Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational, kill-dominated fishery

Paul Radomski, James R. Bence, and Terrance J. Quinn II

Abstract: We used simulations to compare the distributions of estimation errors for virtual population analysis using forward calculation (FVPA) and three variants of statistical kill-at-age analysis (KAA). The KAA variants assumed constant, time-blocked, and nonadditive selectivity. Simulations were based on a recreational walleye (*Sander vitreus*) fishery in Lake Mille Lacs, Minnesota. The focus of our experiments was on how model mis-specification (incorrect assumptions about selectivity for KAA or that kill had no error for FVPA) interacted with the magnitude of measurement errors and fishing mortality. We found that KAA models outperformed FVPA when they assumed the correct selectivity pattern, even when kill was measured without error. Of particular concern was a strong tendency by FVPA to overestimate stock size when kill was measured with substantial error. When KAA was based on an incorrect assumption regarding fishery selectivity and kill was measured with little error, wide distributions of errors and substantial biases sometimes resulted. KAA models that allowed fishery selectivity to change over time performed about as well as a constant-selectivity KAA model when selectivity was constant, and they performed much better when selectivity changed over time. Careful consideration of alternative fishery selectivity models should be a fundamental part of any age-structured assessment.

Résumé : Des simulations nous ont permis de comparer les distributions des erreurs d'estimation dans une analyse virtuelle de population utilisant un calcul prospectif (FVPA) et trois variantes d'une analyse statistique KAA (kill-at-age, mortalité à un âge donné). Les variantes de l'analyse KAA présupposent une sélectivité constante, déterminée en fonction du temps et non additive. Nos simulations se basent sur la pêche sportive de dorés (*Sander vitreus*) dans le lac Mille-Lacs, Wisconsin. Nos expériences visent à étudier comment les paramètre erronés des modèles (présuppositions fausses de la sélectivité pour KAA et présupposition d'absence d'erreur de mortalité dans FVPA) interagissent avec l'importance des erreurs de mesure et de la mortalité due à la pêche. Les modèles KAA fonctionnent mieux que les modèles FVPA lorsqu'on présuppose les bons patterns de sélectivité, même quand la mortalité est mesurée sans erreur. Il est particulièrement inquiétant que les modèles FVPA aient une forte tendance à surestimer la taille du stock lorsque la mortalité est estimée avec une erreur importante. Lorsque l'analyse KAA est basée sur des présuppositions fausses concernant la sélectivité de la pêche et que la mortalité est mesurée avec peu d'erreur, il peut en résulter quelquefois de distributions étendues des erreurs et des tendances erronées importantes. Les modèles KAA qui permettent à la sélectivité de la pêche de changer dans le temps fonctionnent presque aussi bien que le modèle KAA à sélectivité constante lorsque la sélectivité est constante et fonctionnent beaucoup mieux lorsque la sélectivité varie dans le temps. Toute évaluation qui tient compte de la structure en âge devrait considérer de façon attentive les différents modèles de rechange de sélectivité de la pêche et en faire une partie essentielle de l'analyse.

[Traduit par la Rédaction]

Introduction

A widely accepted standard for stock assessment is age-structured modeling to integrate information about a fish population (Megrey 1989; Hilborn and Walters 1992; Quinn and Deriso 1999). Kill limits and other biological reference points derived from such assessments form the scientific basis for management of many of the world's most important fisheries. One common modeling family is virtual population analysis (VPA) (Jones 1963; Gulland 1965; Murphy 1965) and its variants: cohort analysis (Pope 1972) and ADAPT (adaptive framework approach to VPA, Gavaris 1988). A second family is statistical age-structured assessment models (Fournier and Archibald 1982; Deriso et al. 1985; Methot 1989). We denote this family as KAA (kill-at-age analysis) to emphasize that it is the number of fish killed, not the number

Received 30 December 2003. Accepted 13 September 2004. Published on the NRC Research Press Web site at http://cjfas.nrc.ca on 22 March 2005.
J17893

P. Radomski.[1] Minnesota Department of Natural Resources, 1601 Minnesota Drive, Brainerd, MN 56401, USA.
J.R. Bence. Michigan State University, 11a Natural Resource Building, East Lansing, MI 48824-1222, USA.
T.J. Quinn II. Juneau Center, School of Fisheries and Ocean Sciences, 11120 Glacier Highway, Juneau, AK 99801-8677, USA.

[1]Corresponding author (e-mail: paul.radomski@dnr.state.mn.us).

 doi: 10.1139/F04-228 © 2005 NRC Canada

caught, that is important for assessment; many fish caught in recreational fisheries may be released unharmed. VPA methods have been popular in eastern North America and Europe, whereas KAA methods have been popular in western North America, South Africa, Australia, and New Zealand; in many other areas, no clear favorite has emerged.

The fundamental difference between the two is that the VPA family assumes that kill-at-age is measured without error. Of course practitioners know that kill is measured with some error, and their use of VPA reflects a belief or hope that results will be robust given the level of measurement error. VPA can use backward or forward recursive calculations to obtain abundance at age. Backward VPA approaches allow for constraints on terminal abundances or fishing mortalities. Such constraints (linkage assumptions) often lack statistical backing or are ad hoc in nature (Quinn and Deriso 1999), but their imposition can increase stability in the estimation routines by limiting the number of parameters that are directly estimated. Both VPA approaches allow year- and age-specific fishing mortality rates to be calculated from the resulting estimates of abundance at age. Earlier VPA implementations amounted to a calculation procedure, not an estimation procedure. Recent VPA implementations statistically incorporate auxiliary data to estimate population size at age, yet they still assume kill is measured without error. Given A ages and Y years of data, there are AY catch-age observations and $A + Y - 1$ year classes present in the database. If, as assumed by VPA, kill is measured without error and natural mortality is known, then estimation of $A + Y - 1$ year-class parameters in a stock assessment model (see Appendix A) is sufficient to determine abundances at all ages. Some implementations of VPA further reduce the number of formal parameters by imposing assumed relationships for terminal fishing mortality rates.

In contrast, if catch is measured with error and natural mortality is known, as assumed by KAA, then estimation of $A + Y - 1$ year-class parameters and AY fishing mortality parameters is necessary to determine abundances at all ages. The total of $AY + A + Y - 1$ parameters is greater than the number of observations, requiring assumptions to reduce the number of parameters so that the model is not overdetermined. KAA reduces the number of fishing mortality parameters that are estimated by assuming separability of fishing mortality into age (selectivity and (or) vulnerability) and year (fishing intensity) factors for a block of years or by assuming some other statistical model for fishing mortality with reduced parameters. Thus, KAA presents concerns about how well selectivity has been modeled and whether results will be robust when the modeled fishing process diverges from the actual one. Even with these selectivity assumptions, KAA requires more formal parameters to determine abundance at age than VPA.

