaneous fishing mortality rate was a function of catchability, fishing effort, and age-specific selectivity (eq. T2.5). We used three levels of fishing mortality where F at fully selected ages was approximately $2M$ (high), M (medium), and $0.5M$ (low). We allowed fishing mortality to change over time by allowing effort to change (Fig. 1) and by incorporating several processes of time-varying catchability (see below). The pattern of increasing then decreasing effort provides contrast in fishing mortality rates, which is different than the pattern of catchability, and can be thought of as simulating periods of fishery development and subsequent regulation that reduces effort (although this is probably not the ideal evolution of management for a fishery). For a given level of fishing mortality, each of the models used the same effort series and each effort series had the same amount of contrast in absolute terms. The selectivity pattern for the fishery was dome-shaped to simulate a gill net fishery (Fig. 2).

We included five models for time-varying catchability, incorporating a range of possible ways that catchability could vary over time. The log of catchability was modeled as

Table 1. Symbols and descriptions of variables for data-generating and estimation models.

Symbol	Description	Value (if needed in the data-generating model)
$\bar{R}$	Average recruitment (numbers)	1 000 000
$N_{y,a}$	Abundance by age and year (numbers)	
B_y	Biomass (weight)	
$Z_{y,a}$	Total instantaneous mortality rate by age and year (year^{-1})	
$F_{y,a}$	Instantaneous fishing mortality rate by age and year (year^{-1})	
M	Instantaneous natural mortality rate (year^{-1})	0.25
$s_{a,f}$	Fishery age-specific selectivity	See Fig. 2
$s_{a,s}$	Survey age-specific selectivity	See Fig. 2
E_y	Fishery effort (effort)	See Fig. 1
$q_{y,f}$	Fishery catchability (effort^{-1})	
q_s	Survey catchability (effort^{-1})	0.0001
$\bar{q}_f$	Mean fishery catchability (effort^{-1})	0.05
$C_{y,a}$	Expected fishery catch-at-age (numbers)	
$I_{y,a}$	Expected survey catch-at-age (numbers)	
$\tilde{C}_y$	Observed total fishery catch (numbers)	
$\tilde{I}_y$	Observed total survey catch (numbers)	
$u_{y,a,f}$	Proportion of catch-at-age in fishery	
$u_{y,a,s}$	Proportion of catch-at-age in survey	
w_a	Mean weight-at-age (weight)	0.16, 0.45, 0.82, 1.2, 1.55, 1.86, 2.11, 2.3
γ_a	Deviations from mean recruitment	
δ_y	Deviations for white noise catchability	
ε_y	Deviations for first-order autoregressive catchability	
ω_y	Deviations for random walk catchability	
α	Parameter for density-dependent catchability (low, medium, high)	175; 90; 35
β	Parameter for density-dependent catchability (low, medium, high)	0.53; 0.49; 0.42
a, b	Parameters for linear increase in catchability	0.032, 0.00225
q_1, q_2	Parameters for abrupt change in catchability	0.0402, 0.0598
f_y	Fishing intensity by year (year^{-1})	
ρ	Correlation parameter for autoregressive catchability	0.9
σ_γ	Log-scale standard deviation (SD) for recruitment variation	1.0
σ_τ	Fishery measurement error SD	0.1
σ_υ	Survey measurement error SD	0.25; 1.0
σ_δ	SD for white noise catchability deviations	0.2
σ_ε	SD for autoregressive catchability deviations	0.16
σ_ω	SD for random walk catchability deviations	0.165

Note: Units for most model quantities are arbitrary, but we denote the type of units in parentheses after the description. Unitless quantities have no parentheses. Units for effort and catchability are arbitrary, and therefore, units on quantities that describe catchability are not displayed.

white noise to simulate a fishery where catchability varied from year to year about a constant mean (eq. T2.6), perhaps resulting from environmental effects, but where year-to-year deviations were not correlated. We also included four treatments that had varying amounts of autocorrelation: first-order autoregressive (AR(1)), density-dependent, linear increase, and abrupt change. The AR(1) process was also on the log scale (eqs. T2.7*a*, T2.7*b*). The AR(1) process simulates a case where catchability varies about its mean in an autocorrelated way but is not related to population abundance. We set the correlation (ρ) of the AR(1) process to 0.9 and the SD (σ_ε) to 0.16. This level of autocorrelation and SD produced 15-year time series that on average had a sample SD of about 0.2. This sample SD is lower than the expected stationary SD because the time series is not long enough to display the full range of dynamics of an autocorrelated process with these parameters. Density-dependent catchability followed a power relationship where catchability declined with increasing abundance (eq. T2.8; Fig. 3; Paloheimo and Dickie 1964). Because each of the different levels of fishing mortality had different average levels of abundance, we used three sets of parameters (α and β) to define the density-dependent power function, one for each level of fishing mortality. In the linear increase scenario, catchability increased linearly over time (eq. T2.9), which could represent learning by fishers or increases in gear efficiency. In the abrupt change scenario, catchability was constant until year 8 of the time series and increased to a higher level, where it re-

Table 2. Data-generating and estimation model equations.