Some researchers believe that VPA is more parsimonious than KAA because fewer formal parameters need to be estimated. Others believe the opposite because the relationships among fishing mortality rates in VPA (without linkage assumptions) are constrained only by the kill data. In this view, each of the unconstrained AY fishing mortality rates increases the effective number of parameters that needs to be estimated. If the kill data can be matched in more than one way, the lack of further constraints on fishing mortality could be a disadvantage for VPA. Furthermore, in the presence of measurement error, VPA can be viewed as an exact solution to a set of catch equations, in which fishing mortality is estimated for each age and year. From this viewpoint, VPA estimates many more parameters than KAA. In contrast, the statistical elegance of KAA may not be reflected in the performance of the method compared with VPA if the specified model for fishing mortality is not a reasonable approximation of the actual processes that are operating.

Although a variety of limitations and assumptions for VPA and KAA are well understood, it is rare to see comparisons of both approaches for a managed fishery. International Council for Exploration of the Sea (1993) and the National Research Council (NRC) (1998) evaluated the performance of several modeling approaches by simulating fish populations and fisheries such that common modeling assumptions would be violated (e.g., ageing error, fisheries catchability changes, kill under-reporting, and age selectivity changes). These studies produced relatively few simulated data sets, with detailed assessments developed by analysts familiar with the different methods. While these studies provided a realistic assessment of how well different analytical approaches and their practitioners could perform on specific simulated data sets, they did not provide general guidance on the applicability of KAA and VPA under a range of different conditions. Punt et al. (2002) compared KAA, several variants of VPA, and several other methods in an extensive simulation study. The model they used to generate simulated data was moderately complex, being spatially, size-, and age-structured, allowed for technological interactions, and was parameterized to reflect assumptions specific to four fisheries. Their generating models did not match the assumptions made by any of the estimating models, as they assumed that total yield was measured without error (although total weight of fish killed had some error because discard was assumed to be measured with error), and their comparisons were based on short (8-year) simulated time series. They found that the KAA estimator generally outperformed VPA, although not always.

The objective of our study was to probe the benefits and shortcomings of several variants of KAA and one implementation of forward VPA (FVPA) in the presence of varying levels of measurement error and fishing mortality. The usefulness and reliability of each approach was tested using simulation studies, in which the assessments were done in an automated and standard way on simulated data sets. Our simulations were based on the recreational-dominated walleye (*Sander vitreus*) fishery of Lake Mille Lacs, Minnesota, and should provide useful guidance for both recreational and commercial fisheries.

We were particularly interested in comparing the relative performance of a VPA method and KAA, when in fact kill was measured with appreciable error. Our intent was to document the trade-offs between the assumption of known catch-at-age in VPA and the selectivity assumptions in KAA. Consequently we sought to minimize other differences between VPA and the KAA methods. We therefore chose a FVPA implementation, which uses a statistical approach and which does not constrain terminal fishing mortality rates through linkage assumptions. We recognize that linkage assumptions can improve estimation by VPA, depending in part on how well the assumed constraints reflect that actual fishing pro-

© 2005 NRC Canada

cess. However, use of such procedures in our simulations would have confounded our intended comparison by including additional assumptions regarding selectivity in the VPA method. Our use of a statistical FVPA approach also allowed us to handle survey data and estimate survey catchability in the same way as was done by the KAA methods.

Given that KAA replaces the assumption of known kill with assumptions regarding selectivity, simulated data were generated making different selectivity assumptions and our variants of KAA modeled selectivity in different ways. Thus, we also evaluated the performances of KAA assessment models when their selectivity models were misspecified. Our systematic comparison of methods in the face of different levels of measurement error combined with different degrees of model misspecification is a unique contribution of this study.

Methods

We simulated population dynamics, expected kill-at-age, and expected gillnet survey indices of abundance based on our understanding of the recreational-dominated walleye fishery of Lake Mille Lacs. We added measurement error to the expected kill-at-age and age-specific indices of abundance to generate simulated data sets. We fitted a suite of age-structured estimation models to each simulated data set. In generating simulated data, we varied fishing mortality rates, measurement error associated with the kill-at-age data, measurement error associated with the survey, and assumptions about selectivity (and the degree of separability) of the fishery to evaluate how these factors influenced the relative performance of the different estimating models. Equations defining the simulation model are presented in Appendix A, and symbols used in the simulation model and in estimation models that are not defined in the text are given there. Equations in Appendix A are denoted as A*x*. Where equivalent quantities are used in the estimation models, as in the simulation model, they are not redefined, and estimated quantities are distinguished from simulated ones by placing a carat over the symbol.

Simulation model

Our generating model assumes a single fishery and single survey with characteristics like those of the recreational fishery and gillnet survey, respectively, on Lake Mille Lacs. Lake Mille Lacs is a 54 000-ha glacial lake in central Minnesota, and among large lakes in the state, it is the most productive for walleye (3.6 $kg \cdot ha^{-1} \cdot year^{-1}$). Although annual fish quotas have been set to allocate the total allowable kill between state-licensed sport anglers and Ojibwe Indian netters and spearers since 1997, the fishery kill has been dominated by the sport fishery over the time period used in age-structured assessments (Radomski 2003). The sport fishery has been monitored by annual creel surveys since 1983, and estimates of fishery kill-at-age have substantial measurement error (kill coefficient of variation (CV) greater than 10%). Fish populations are also annually monitored with gillnet, electrofishing, and trawl assessment surveys. The gillnet survey provides the longest time series and is the most extensive of these fishery-independent data sources, and relative abundance estimates have substantial measurement error (survey CV greater than 20%).

The population model covers 21 years, from year 1 to year 21, and produces simulated, "actual" abundances at the start of each year for ages 1 through 9 and a 10+ group. Abundance at age-1 (recruitment) in years 2 through 21 and abundance at all ages in the first year took the same values in each simulation, based on estimates of these quantities from age-structured assessment models of the Lake Mille Lacs walleye fishery. Subsequent abundances for ages (a) 2 through 9 and years (y) 2 through 21 follow a standard deterministic mortality model (Appendix A). In our experiments, we considered three scenarios for fishing mortality on fully selected ages (F_y). The status quo pattern was based on fishing mortality rates estimated for the Lake Mille Lacs walleye fishery in past assessments. We also considered two other fishing mortality scenarios, which we think come close to spanning the range that managed walleye fisheries are likely to experience. The low fishing mortality scenario had fully selected fishing mortality set to half of those in the status quo option. The high fishing mortality scenario had status quo rates through year 13 and a linearly increasing fishing mortality on fully selected ages over the subsequent years (Table 1).

We generated data from a model in which fishing mortality was fully separable into year and age effects (i.e., constant selectivity $S_{a,y} = S_a$) and from three models that deviated from full separability (Appendix A and Fig. 1). The first alternative, time-blocked selectivity, was where the selectivity pattern changed every 5 years, but was constant within 5-year time blocks. The next alternative was a nonadditive model with selectivity varying as a function of fishing mortality (Myers and Quinn 2002). Finally, random variability was added each year to a base selectivity pattern. Annual variability was chosen to be larger for younger ages and of a magnitude to create real contrast in annual selectivity (as shown in Fig. 1). The constant selectivity pattern was dome-shaped and based on stock assessment results for the Mille Lacs Lake walleye fishery (Fig. 1). This same selectivity pattern also served as the base pattern for the nonadditive and randomly varying selectivity models and was the selectivity pattern for the first time block in the time-blocked selectivity model. Selectivity for the remaining three blocks was selected to provide contrast in the simulation results, with increased selectivity at most older ages for each succeeding time block.

Observed data for kill-at-age and survey indices-at-age were simulated for years 1 through 20 and years 1 through 21, respectively. Based on modeled abundance and mortality rates, actual numbers killed at age were calculated from the Baranov catch equation (eq. A8). Observed kill-at-age with a median equal to the actual kill was generated by multiplying this kill by a lognormal measurement error term (eq. A9). (Preliminary simulations, using a bias correction factor, set the mean of observed kill equal to actual kill and produced qualitatively similar results.) Gillnet survey indices of abundance were simulated similarly, assuming that the indices were proportional to actual abundance at age, up to a multiplicative lognormal measurement error (eq. A10). The proportionality constants, or catchabilities, were constant over time, but age-specific (Table 1). The kill-at-age and survey

© 2005 NRC Canada

Table 1. Parameter values used in the simulator.

Year-specific					Age-specific			
Year	Recruitment of age 1	Status quo scenario	Fishing mortality scenario		Age	Population at year 1	Natural mortality rate	Survey catchability
			High	Low				
1	565 088	0.25	0.25	0.12	1	565 088	0.80	2.817×10^7
2	5 914 862	0.13	0.13	0.06	2	1 080 982	0.45	1.735×10^6
3	273 807	0.28	0.28	0.14	3	127 276	0.24	4.374×10^6
4	460 370	0.26	0.26	0.13	4	165 449	0.24	6.751×10^6
5	2 093 396	0.44	0.44	0.22	5	1 119 917	0.24	8.493×10^6
6	8 863 597	0.23	0.23	0.11	6	148 893	0.24	1.222×10^5
7	956 742	0.36	0.36	0.18	7	143 264	0.24	1.277×10^5
8	799 512	0.31	0.31	0.15	8	50 421	0.24	1.513×10^5
9	3 424 054	0.55	0.55	0.28	9	33 191	0.24	1.589×10^5
10	2 402 837	0.41	0.41	0.20	10+	44 814	0.24	1.589×10^5
11	2 062 625	0.13	0.13	0.06				
12	1 596 742	0.24	0.24	0.12				
13	1 284 108	0.29	0.29	0.14				
14	1 402 466	0.17	0.34	0.08				
15	478 640	0.21	0.39	0.10				
16	627 093	0.36	0.44	0.18				
17	1 890 010	0.19	0.49	0.09				
18	1 726 248	0.26	0.54	0.13				
19	5 283 640	0.29	0.59	0.14				
20	313 023	0.38	0.64	0.19				
21	1 474 974							

Note: Year-specific parameters included a set of recruitments that were used for every simulation and trajectories of fully selected fishing mortality. Age-specific parameters included the abundance at age in the first year and natural mortality and survey catchability. Selectivity is shown in Fig. 1.

indices of abundance at age errors were generated independently among ages and years, in accord with the assumptions of widely used least-squares fitting approaches.

Estimating models

We developed a single FVPA model and three statistical KAA models using standard methods (Quinn and Deriso 1999). For all the KAA models, the abundances at age in the first year and abundance at age 1 in subsequent years 2 through 21 were estimated as parameters during fitting of the models. The subsequent dynamics of each cohort are described by eqs. A1 through A3, with model-estimated quantities substituted for the true values. In all cases, M was assumed known and took the correct values (Table 1).

In each model, estimated or predicted kill-at-age was calculated from eq. A8 with known quantities replaced by model estimates and predicted survey indices of abundance by $\hat{I}_{a,y} = \hat{q}_a\hat{N}_{a,y}$ (i.e., eq. A10 with model estimates replacing true values and the error term dropped). Survey catchability (q_a) was estimated for ages 1–9, and catchability of the group 10+ was assumed correctly to be equal to the catchability of age 9.

The KAA models differed in how fishing mortality was modeled. The first KAA estimating model assumed constant selectivity over time (eq. A4). The second estimated selectivity in 5-year blocks (eq. A5). The third KAA model included estimation of the one degree of freedom for the nonadditivity parameter (Myers and Quinn 2002). Thus, each of the estimating models matched one of the models used to generate the simulated data. For the constant selectivity model and the time-blocked selectivity model, selectivity at ages 4 and 5 was set to 1, and F_y was estimated for each year 1 through 20. Selectivity for ages other than 4 and 5 was estimated either for the entire assessment period (constant selectivity model) or for each time block (time-block model). For the nonadditive model, eqs. A3 and A6 were reparameterized as suggested by Myers and Quinn (2002) to describe $\ln(F_{a,y})$ by

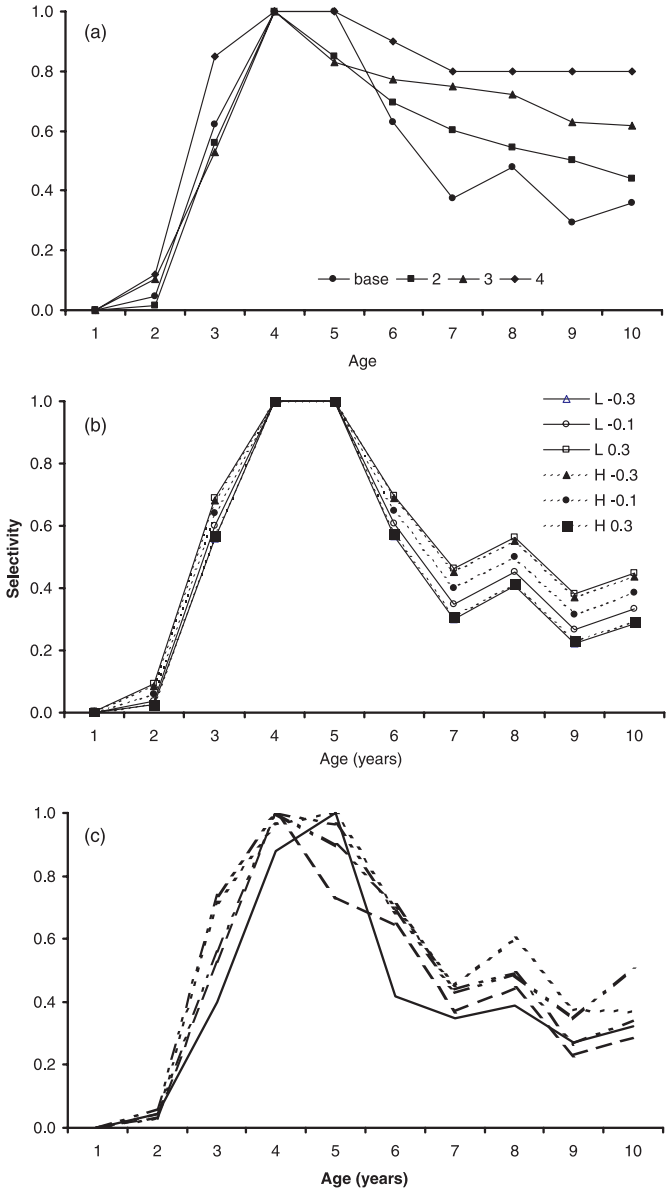
$$(1) \qquad f_{a,y} = \bar{f} + \tilde{f}_y + \tilde{s}_a^{\,1+y\bar{f}_y}$$