	Model equations	Application
Population model		
(T2.1)	$N_{1,a} = \bar{R}\mathrm{e}^{-\sum_{a=1}^{A-1} Z_{1,a} + \gamma_a}$; $\gamma_a \sim N(0, \sigma_\gamma^2)$	Generation
(T2.2*a*)	$N_{y+1,a+1} = N_{y,a}\,\mathrm{e}^{-Z_{y,a}}$	Both
(T2.2*b*)	$N_{y+1,8} = N_{y,7}\,\mathrm{e}^{-Z_{y,7}} + N_{y,8}\,\mathrm{e}^{-Z_{y,8}}$	Both
(T2.3)	$B_y = \sum_a N_{y,a} w_a$	Both
(T2.4)	$Z_{y,a} = M + F_{y,a}$	Both
(T2.5)	$F_{y,a} = q_y E_y s_a$	Both
Catchability model		
(T2.6)	White noise: $\log q_{y,f} = \log \bar{q}_f + \delta_y$; $\delta_y \sim N(0, \sigma_\delta^2)$	Both
(T2.7*a*)	First-order autoregressive: $\log q_{1,f} \sim N\left(\log \bar{q}_f, \frac{\sigma_\varepsilon^2}{1-\rho^2}\right)$	Generation
(T2.7*b*)	$\log q_{y+1,f} = \log \bar{q}_f + \rho(\log q_{y,f} - \log \bar{q}) + \varepsilon_y$; $\varepsilon_y \sim N(0, \sigma_\varepsilon^2)$	Generation
(T2.8)	Density-dependent: $q_{y,f} = \alpha N_y^{-\beta}$	Both
(T2.9)	Linear increase: $q_{y,f} = a + by$	Generation
(T2.10)	Abrupt change: $q_{y,f} = \begin{cases} q_1 & \text{if } y < 8 \\ q_2 & \text{if } y \geq 8 \end{cases}$	Generation
(T2.11)	$\log q_{y+1,f} = \log q_{y,f} + \omega_y$; $\omega_y \sim N(0, \sigma_\omega^2)$	Estimation
(T2.12)	Freely estimate f_y (ignore fishery effort): $F_{y,a} = f_y s_{a,f}$	Estimation
Observation model		
(T2.13)	$C_{y,a} = \frac{F_{y,a}}{Z_{y,a}}(1 - \mathrm{e}^{-Z_{y,a}})N_{y,a}$	Both
(T2.14)	$\tilde{C}_y = \mathrm{e}^{\tau_y} \sum_a C_{y,a}$; $\tau_y \sim N(0, \sigma_\tau^2)$	Both
(T2.15)	$I_{y,a} = q_s s_a N_{y,a}$	Both
(T2.16)	$\tilde{I}_y = \mathrm{e}^{\upsilon_y} \sum_a I_{y,a}$; $\upsilon_y \sim N(0, \sigma_\upsilon^2)$	Both

mained for the rest of the time series (eq. T2.10). This scenario simulated the adoption of a more efficient technology by the fishery. All models were parameterized to have the same expected catchability (over the time series) and similar variances of log q_f (Fig. 4). We achieved this by simulating data sets and adjusting the catchability parameters until the mean and variance of catchability were the same as in the white noise case. We used a value of 0.2 for the SD as the standard for all other catchability models. This value is similar to estimates of the SD of catchability for commercial fisheries in New Zealand (Francis et al. 2003) but was less than median values of the SD of fishery CPUE estimated by Harley et al. (2001) for International Council for the Exploration of the Sea fisheries of 0.4–0.8, which should be an upper bound for catchability variation.

Fishery catch was calculated with the Baranov catch equation (eq. T2.13; Quinn and Deriso 1999). We multiplied total catch by a lognormal measurement error to calculate observed fishery catch (eq. T2.14); the measurement error SD for fishery catch was 10%. Observed age compositions were generated by drawing a sample from a multinomial distribution of size *n* (100 for the fishery) with proportions equal to the expected catch-at-age in the fishery. Survey CPUE-at-age was calculated as the product of survey catchability, abundance, and survey selectivity (eq. T2.15), and observed survey CPUE was the product of total survey CPUE and a lognormal measurement error (eq. T2.16). Our simulation model contained three levels of survey quality with differing levels of measurement error: good SD = 0.25, poor SD = 1.0, and no survey. Catchability of the survey was constant over time. Observed survey age compositions were generated by drawing a random sample from a multinomial distribution of size 75 with proportions equal to the expected CPUE-at-age in the survey. We chose a slightly smaller effective sample size for the survey to simulate a situation where ages are determined for fewer individuals in the survey than in the fishery catch.

Estimation model

The estimation models were largely the same as the simulation models except for how catchability was estimated and how numbers-at-age in the first year and recruitments were

Table 3. Objective function equations for statistical catch-at-age analysis simulation study.

(T3.1)	$L = \sum_i \ell_i$	Objective function
(T3.2)	$\ell_1 = \frac{1}{2\sigma_\tau^2}\sum_y (\log(\tilde{C}_y) - \log(\hat{C}_y))^2$	Fishery catch
(T3.3)	$\ell_2 = \frac{1}{2\sigma_\upsilon^2}\sum_y (\log(\tilde{I}_y) - \log(\hat{I}_y))^2$	Survey catch-per-unit-effort
(T3.4)	$\ell_3 = -n_f \sum_y \sum_a u_{y,a,f} \log(\hat{u}_{y,a,f})$	Proportion-at-age in the fishery catch
(T3.5)	$\ell_4 = -n_s \sum_y \sum_a u_{y,a,s} \log(\hat{u}_{y,a,s})$	Proportion-at-age in the survey catch
(T3.6)	$\ell_5 = \frac{1}{2\sigma_\delta^2}\sum_y (\hat{\delta}_y)^2$	White noise catchability
(T3.7)	$\ell_5 = \frac{1}{2\sigma_\omega^2}\sum_y (\hat{\omega}_y)^2$	Random walk catchability

Note: Equations T3.3 and T3.5 were only used in estimation models that considered survey data. Equations T3.6 and T3.7 were only used in estimation models that modeled fishery catchability as white noise or a random walk, respectively.

Fig. 1. Effort series used for low (dotted line), medium (broken line), and high (solid line) fishing mortality scenarios. The average fishing mortality rates for fully selected age classes were approximately 2M for the high scenario, M for the medium scenario, and 0.5M for the low scenario.