so that in addition to γ, $\bar{f}$, $\tilde{f}_y$ (constrained to sum to 0), and $\tilde{s}_a$ (constrained to sum to 0) were estimated instead of F_y and S_a. The KAA constant selectivity model had 67 parameters (9 parameters for population at age >1 for the first year, 21 parameters for population at age 1, 20 fishing mortality parameters (F_y), 8 selectivity parameters (S_a), and 9 parameters for survey catchability q_a); the time-block model had 91 (9 parameters for population at age >1 for the first year, 21 parameters for population at age 1, 20 fishing mortality parameters, 32 selectivity parameters, and 9 parameters for q_a); and the nonadditive model had 71 parameters (9 parameters for population at age >1 for the first year, 21 parameters for population at age 1, 21 fishing mortality parameters, 11 selectivity parameters, and 9 parameters for q_a).

In each case, the parameters of the KAA models were estimated numerically by minimizing the weighted sum of squares:

© 2005 NRC Canada

Fig. 1. Fishery selectivity patterns used in generating simulated data. (*a*) Base selectivity pattern C and selectivity patterns (B) used for time blocks two through four. (*b*) Selectivity pattern N in low (L) and high (H) fishing mortality years (F_y = 1/2 $\bar{f}$ or = 2 $\bar{f}$) for the nonadditive model with Tukey's nonadditivity parameter (γ) of –0.3, –0.1 and 0.3. (*c*) Five outcomes from the random selectivity model R.



$$\text{(2)} \qquad \mathrm{RSS} = \mathrm{RSS}(K) + \lambda \mathrm{RSS}(S)$$

$$\mathrm{RSS}(K) = \sum_{a,y} (\ln K^*_{a,y} - \ln \hat{K}_{a,y})^2$$

$$\text{(3)} \qquad \mathrm{RSS}(S) = \sum_{a,y} (\ln I^*_{a,y} - \ln \hat{I}_{a,y})^2$$

where λ is a weighting term and RSS represents the objective function.

For the FVPA estimating model, the initial abundances of each cohort (abundances-at-age in the first year) and subsequent recruitment at age-1 were estimated as parameters. For computational efficiency, Pope's (1972) approximation was used to project subsequent abundances:

$$\text{(4)} \qquad \hat{N}_{a+1,y+1} = \hat{N}_{a,y} \exp(-M_a) - K^*_{a,y} \exp(-M_a/2)$$

and

$$\text{(5)} \qquad \hat{N}_{10+,y+1} = \hat{N}_{10+,y} \exp(-M_{10+}) - K^*_{10+,y} \exp(-M_{10+}/2) + \hat{N}_{9,y} \exp(-M_9) - K^*_{9,y} \exp(-M_9/2)$$

As with the KAA methods, the correct natural mortality rates were assumed to be known. The FVPA model had 38 parameters directly estimated (9 parameters for population at age >1 for the first year, 20 parameters for population at age 1, and 9 parameters for q_a). The additional 200 fishing mortality rates were calculated by solving eqs. A1 and A2.

The survey indices of abundance were predicted, as for the KAA models, by $\hat{I}_{a,y} = \hat{q}_a \hat{N}_{a,y}$. As with the KAA models, it was assumed that 10+ age group survey catchability was equal to the catchability of the age 9 group. The parameters (abundance-at-age in the first year, recruitment in subsequent years, and the survey catchabilities (excluding the 10+ group)) were estimated by numerically minimizing RSS(*S*), as expressed in eq. 3. The FVPA estimating model was constrained so that population at age and resulting fishing mortality for age values could only be positive.

The estimations were performed using the Solver routine in Microsoft® Excel (Microsoft Corporation, Redmond, Washington). To assure that the software performed properly, we replicated experiment 1 (see below) using AD Model Builder (Otter Research Ltd., Sidney, British Columbia). We also compared parameter estimates from both software products when presented identical test data sets. Both software products produced similar results, providing partial validation of the simulation results.

Experimental design

For each combination of fishing mortality, survey and kill CVs, and selectivity model, we generated 100 simulated data sets and fitted a suite of assessment models to each data set. Preliminary testing suggested that 100 simulations produced estimates for parameters of error distributions sufficiently reliable for our purposes, and we subsequently verified this by bootstrapping (see "Evaluating the performance of the estimating models"). To evaluate retrospective patterns, we refitted the assessment models to each data set four additional times as the last 4 years of data were deleted sequentially starting with the last year. Unless stated otherwise, the value of λ in eq. 2 for the KAA estimators was set to the ratio of variances of kill and survey (log scale) used to generate simulated data. Thus, we usually assumed that the KAA analyst is aware of the relative magnitudes of variances associated with the kill and survey data. For example, if the kill error had a CV of 5% and the survey error had a CV of 25%, then the weighting on survey residual sum of squares was set at about 0.04 (for a lognormal variable, the log-scale standard deviation is approximately equal to the CV).

Five experiments were conducted. Experiment 1 was intended to guide the design of additional simulation experi-

© 2005 NRC Canada

ments and provide results for FVPA and KAA models over a wide range of error variances in the data, when the separable fishing mortality model was correct. The design of this experiment was a factorial cross of kill CV (0%, 5%, 25%, 50%) and survey CV (0%, 5%, 25%, 50%). Simulated data were generated using constant kill selectivity following the base pattern and status quo exploitation (Table 1). The value of λ in eq. 2 for the KAA estimator was set to the ratio of variances of kill and survey (log scale), except for when only kill CV = 0% (λ = 0.001), when only survey CV = 0% (λ = 1000), and when both kill and survey CV = 0% (λ = 1). In this experiment, the KAA method was faced only with measurement error. Here (and in our other simulations), FVPA can be viewed as being misspecified except when the actual kill CV was zero, since the assumption of no error in the kill data is incorrect.

Experiment 2 generated data from the constant selectivity model using the base selectivity pattern (Table 1), and thus the KAA method again only faced measurement error. The experiment was intended to explore how fishing mortality might interact with different levels of measurement error to influence the relative performance of KAA and FVPA assessments, when the KAA model was correct. It followed a factorial design that varied fishing mortality (low, status quo, and high), kill CV (5% and 25%), and survey CV (25% and 50%). As in experiment 1, FVPA and the KAA estimator that assumed a constant kill selectivity were applied to each resulting data set. We used the results of this experiment to define a reasonable but more restricted range of these factors for which to evaluate the alternative estimators when the assumption of separability was violated in later experiments.

Experiment 3, our main experiment, was also factorial, including the type of selectivity pattern (time-blocked, nonadditive (γ = –0.1), and random year-to-year variation), kill CV (5% and 25%), and fishing mortality (status quo and high). Survey CV was 25% in these simulations. All four estimating models were used (FVPA and the three KAA models). Thus, in this experiment the KAA methods were faced with measurement error alone and in combination with model misspecification. Our intent here was to evaluate the relative performance of the different methods when selectivity was no longer constant. We were particularly interested in how various misspecifications of the selectivity pattern would degrade the performance of the different KAA methods and how they would now perform relative to the FVPA.

Experiment 4 was designed to evaluate how more extreme nonadditive fishing mortality than that used in experiment 3 affected assessment performance. The γ parameter, controlling how selectivity varied in response to fishing mortality for the nonadditive model (eq. A6), was set to two values (–0.3, 0.3) and crossed with fishing mortality (status quo and high) in a factorial design. All four estimation models were fitted to each data set, and all simulations used a kill CV of 25% and a survey CV of 25%. An additional level of γ (–0.1) is available from the results of experiment 3 for these same combinations of the other factors. Here, all methods except the nonadditive KAA faced both measurement error and model misspecification.

Experiment 5 evaluated the consequences of setting λ = 1 in the objective function (eq. 2) for the KAA models, a common approach when variances are not known, rather than to the ratio of the log-scale variances. Simulations were done with the altered objective function, using the time-blocked selectivity models crossed with two levels of measurement error ((*i*) kill CV = 5%, survey CV = 25%, λ = 0.04 and (*ii*) kill CV = 25%, survey CV = 50%, λ = 0.25). Because survey CV was greater than kill CV in these simulations, setting λ = 1 has the effect of increasing the emphasis on the survey data relative to the kill data. A status quo fishing mortality rate was used in these simulations, and the three KAA estimating models were fitted to each data set. Results from experiment 3 provide comparative results for the lower level of measurement error, when λ was set to the ratio of the log-scale variances. Here, all but the time-blocked KAA models were misspecified both for the new simulations and the comparative cases from experiment 3. All the KAA models can be viewed as misspecified for the simulations with λ = 1, since incorrect relative variances are being assumed.

Evaluating the performance of the estimating models

The purpose of most stock assessments is to generate estimates of population parameters from the most recent time period. We evaluated the performance of the estimators based primarily on the distribution of relative errors in spawning biomass at the start of year 21 and on the distribution of relative errors in exploitation (ages 2 and older) during year 20. For the calculation of spawning biomass, mean weight-at-age and maturity schedules from Lake Mille Lacs were used. Relative error is defined as the percentage r = (estimated – true)/true × 100 for each quantity for a given simulation. In addition to plotting distributions of these relative errors, we also summarized overall performance by the median of the relative error distribution (MRE) and by the median of the absolute values of the relative errors (MARE). We used MRE to indicate median relative bias, the tendency to consistently underestimate or overestimate. We used MARE to summarize the uncertainty or imprecision in the estimates. We used MRE instead of mean relative error and MARE instead of root mean square error because we encountered occasional extreme outliers (positive errors for spawning biomass) that required use of these robust statistics. For all combinations of generating models, estimating models, and error levels represented in our experiments, we calculated bootstrap standard errors (SE) for MARE and MRE for both spawning biomass at the start of year 21 and exploitation during year 20 (based on 1000 bootstrap samples in each case). We examined plots of these SEs versus the absolute values of MRE or versus MARE to gauge our ability to distinguish differences.

In addition, we examined time trends in relative errors for exploitation and spawning biomass and examined retrospective patterns in these variables (Parma 1993), although we present these time trend results only for selected cases.

Results

Uncertainty in estimates of MRE and MARE was evaluated by calculating SEs based on bootstrap sampling. For both spawning biomass at the start of year 21 and exploita-

© 2005 NRC Canada

Table 2. Comparison of forward virtual population analysis (FVPA) and a statistical kill-at-age model (KAA-C) that assumed constant selectivity over time in experiment 1.

	KAA-C				FVPA			
	Survey CV				Survey CV			
Kill CV	0	5	25	50	0	5	25	50
Spawning biomass MRE								
0	0.00	–0.18	–0.17	–0.21	0.75	0.38	–0.33	1.94
5	0.14	–0.40	–0.09	–1.73	1.40	0.96	0.26	2.02
25	0.36	–0.98	–1.11	0.43	15.76	16.06	15.93	24.18
50	0.61	–0.70	–2.11	0.63	83.34	84.79	102.62	119.53
Spawning biomass MARE								
0	0.00	0.18	0.52	0.96	0.75	1.80	9.67	19.14
5	0.40	1.88	4.55	5.86	1.52	2.35	10.26	19.06
25	2.10	3.43	9.42	14.87	15.76	16.06	17.11	26.95
50	3.93	5.01	10.64	18.82	83.34	84.79	102.62	119.53
Exploitation rate MRE								
0	0.00	0.14	0.14	0.17	–0.75	–0.65	–0.51	–1.39
5	–0.09	0.53	0.20	1.41	–0.96	–1.36	–1.43	–1.37
25	0.14	1.83	2.15	2.77	–7.41	–7.79	–9.89	–13.66
50	0.20	2.16	3.36	2.55	–33.49	–33.89	–38.62	–43.44
Exploitation rate MARE								
0	0.00	0.15	0.41	0.78	0.75	1.49	7.31	15.11
5	0.18	1.71	3.83	5.07	1.79	2.48	8.65	14.86
25	0.78	4.32	8.90	11.57	9.59	11.15	16.04	21.93
50	1.52	6.41	11.33	17.32	33.49	33.89	39.87	44.02

Note: Median relative error (MRE) and median absolute relative error (MARE) of estimates of the most recent spawning biomass and exploitation rate are given as a function of increasing kill and survey error (kill coefficient of variation (CV) = 0%, 5%, 25%,and 50%; survey CV = 0%, 5%, 25%, and 50%) with the status quo exploitation pattern and constant kill selectivity in the simulation model.

tion in year 20, the SEs for MRE were usually well below 5. For exploitation, the maximum estimated SE was 7, which occurred for an absolute MRE of 18. For spawning biomass, SEs exceeded 7.5 in only two cases, taking a value of 16 for an absolute MRE of 20 and an SE of 24 for an absolute MRE of 120. Thus, we focus on differences in MRE of 10 or more or smaller differences representing general patterns in the results. SEs for MARE were more closely tied to the estimated magnitude of MARE, generally being about 10% of the MARE for both spawning biomass and exploitation. Thus, in the case of MARE we focus on differences on the order of 20% of the larger value being compared or general patterns for smaller differences.

Experiment 1 generated data from a purely separable model with constant selectivity, over a range of kill and survey errors, and with the status quo fishing mortality pattern. Not surprisingly, both the FVPA and KAA estimators generally performed better (MRE closer to zero, MARE smaller) when kill and survey CVs were lower (Table 2). The KAA model outperformed FVPA for all combinations of kill and survey CVs (Table 2). The difference in performance was due to both spread of the error distributions and how close the distributions were to being centered on zero. KAA estimates had MRE near zero under all conditions, whereas FVPA had MRE substantially above zero for spawning biomass and substantially below zero for exploitation when kill CV = 25%.