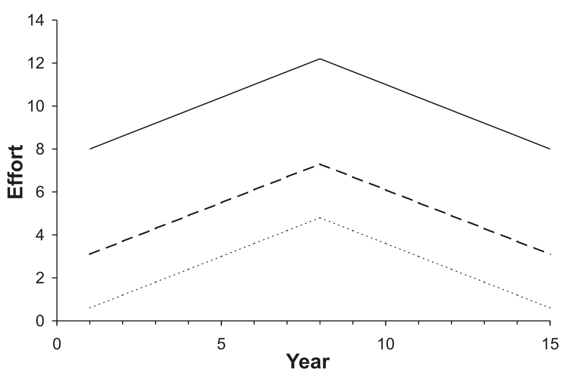
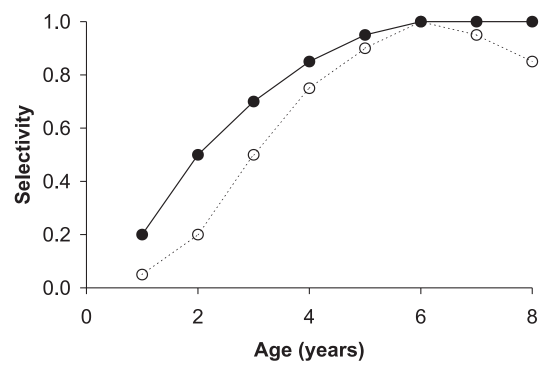


Fig. 2. Fishery (dotted line and open circles) and survey selectivity (solid line and solid circles) patterns used in the data-generating model.



handled. Common parameters among models included $N_{1,1},\ldots,N_{15,1}$ (recruitment), $N_{1,2},\ldots,N_{1,8}$ (numbers-at-age in the first year), and $s_{1,f},\ldots,s_{7,f}$ (fishery selectivity); models with surveys also included $s_{1,s},\ldots,s_{7,s}$ (survey selectivity) and q_s (survey catchability). Numbers-at-age in the first year and recruitment for each year were estimated as parameters during the model fitting process. After the first year and age, abundance-at-age followed a standard exponential mortality model with the last age representing all fish that age and older (eqs. T2.2a, T2.2b).

The total mortality rate ($Z_{y,a}$) was the sum of M and $F_{y,a}$ (eq. T2.4); M was assumed known at 0.25 (the true value from the simulation models). Fishing mortality followed a separable model for all of our estimation models. Fishery and survey selectivities were estimated as individual parameters by setting selectivity at the oldest age class to one; selectivity for other ages were only constrained to be positive values, so fixing the selectivity of the oldest age class to one did not assume an asymptotic selectivity pattern. Estimation models contained four methods of estimating catchability: white noise, random walk, density-dependent, and no catchability (directly estimating fishing mortality) with survey data. The first estimation model allowed fishery catchability to vary with white noise about a constant mean (eq. T2.6). The second estimation model allowed fishery catchability to vary according to a random walk (eq. T2.11); in this estimation model, catchability in the first year is an estimated parameter. The third estimation model allowed catchability to be a density-dependent function (eq. T2.8). The density-dependent model did not contain any random deviations in order to match the assumptions from the density-dependent catchability data-generating model. In our fourth estimation model, we estimated the fishing mortality rate for fully selected age classes as a parameter and then applied the estimated fishery selectivity to calculate age-specific fishing mortality rates (eq. T2.17). This method does not use fishery effort as a data source. The estimation models also predicted proportions of fishery and survey catch-at-age.

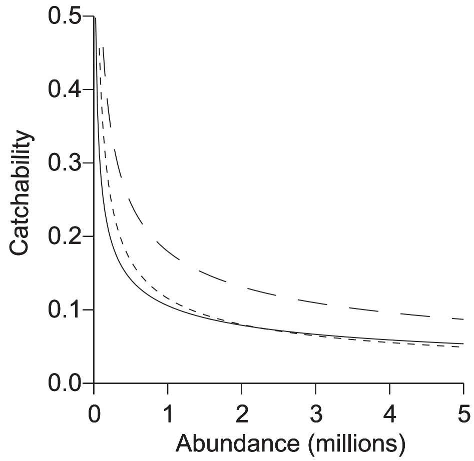


Fig. 3. Density-dependent fishery catchability for low (short dash), medium (long dash), and high (solid line) fishing mortality scenarios in the data-generating model.

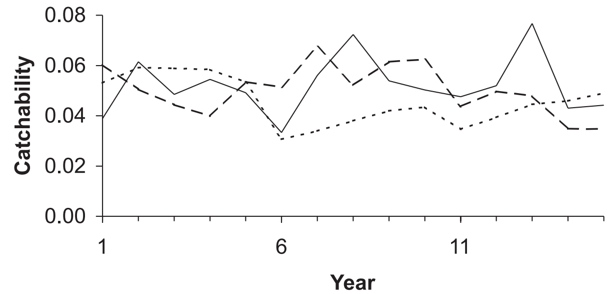


Fig. 4. Examples of 15-year time series of catchability generated by a white noise process (solid line), an autocorellated process (long dashed line), and a density-dependent process (short dashed line). Each of these time series has a sample standard deviation of about 0.2 on the log scale.

Model fitting and convergence

We fitted the models using a likelihood-based approach where we used a numerical search to find parameter values that minimized our objective function. The objective function was the sum of the likelihood components and each component was the negative of the log-likelihood for a single data source or a penalty related to time-varying catchability (eq. T3.1).

Our estimation models assumed lognormal distributions of errors for total catch for the fishery (eq. T3.2) and survey CPUE (eq. T3.3) and multinomial distributions for age compositions of the fishery (eq. T3.4) and the survey (eq. T3.5; Fournier and Archibald 1982). Effective sample sizes and SDs of the fishery and survey catch and age compositions were set to their true values from the data-generating models. The likelihood components for survey CPUE and age composition were only included in models that included survey data.