With no error in the kill, both KAA and FVPA were correctly specified, yet under these conditions FVPA still had higher MARE than KAA (Table 2). Not surprisingly, with no measurement error, FVPA had MRE that diverged only slightly from zero (<1 in absolute value) because of the use of Pope's approximation. With no or little (CV = 5%) error in the kill, there is no clear pattern for MRE to increase with increasing survey error. Thus the larger MARE for FVPA than for KAA under these conditions (especially when survey CV = 25%) reflects a broader distribution of errors for FVPA. This can be viewed as a cost incurred by VPA for effectively estimating more quantities than are needed to describe fishing mortality patterns.

Misspecification by FVPA that there is no error in the kill led to even poorer performance relative to KAA when there was substantial error in the kill. At high kill CV (25% and 50%), the FVPA estimator's high positive MRE for spawning biomass is the dominant influence on MARE, although FVPA still had broader distributions of errors than KAA. For FVPA, the movement of the center of the distribution away from zero is the primary cause for the increase in MARE with increasing error in the kill (Table 2); error distributions only became modestly broader (data not shown).

Experiment 2 also generated data from a purely separable model with constant selectivity, but used a range of fishing mortality and two sets of CV values for fishery kill and survey index data (5%, 25% versus 25%, 50%). Results gave

© 2005 NRC Canada

Table 3. Comparison of forward virtual population analysis (FVPA) and a statistical kill-at-age model (KAA-C) that assumed constant selectivity over time in experiment 2.

	KAA-C			FVPA		
	Fishing mortality			Fishing mortality		
Kill CV, survey CV (%)	Low	Status quo	High	Low	Status quo	High
Spawning biomass MRE						
5, 25	0.44	–0.09	–1.45	1.07	0.26	1.08
5, 50	0.68	–1.73	–1.12	–2.21	2.02	2.45
25, 25	0.67	–1.11	–0.76	23.32	15.93	13.15
25, 50	2.50	0.43	–1.99	19.92	24.18	16.00
Spawning biomass MARE						
5, 25	7.69	4.55	4.55	17.33	10.26	9.88
5, 50	9.62	5.86	6.06	34.45	19.06	17.25
25, 25	16.67	9.42	8.85	28.45	17.11	14.10
25, 50	26.82	14.87	12.72	42.25	26.95	20.84
Exploitation rate MRE						
5, 25	–0.07	0.20	0.47	–0.84	–1.43	–2.71
5, 50	2.78	1.41	0.96	8.74	–1.37	–5.84
25, 25	–1.45	2.15	1.64	–18.62	–9.89	–7.07
25, 50	–1.54	2.77	1.92	–17.79	–13.66	–12.69
Exploitation rate MARE						
5, 25	6.64	3.83	2.78	13.00	8.65	7.10
5, 50	9.52	5.07	4.05	29.22	14.86	12.85
25, 25	15.10	8.90	5.88	24.69	16.04	13.83
25, 50	25.31	11.57	8.78	40.28	21.93	18.84

Note: Median relative error (MRE) and median absolute relative error (MARE) of estimates of the most recent spawning biomass and exploitation rate are given as a function of fishing mortality (low, status quo, and high), kill coefficient of variation (CV) (5% and 25%), and survey CV (25% and 50%). Constant kill selectivity was operative.

the same general picture as was seen in experiment 1 (Table 3). Not surprisingly, MARE decreased with decreases in fishery or survey CV or with increases in fishing mortality. The FVPA estimator always performed less well than the KAA estimator for all levels of fishing mortality and error variance. The magnitude of change in MARE in response to these factors was greater for FVPA than KAA. In addition to the larger MARE, the somewhat surprising tendency for FVPA to have positive relative errors (as indicated, for example, by MRE) in spawning biomass and negative relative errors in exploitation was evident when kill CV was 25% for all levels of fishing. As in experiment 1, the improved performance of FVPA when kill CV was lower was largely because the distributions of relative errors tended to be centered closer to zero.

The main experiment, experiment 3, generated data based on different assumptions about how selectivity varied over time. The type of temporal variation in selectivity had a large influence on the relative performance of the different estimating models, and this interacted strongly with the kill CV (Table 4; Figs. 2 and 3). When selectivity varied over time, following either the nonadditive model or 5-year time blocks, KAA estimating models that made the corresponding assumption about selectivity performed relatively well, in the sense that MRE was near zero and MARE was either the lowest or near the lowest among the estimators that were considered. When the kill CV was 5%, misspecification by the KAA estimators in how selectivity changed over time led some estimators to have strong tendencies to either over- or under-estimate spawning biomass and exploitation rate (Figs. 2*a*, 2*c*, 3*a*, 3*c*). This was most notable for the nonadditive and constant selectivity estimators when actual selectivity varied in 5-year time blocks and also for the constant selectivity estimator when actual selectivity was nonadditive ($\gamma = -0.1$). Differences among the KAA estimators were substantially less pronounced when the kill CV was 25% (e.g., Fig. 3*a* versus Fig. 3*b*), suggesting misspecification of selectivity is of less consequence when kill is measured with more error and that higher error rate is acknowledged.

As in experiments 1 and 2, the FVPA estimator generally produced a broader distribution of relative errors (Figs. 2 and 3) than the KAA estimators. In agreement with the earlier experiments, the FVPA estimator tended to produce positive errors in spawning biomass and negative errors in exploitation when the kill CV was 25% (Figs. 2 and 3). The broad distribution, combined with a tendency to overestimate spawning biomass, caused FVPA to have higher MARE than any of the KAA estimators when kill CV was 25% (Table 4). A notable exception to the broader distribution for errors from FVPA was that all the KAA estimators produced nearly as broad or broader distributions of errors when selectivity varied randomly from year to year and the kill CV was 5%. Under these conditions, the misspecification of selectivity by the KAA models outweighed benefits of their parsimony and allowance for error in the kill. For this situation,

© 2005 NRC Canada

Table 4. Results from experiment 3 of median absolute relative error (MARE) of estimates of the most recent spawning biomass and exploitation rate.

		Estimating model							
		Kill CV = 5%				Kill CV = 25%			
Fishing mortality	Selectivity	V	C	B	N	V	C	B	N
Spawning biomass MARE									
Status quo	C	10.26	4.55	4.62	5.84	17.11	9.42	9.03	9.63
Status quo	B	8.28	37.25	7.93	34.20	10.34	9.43	8.27	9.07
Status quo	N	10.17	25.86	8.86	6.55	18.79	9.34	9.69	9.99
Status quo	R	10.77	11.26	8.83	17.39	18.16	10.32	9.44	10.01
High	C	9.88	4.55	4.91	5.00	14.10	8.85	8.62	9.76
High	B	6.63	34.34	5.82	33.03	9.96	6.27	7.36	6.97
High	N	9.72	43.63	11.83	5.15	15.17	9.91	9.78	8.78
High	R	10.31	14.13	11.67	15.47	14.90	9.94	10.38	10.50
Exploitation rate MARE									
Status quo	C	8.65	3.83	3.84	4.36	16.04	8.90	8.11	8.66
Status quo	B	6.92	46.58	5.34	43.67	11.76	11.77	5.68	12.07
Status quo	N	8.98	13.61	5.52	4.57	16.72	9.30	8.86	8.03
Status quo	R	9.03	8.47	7.37	13.04	16.83	9.88	9.03	9.64
High	C	7.10	2.78	2.90	3.16	13.83	5.88	6.42	5.91
High	B	5.30	36.91	3.04	36.07	10.52	8.09	5.39	8.68
High	N	7.57	15.72	5.02	3.18	14.50	7.45	6.81	5.99
High	R	6.90	9.13	6.98	9.17	13.85	7.28	7.44	7.74