For estimation models that used fishery effort as a data source, fishery CPUE was not explicitly modeled. Instead, fishing mortality was an explicit function of effort, and catch was linked to abundance and fishery effort by estimating the catchability coefficient. We assumed lognormal deviations for catchability in the white noise (eq. T3.6) and random walk (eq. T3.7) estimation models. The SD for the white noise catchability was set to the true expected value, which was 0.2 for all data-generating models. For the random walk model, we set the SD to 0.165, the SD that on average created a time series with a sample SD of 0.2. This component in the objective function can be thought of as a penalty that produces a shrinkage estimator (in the Frequentist case) or as a Bayesian prior and penalizes large deviations from mean catchability (for the white noise model) or large year-to-year deviations (in the random walk model). Estimation models that contained density-dependent catchability or ignored effort data did not contain likelihood component ℓ_5.

We minimized the objective function iteratively using an efficient quasi-Newton implementation in AD Model Builder software (Otter Research Limited 2000) that takes advantage of automatic differentiation. We minimized the objective function in stages, where the initial stages were penalized if the model estimates deviated from the expected average fishing mortality rates under each scenario (early stages can be viewed as providing starting values for subsequent stages). This constraint was removed for the final stage of fitting and therefore did not penalize final model estimates. Iterative adjustment of the parameters terminated when the maximum gradient of parameters with respect to the objective function was less than 0.0001, or more than 1000 function evaluations had occurred. We denoted any terminated parameter estimates where the maximum gradient component was less than 0.0005 as converged, based on trial investigations after the completion of the simulations that used different parameter starting values.

Evaluation of estimation model performance

In stock assessments, estimated quantities in the last year are often most important for forecasting and management. Therefore, we evaluated estimation model performance by calculating the relative error (RE) of estimated biomass in the last year:

$$(1) \qquad \mathrm{RE} = \frac{\text{estimated} - \text{true}}{\text{true}} \times 100$$

We report only results for stock size measured in biomass. Other common measures of stock size (e.g., measures of exploitable abundance) showed similar patterns and estimates of fully selected F or exploitation rate reflected similar but inverse patterns (i.e., if estimated biomass was higher than the true value, estimated F was usually lower than the true value and vice versa). We evaluated systematic over- or under-estimation using the median of the relative error (MRE). If MRE equals zero, half of the estimates are higher than the true value and half are lower than the true value. Throughout the rest of the paper, we use the term unbiased as meaning median unbiased (i.e., MREs near zero). We also compared estimation model performance using the median of the absolute values of relative error (MARE), which indicates the width of the distribution of REs if the median is zero. In situations where the REs are either all (or mostly) positive or negative, the MRE will equal the MARE. We compared relative performance of the estimation models by calculating the difference of their MAREs and report these differences as percentages because MAREs are reported as percentages. We used MRE and MARE instead of mean relative error and root mean squared error because mean values were heavily influenced by several cases with large relative errors (>100). We checked whether these outliers represented false convergence by restarting the estimation with different starting values. Convergence was verified and we obtained the same parameter estimates.

Results

All estimation models performed best in situations with high fishing mortality and low survey SD and worst in cases with low fishing mortality and no survey (Table 4, high, good). The performance of a given estimation model depended on the level of fishing mortality, survey quality, and data-generating model. In almost all cases, estimation models that made use of both survey CPUE and fishery effort outperformed models that used only fishery effort or survey CPUE. Performance of the estimation model that ignored fishery effort data was independent of the underlying catchability model that generated the data and was only a function of survey quality and fishing mortality. The estimation model that ignored effort data was relatively unbiased (MRE near zero) in all cases, but the MARE was often significantly higher than for estimation models assuming white noise and random walk catchability, and the relative performance of this method was highly dependent on survey quality. Small differences in performance among different catchability scenarios arose because differences in catchability dynamics created different patterns of fishing mortality and therefore slightly different population dynamics. For the other estimation models, the results can be separated into two categories: (*i*) all estimation models were relatively unbiased (scenarios without consistent trends in catchability; Table 4, WN, AR, DD) and (*ii*) some estimation models had substantial bias (linear increase and abrupt change; Table 4, LI, AC). Although the density-dependent estimation model was relatively unbiased in many cases, it performed relatively poorly overall because it failed to converge for 25%–85% (depending on the combination of underlying catchability model, survey SD, and level of fishing mortality) of the simulated data sets that did not contain density-dependent catchability; the other estimation models usually failed to converge less than 1% of the time. This lack of convergence likely occurred because the two parameters describing density-dependent catchability were confounded with one another (i.e., many combinations of α and β could produce equally good fits) for many data sets and thus the optimization procedure could not find a unique best solution. Because of the lack of convergence, results of the density-dependent model were omitted from the rest of the results section.

Scenarios without catchability trends

In cases where the data-generating models contained white noise catchability, first-order autoregressive catchability, or density-dependent catchability, all estimation models produced relatively unbiased estimates of biomass in the last year (i.e., MREs near zero; Table 4, WN, AR, DD), with the most biased estimation model in these scenarios having an MRE of only –8.7% (random walk estimation model fitting density-dependent generation model with low mortality and no survey). There were larger differences in precision among the estimators and this was reflected in MARE and the tightness of the distributions of relative errors (Table 4). For cases where the estimation model was the same as the data-generating model (white noise), the estimation model that matched the data-generating model performed best (i.e., had the lowest MARE and tighter distributions). In the case of the AR(1) data-generating model, the random walk model performed best in most cases. Differences in MARE among estimation models that modeled catchability as white noise or a random walk or ignored fishery effort were usually less than 5% for cases with good surveys (Fig. 5). However, MAREs of random walk and white noise estimation models were 10%–30% lower than estimation models that ignored fishery effort in cases with a poor survey. Differences in relative performance of estimation models were largely accounted for by differing performance of random walk and white noise catchability models because the performance of the estimation model that ignored fishery effort data was relatively constant for a given level of fishing mortality and survey quality.

Scenarios with catchability trends

The white noise and random walk estimation models were biased in cases where catchability increased linearly or increased abruptly, but the amount of bias depended on survey quality, fishing mortality rate, and data-generating model. The MREs of biomass in the last year for estimation models with white noise and random walk catchability were above zero in all cases, indicating a positive bias (Table 4, LI, AC). The positive bias seen in our simulations undoubtedly reflects the direction of change in catchability built into our simulations, where the estimation models did not fully account for the increase in fishery catchability. Neither the white noise nor the random walk estimation models performed well in cases with no survey, trending catchability, and low mortality. The amount of bias was highest in cases where fishery catchability changed abruptly and fishing mortality rates were low and decreased as the level of fishing mortality increased and as survey quality improved.

Although the random walk estimation model was biased, it usually had a lower MARE than the other estimation models, but performance relative to the other estimation models depended on the treatment. In cases with a good survey, the MARE of the estimation model that ignored fishery effort and the MARE of the random walk estimation model were within 5% of one another (Fig. 5). However, in cases with a poor survey, the random walk model usually had MAREs 8%–20% lower than the estimation model that ignored fishery effort. The estimation model that ignored fishery effort data only outperformed the random walk model in the scenario with an abrupt change in catchability and low or medium fishing mortality. The estimation model that ignored fishery effort and the random walk estimation model clearly outperformed the white noise estimation model in these cases and had MAREs 12%–50% lower than the white noise estimation model (Fig. 5).

Discussion

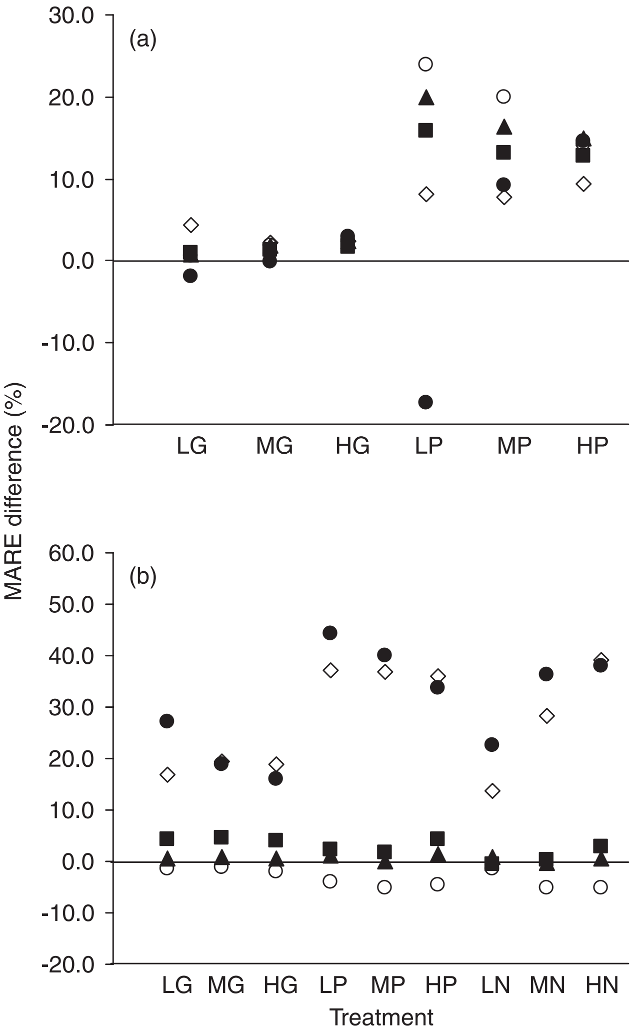
Often stock assessment scientists will not use or will substantially downweight (i.e., specify an arbitrarily large SD)

Table 4. Simulation results for statistical catch-at-age estimation model performance in cases where data-generating models included white noise catchability (WN), first-order autoregressive catchability (AR), density-dependent catchability (DD), linearly increasing catchability (LI), and an abrupt increase in catchability (AC).