Note: Forms of the selectivity pattern were time-blocked (B), constant kill selectivity (C), nonadditive with Tukey's nonadditivity parameter (γ) set to –0.1 (N), and random year-to-year variation (R). Kill coefficient of variation (CV) was either 5% or 25%, and survey CV was always 25%. Fishing mortality was either status quo or high. Estimating models were V for forward virtual population analysis, C for the statistical kill-at-age model (KAA) that assumed constant kill selectivity over time, B for KAA that estimated selectivity in 5-year blocks, and N for KAA that included the estimation of the one degree of freedom for nonadditivity parameter.

the resulting MARE for FVPA was comparable to that of the best KAA models. For other cases with kill CV = 5%, FVPA did not always have higher MARE than the KAA estimators because of the substantial divergence of MRE from zero for some of the KAA estimators.

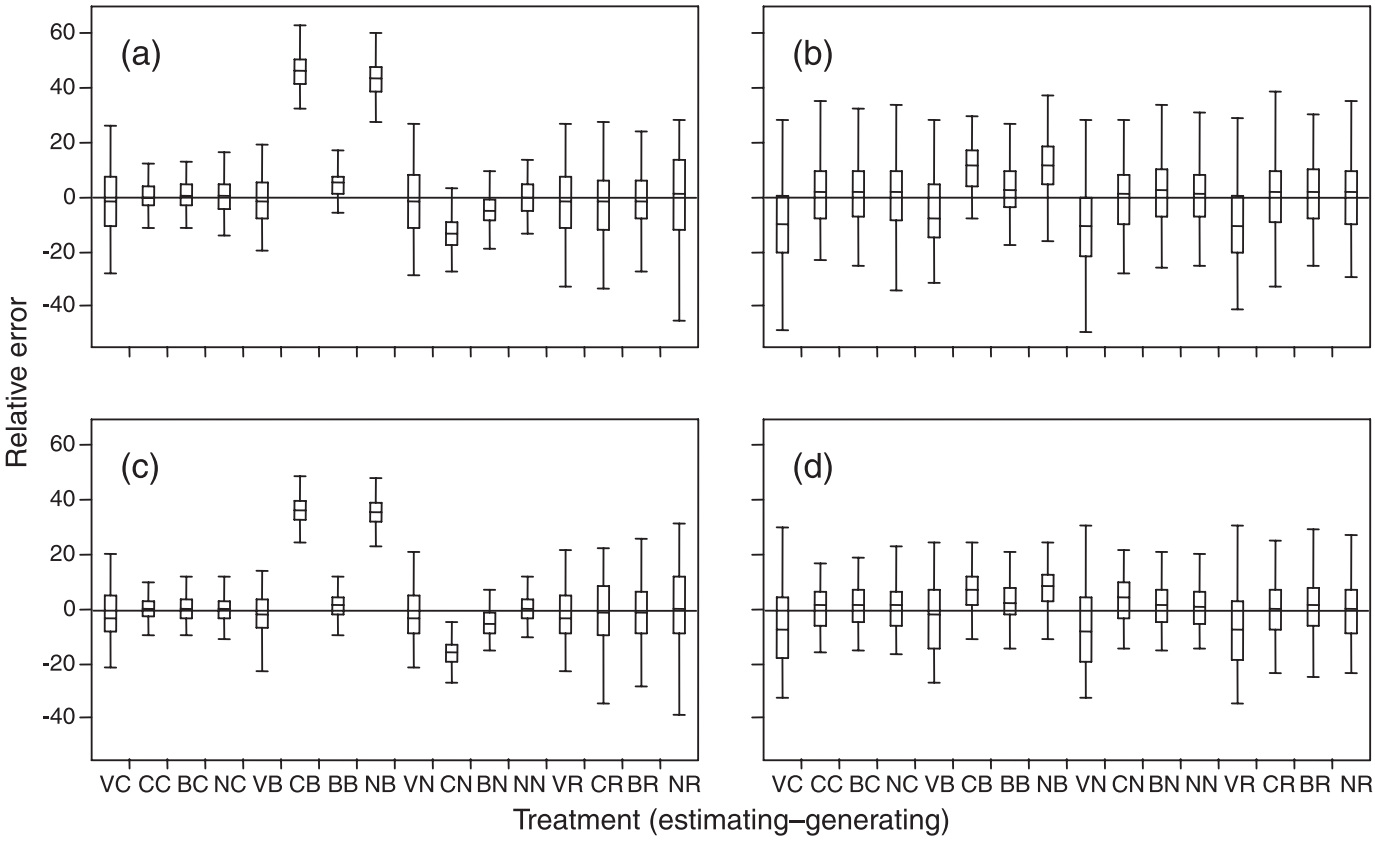
There was no clear best omnibus estimator over the range of generating models and variables considered in experiment 3. When selectivity was in fact constant, both the time-blocked and nonadditive estimators performed as well or nearly as well as the constant selectivity estimator (Table 4). When the CV in kill was 25%, there was little difference in the performance among the KAA estimators, and all were superior to FVPA (Table 4). When the kill CV was 5%, the time-blocked estimator generally had lower MARE than FVPA when selectivity changes were sustained over a period of time (either constant within a time block or gradual responses to nonadditive fishing mortality). FVPA became a competitive estimator only when kill CV was 5% and variations in selectivity were erratic (random) from year to year.

In experiment 4, we further explored the consequences of nonadditive selectivity by generating data from the nonadditive model setting γ to –0.3 or 0.3, instead of –0.1. The distributions of relative errors for the different estimating models depended jointly on the degree of nonadditivity and the variable being considered. Again, the estimator with correctly specified selectivity (nonadditive) performed well, with relatively narrow error distributions centered near zero leading to low MARE (Table 5). The time-blocked selectivity estimator tended to produce narrow distributions of errors, but these tended to be positive when $\gamma < 0$ and negative when $\gamma > 0$. These factors combined so that the time-blocked estimator often had MARE nearly as small as or smaller than the nonadditive estimator. The constant selectivity estimator generally performed relatively poorly (larger MARE) compared with the other KAA estimators. This was particularly evident when $\gamma < 0$ and for spawning biomass. Large values of MARE were largely the result of MRE diverging from zero for this estimator, taking positive and negative values (for spawning biomass) when γ was negative or positive, respectively. Relative errors in FVPA estimates of spawning biomass had a wide distribution, and as in other experiments there was a tendency for positive relative errors. As a consequence, MARE was generally nearly as large or larger for FVPA than for the worst of the KAA estimators (Table 5). Relative performance of the estimators was generally similar for exploitation as for spawning biomass, with the pattern in MRE being reversed (Table 5).