			Estimation model					
			MRE (%)			MARE (%)		
q Model	Mortality	Survey	WN	RW	FF	WN	RW	FF
WN	Low	Good	–0.7	1.2	1.9	**19.3**	20.8	21.4
WN	Low	Poor	–0.2	–0.6	1.9	**21.4**	25.3	49.2
WN	Low	None	–0.4	–2.5		**23.7**	25.1	
WN	Medium	Good	0.2	1.4	2.7	**13.5**	14.7	16.6
WN	Medium	Poor	1.3	0.2	3.0	**15.9**	21.1	41.1
WN	Medium	None	1.3	–0.9		**17.9**	23.0	
WN	High	Good	1.1	0.7	3.2	**10.4**	12.4	14.9
WN	High	Poor	1.5	0.1	5.1	**12.7**	17.2	31.4
WN	High	None	1.8	0.8		**13.8**	19.1	
AR	Low	Good	–0.8	–0.2	0.4	26.2	**22.0**	23.0
AR	Low	Poor	0.0	–2.6	1.4	35.3	**33.2**	49.0
AR	Low	None	–0.2	–3.8		**39.6**	40.1	
AR	Medium	Good	0.1	0.5	3.5	20.6	**15.9**	17.2
AR	Medium	Poor	0.1	–4.9	2.1	29.5	**29.9**	41.1
AR	Medium	None	1.5	–4.3		32.9	**32.6**	
AR	High	Good	0.6	0.9	2.7	17.1	**13.2**	14.9
AR	High	Poor	0.6	–2.0	2.4	23.7	**19.6**	32.3
AR	High	None	1.0	–4.2		26.5	**23.7**	
DD	Low	Good	–2.2	1.6	1.7	23.1	**22.4**	23.2
DD	Low	Poor	–4.7	–2.3	4.3	32.4	**31.2**	51.1
DD	Low	None	–5.8	–8.7		36.8	**36.0**	
DD	Medium	Good	–3.0	1.3	2.8	17.0	**16.1**	17.9
DD	Medium	Poor	–3.2	–4.0	2.9	**24.6**	24.7	41.1
DD	Medium	None	–4.2	–7.1		**27.4**	27.7	
DD	High	Good	–0.3	1.1	3.2	13.0	**12.5**	15.0
DD	High	Poor	–0.1	–1.4	4.9	19.4	**17.9**	32.9
DD	High	None	–1.0	–3.4		21.5	**20.8**	
LI	Low	Good	35.4	5.1	2.3	35.4	**18.6**	23.0
LI	Low	Poor	78.4	41.2	4.1	78.4	**41.3**	49.3
LI	Low	None	75.6	62.0		75.6	**62.0**	
LI	Medium	Good	33.9	7.3	2.5	33.9	**14.4**	16.6
LI	Medium	Poor	67.5	30.1	3.7	67.5	**30.7**	38.5
LI	Medium	None	71.9	43.6		71.9	**43.6**	
LI	High	Good	30.8	7.2	3.1	30.8	**11.9**	14.3
LI	High	Poor	55.1	17.5	4.2	55.1	**19.0**	28.4
LI	High	None	61.2	21.8		61.2	**22.1**	
AC	Low	Good	52.5	16.6	2.1	52.5	25.4	**23.5**
AC	Low	Poor	110.3	65.9	6.0	110.3	65.9	**48.6**
AC	Low	None	149.7	127.0		149.7	**127.0**	
AC	Medium	Good	35.4	8.6	3.4	35.4	16.5	**16.4**
AC	Medium	Poor	69.3	27.9	2.5	69.3	**29.4**	38.6
AC	Medium	None	80.5	44.0		80.5	**44.0**	
AC	High	Good	27.5	3.6	2.6	27.5	**11.4**	14.4
AC	High	Poor	48.4	8.5	2.7	48.4	**14.6**	29.1
AC	High	None	53.7	10.7		53.7	**15.7**	

Note: Shown are median relative error (MRE) and median of the absolute values of relative error (MARE) for estimated biomass in the last year (year 15) from four statistical catch-at-age estimation models: white noise (WN), random walk (RW), density-dependent (power relationship; DD), and freely estimated F at maximum selectivity (i.e., not fitted to fishery effort data; FF). Data-generating models included three levels of fishing mortality (high ($F = 2M$), medium ($F = M$), and low ($F = 0.5M$)), and three levels of survey precision (good (SD = 25%), poor (SD = 100%), and no survey). Bias is indicated by MREs different from zero, and lower MAREs indicate more accurate model predictions. Estimation models with the lowest MARE for each treatment are indicated in bold.

Fig. 5. Relative performance of (*a*) the estimation model that ignores fishery effort and (*b*) the white noise estimation model versus the random walk estimation model measured by the difference of the median of the absolute value of the relative errors (MARE). Points above zero indicate that the estimation model that ignored fishery effort data or the white noise estimation model had a larger MARE than the random walk estimation model and vice versa. Data-generating models are indicated by the symbol shape: white noise, ○; autoregressive, ■; density-dependent, ▲; linear increase, ◇; abrupt change, ●. Two letters identify each treatment: the first letter indicates the level of fishing mortality (L, low; M, medium; H, high) and the second letter indicates the level of survey quality (G, good; P, poor; N, none).



fishery effort or CPUE data in an SCA if a fishery-independent index of abundance is available for a given stock. Indeed, based on the results of their simulations, the NRC (1998) recommended that fishery-dependent indices of abundance should be ignored if an independent index of abundance is available, although many assessments will use both fishery-dependent and -independent data if they are available (Millar and Methot 2002; Francis et al. 2003). However, our results argue against automatically ruling out the use of fishery-dependent indices of abundance when a survey is present. In cases where the survey SD is large, we believe that use of fishery-dependent indices is justified if they are believed to contain information on stock size. Of course, fishery effort should be adjusted for known changes in fishing efficiency, and the estimation model should allow for flexible changes in catchability over time, as was the case for our random walk estimator. The reliability of fishery effort data may be suspect in some fisheries, and in these cases, it may make sense to ignore fishery effort. Using methods that do not allow for trends in catchability can lead to severely biased SCA estimates, and modeling fishery catchability as white noise (which is often done) may not provide the necessary flexibility for models to accurately depict system dynamics. Also, there may be a tendency to overstate the precision of survey data and understate the precision of fishery data in SCAs (Francis et al. 2003), which may degrade accuracy of SCA estimates.

Our recommendations are contrary to those of NRC (1998) because we evaluated a wider range of structural models for time-varying fishery catchability within SCAs, but our results yield similar insights for the cases they explored. In the NRC (1998) study, fishery catchability increased over time combined with density dependence; their survey had a SD of 30% (near the level of our "good" survey). Also, the NRC (1998) study mainly included SCA estimation models that contained white noise models for catchability or ignored fishery effort data (see Restrepo (1998) for details of models used in the NRC 1998 study). The exception was one estimation model where fishery catchability was modeled as a mixture of random walk and white noise processes. However, the SD of the white noise term was large relative to the SD of the random walk term (Ianelli and Fournier 1998), which likely caused the model to perform similarly to a white noise model. Similar to the results of NRC (1998), we also found that SCA models that ignored fishery effort data outperformed SCA models that modeled fishery catchability as white noise in cases with trending catchability.

Independent survey indices of abundance or relative abundance are extremely important for obtaining accurate SCA estimates, especially in situations with low fishing mortality. Our results agree with the NRC (1998) recommendation to use survey data if they are available. In our study, estimation models that utilized fishery effort data and survey data (even with a SD of 100%) outperformed models that used only fishery effort data, especially in cases where catchability trended over time and fishing mortality was not high.

It is important to standardize effort series to remove catchability trends to as large an extent as possible. Our experiments showed that SCA estimates were most biased when trends or abrupt changes in fishery catchability occurred and that all our estimation models performed reasonably well in cases where catchability did not trend over time. However, trending fishery catchability is probably a common phenomenon. Many mechanisms could lead to trends in fishery catchability, e.g., increasing power of the fishery, increasing aggregation of fish stocks and fishers, or trending recruitment dynamics and density-dependent catchability. Salthaug and Aanes (2003) presented a method to correct CPUE for the spatial distribution of fishing effort, which has been shown to affect catchability (Winters and Wheeler 1985; Rose and Kulka 1999). Also, improvements in vessels and other fisherman behaviors can be accounted for either by preprocessing (e.g., analyzing CPUE data to estimate mean CPUE by accounting for vessel characteristics and spatial and temporal patterns of fishing) fishery data or by integrating the standardization process into the stock assessment model (e.g., Maunder 2001; Maunder and Starr 2003; Maunder and Punt 2004).

Our results represent the relative performance of the estimation models rather than the absolute performance because, except for the catchability aspect, the structure and assumed values of some parameters of the estimation models were correct (i.e., the same as the data-generating model). Fishery selectivity may vary over time, which can cause biased estimates from SCA models if it is not accounted for (Radomski et al. 2005). Further, our data-generating models contained a survey with an asymptotic selectivity pattern, which allows SCA models to produce more accurate estimates than other survey selectivity patterns (Bence et al. 1993). Likewise, our models did not contain trends in survey catchability over time or correlation with changes in fishery catchability, which could cause models that used survey indices of abundance to generate less accurate estimates.

The results of our analysis are contingent on correctly specifying M. In reality, the data analyst will probably not know the true M, and M may vary among years and ages. Attempting to estimate M within stock assessment models is becoming more common (Fu and Quinn 2000); however, M is often confounded with catchability and other parameters in SCA models (Schnute and Richards 1995; Fu and Quinn 2000). In some cases, M may be estimable, but this procedure usually requires additional auxiliary data that we did not include in our analyses, such as estimates of population size or other parameters from a mark–recapture study (Quinn and Deriso 1999). For example, Fu and Quinn (2000) showed that M was estimable in a length-structured model, especially when survey catchability was specified at the correct value. For our simulations, M was generally not estimable (results not shown). It seems that often reliable auxiliary information on either M or catchability is essential for fitting an age-structured assessment.

The density-dependent estimator used in this study was not one that would probably be used in real assessments, and explicitly modeling density-dependent catchability seems to be rarely used in stock assessments. The formulation that has been suggested for including density-dependent catchability in SCAs includes white noise process error about the density-dependent relationship (Fournier 1983). We did not include the Fournier (1983) estimator in our simulation because we wanted a clear distinction between the white noise and density-dependent estimators. Because of problems with convergence of our density-dependent estimator in most cases without density dependence, we do not believe that the density-dependent estimation model is a viable omnibus estimator, although it may be worth considering when there are a priori grounds for suspecting density dependence.

The relative performance of the estimators we tested depends on the amount of variation in catchability used in our simulations (SD of 0.2). We used this amount of variability because it is similar to estimates of the SD of catchability for commercial fisheries in New Zealand (Francis et al. 2003), which to our knowledge is the only study that has attempted to characterize variability in catchability across many fisheries within a region. We also conducted a limited set of simulations (results not shown) where the SD for catchability was set to 0.4. Results of these simulations were similar to those for the 0.2 scenarios. However, the relative performance of the estimators would likely change with increasing variation in catchability above these levels. Some useful generalizations can be drawn from the similarity of the different estimators. The estimator that ignored effort data and the white noise and random walk estimators are all special cases of an AR(1) process. In the case of white noise and random walk estimators, the correlation coefficient is zero and unity, respectively; for the estimator that ignored fishery effort data, the SD is infinity. As the SD of catchability increases, the white noise and random walk estimators will perform more like the estimator that ignores effort data if the correct standard deviation is used in the estimation model. We attempted to implement an estimation model where catchability followed an AR(1) process and the correlation and SD parameters were estimated, but in most cases this model failed to converge (results not shown).

We believe that using a random walk to model fishery catchability in an SCA is currently the best available approach when both fishery-dependent and -independent indices of abundance are available. This approach allows the model to accommodate and adjust for trends in catchability over time, which can cause biased estimates of stock size and fishing mortality rates if they are unaccounted for. An alternative approach to simply using a random walk may be to use a composite estimator (i.e., fitting several alternative models and choosing the best model structure based on specific rules). Using a composite estimator may provide improved performance over simply using a random walk. However, increased performance from a composite estimator is likely to be small because the random walk model performed nearly as well as the as the best model in the cases we examined. A variety of alternative approaches exist for selecting among competing models or averaging over such competitors (Burnham and Anderson 2002; McAllister and Kirchner 2002; Spiegelhalter et al. 2002). The extent to which such methods can lead to improved stock assessments is an area warranting future research.

Acknowledgments

The authors thank two anonymous reviewers and Dan Hayes, Mike Jones, and Rob Tempelman for providing helpful comments, which improved this manuscript. This research was supported in part by the Michigan Sea Grant College Program, the Great Lakes Fish and Wildlife Restoration Act administered by the US Fish and Wildlife Service, the Great Lakes Fishery Commission, the International Association for Great Lakes Research, and Michigan State University. This is publication number 2006-01 of the Quantitative Fisheries Center, Michigan State University.

References

Bence, J.R., Gordoa, A., and Hightower, J.E. 1993. Influence of age-selective surveys on the reliability of stock synthesis assessments. Can. J. Fish. Aquat. Sci. **50**: 827–840.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York.

Deriso, R.B., Quinn, T.J., II, and Neal, P.R. 1985. Catch–age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. **42**: 815–824.

Fournier, D.A. 1983. An analysis of the Hecate Strait Pacific cod fishery using an age-structured model incorporating density dependent effects. Can. J. Fish. Aquat. Sci. **40**: 1233–1243.

Fournier, D., and Archibald, C.P. 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. **39**: 1195–1207.

Francis, R.I.C.C., Hurst, R.J., and Renwick J.A. 2003. Quantifying annual variation in catchability for commercial and research fishing. Fish. Bull. **101**: 293–304.

Fu, C., and Quinn, T.J., II. 2000. Estimability of natural mortality and other population parameters in a length-based model: *Pandalus borealis* in Kachemak Bay, Alaska. Can. J. Fish. Aquat. Sci. **57**: 2420–2432.

Harley, S.J., Myers, R.A., and Dunn, A. 2001. Is catch-per-unit-effort proportional to abundance? Can. J. Fish. Aquat. Sci. **58**: 1760–1772.

Hilborn, R., and Walters, C.J. 1992. Quantitative fisheries stock assessment. Chapman and Hall, New York.

Ianelli, J.N., and Fournier, D.A. 1998. Alternative age-structured analyses of the NRC simulated stock assessment data. *In* Analyses of simulated data sets in support of the NRC study on stock assessment methods. *Edited by* V.R. Restrepo. US Department of Commerce, Washington, D.C. NOAA Tech. Memo. NMFS-F/SPO-30. pp. 81–96.

Maunder, M.N. 2001. A general framework for integrating the standardization of catch per unit effort into stock assessment models. Can. J. Fish. Aquat. Sci. **58**: 795–803.

Maunder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. **70**: 141–159.

Maunder, M.N., and Starr, P.J. 2003. Fitting fisheries models to standardized CPUE abundance indices. Fish. Res. **63**: 43–50.

McAllister, M., and Kirchner, C. 2002. Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. Bull. Mar. Sci. **70**: 499–540.

Megrey, B.A. 1989. Review and comparison of age-structured stock assessment models from theoretical and applied points of view. *In* Mathematical analysis of fish stock dynamics. *Edited by* E.F. Edwards and B.A. Megrey. Am. Fish. Soc. Symp. **6**: 8–48.

Methot, R.D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. *In* Proceedings of the Symposium on Application of Stock Assessment Techniques to Gadids. *Edited by* L. Low. International North Pacific Fisheries Commission Bulletin 50. pp. 259–277.

Millar, R.B, and Methot, R.D. 2002. Age-structured meta-analysis of U.S. West Coast rockfish (Scorpaenidae) populations and hierarchical modeling of trawl survey catchabilities. Can. J. Fish. Aquat. Sci. **59**: 383–392.

National Research Council (NRC). 1998. Improving fish stock assessments. National Academy Press, Washington, D.C.

Otter Research Limited. 2000. An introduction to AD Model Builder version 4 for use in nonlinear modeling and statistics. Otter Research Ltd., Sidney, B.C.

Paloheimo, J.E., and Dickie, L.M. 1964. Abundance and fishing success. Rapp. P-V Reun. Cons. Int. Explor. Mer, **155**: 152–163.

Peterman, R.M., and Steer, G.J. 1981. Relation between sport-fishing catchability coefficients and salmon abundance. Trans. Am. Fish. Soc. **110**: 585–593.

Punt, A.E., Smith, A.D.M., and Cui, G. 2002. Evaluation of management tools for Australia's south east fishery. 2. How well can management quantities be estimated? Mar. Freshw. Res. **53**: 631–644.

Quinn, T.J., II, and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York.

Radomski, P.A., Bence, J.R., and Quinn, T.J., II. 2005. Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational kill dominated fishery. Can. J. Fish. Aquat. Sci. **62**: 436–452.

Restrepo, V.R. 1998. Analyses of simulated data sets in support of the NRC study on stock assessment methods. US Department of Commerce, Washington, D.C. NOAA Tech. Memo. NMFS-F/SPO-30.

Rose, G.A., and Kulka, D.W. 1999. Hyperaggregation of fish and fisheries: how catch-per-unit-effort increased as the northern cod (*Gadus horhua*) declined. Can. J. Fish. Aquat. Sci. **56**(Suppl. 1): 118–127.

Salthaug, A., and Aanes, S. 2003. Catchability and the spatial distribution of fishing vessels. Can. J. Fish. Aquat. Sci. **60**: 259–268.

Schnute, J.T., and Richards, L.J. 1995. Influence of error on population estimates from catch-age models. Can. J. Fish. Aquat. Sci. **52**: 2063–2077.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B, **64**: 583–639.

Winters, G.H., and Wheeler, J.P. 1985. Interactions between stock area, stock abundance, and catchability coefficient. Can. J. Fish. Aquat. Sci. **42**: 989–998.

Ziegler, P.E., Frusher, S.D., and Johnson, C.R. 2003. Space–time variation in catchability of southern rock lobster *Janus edwardsii* in Tasmania explained by environmental, physiological and density-dependent processes. Fish. Res. **61**: 107–123.