In experiment 5, we used the time-blocked selectivity generator to evaluate the consequences of specifying weights that were not proportional to the inverse of the variances associated with the kill and survey data, but were instead arbitrarily set to 1 for both data sources. The procedures were somewhat robust to the choice of weights when the CVs were 25% and 50% for the kill and survey, respectively. In this case, the estimators had similar MARE for spawning biomass when weighted correctly, and this remained roughly the same when weights were set equal (Table 6). For exploitation rate, the time-blocked estimator (which correctly mod-

© 2005 NRC Canada

Fig. 2. Box plots of relative error distributions from experiment 3 for exploitation rate in year 20, for (*a*) status quo fishing mortality (*F*) and kill coefficient of variation (CV) of 5%, (*b*) status quo *F* and kill CV of 25%, (*c*) high *F* and kill CV of 5%, and (*d*) high *F* and kill CV of 25%. Boxes denote the interquartile range, and the vertical lines end at the furthest data point not longer than 1.5 times the interquartile range. Treatments are indicated on the *x* axes as *xy*, with *x* taking values of V, C, B, or N and *y* taking values of C, B, N, and R. *x*, estimating model; *y*, generating model; C, constant selectivity; B, blocks of selectivity; N, nonadditive selectivity; R, random variations in selectivity; V, forward virtual population analysis.



eled selectivity) had lower MARE than the other estimators both with the correct weighting and with equal weighting. However, when data were weighted equally, MARE for this estimator increased, whereas MARE for the other estimators decreased (Table 6). For the case where CVs were 5% and 25% for kill and survey, respectively, there was a much stronger interaction between the weighting and whether the selectivity model was correct or not. For the correct time-blocked estimator, MARE changed little when weighting was changed from correct to equal. However, for the other estimators, when weighting was correct MRE differed greatly from zero, and this difference decreased substantially when weighting was equal (Table 6), leading to smaller MARE. Thus, underweighting kill data actually improved performance of estimators that misspecified selectivity.

Temporal patterns

Heretofore, our evaluation of the relative performance of the different estimating models was based on results obtained for the last year. In general, these results reflected performance over the entire time period modeled. For example, considering experiment 3 (Table 4, Fig. 3), MRE and MARE of spawning biomass for the first year modeled were highly related to the same quantities for the last year (Fig. 4), although there was a general tendency for MRE to be closer to zero in the first year, especially for the most extreme departures from zero. There was a general tendency for the MARE to be less in the first year than in the last year, reflecting tighter distributions of errors in earlier years and MRE closer to zero.

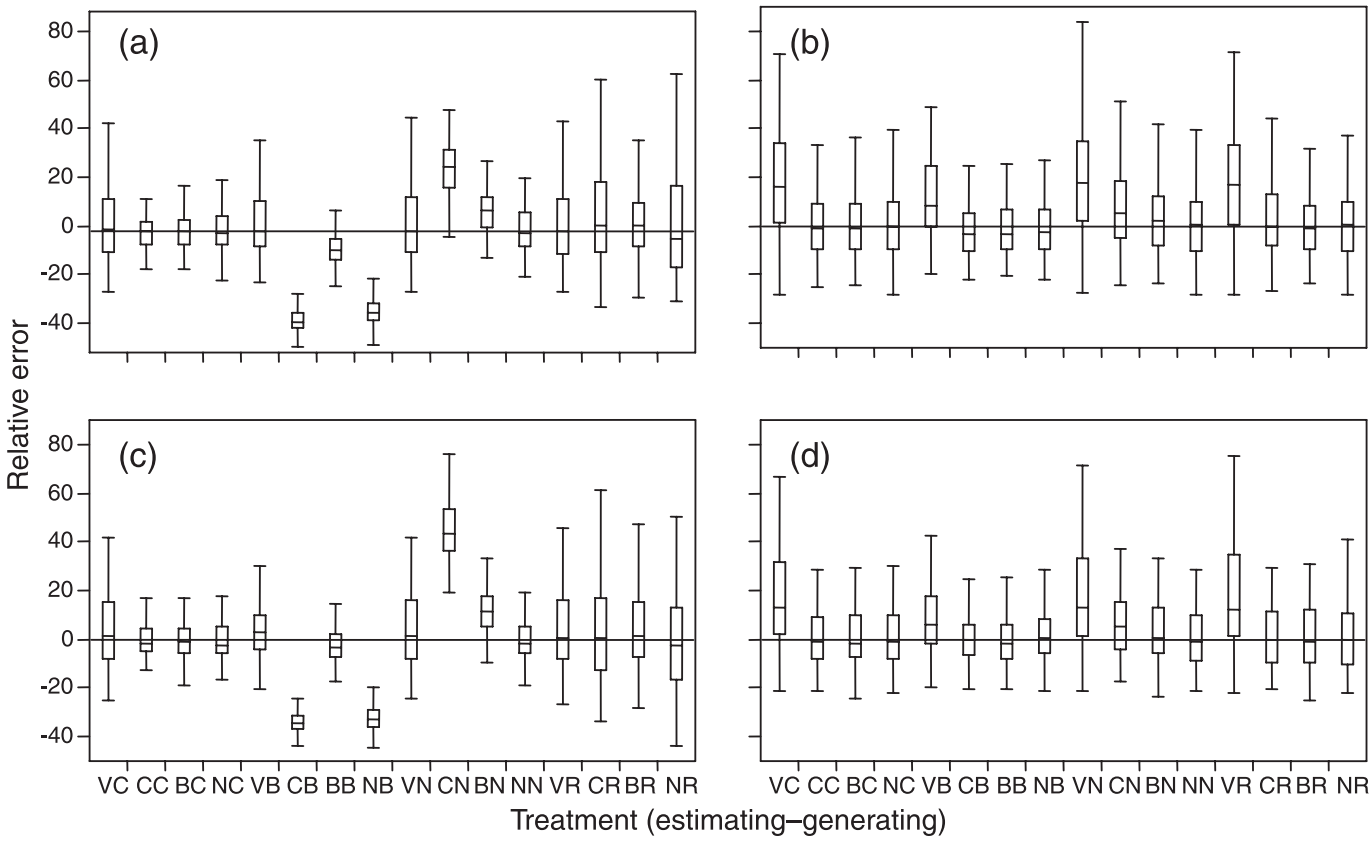
There was a notable exception to the observation that performance of an estimator in the last year was indicative of performance over the time course of the assessment. For the time-blocked generating model, the constant-selectivity estimator produced very strong temporal trends in MRE, illustrated again using spawning biomass from experiment 3 (Fig. 5). Although not illustrated, the nonadditive estimator produced very similar temporal patterns in MREs for this case. The result was strong retrospective patterns. Estimates based on sequentially leaving out years of data starting from the last year produced MREs that increasingly diverged from zero after year 10 (Fig. 6). Interestingly, the magnitude of retrospective errors in Fig. 6 is greater when the kill CV is lower (5% versus 25%), indicating again that greater error in the kill-at-age data ameliorates some of the bias in KAA caused by misspecification of the selectivity model.

Discussion

A general statistical principle is that the accuracy and pre-

© 2005 NRC Canada

Fig. 3. Box plots of relative error distributions from experiment 3 for spawning biomass at the start of year 21 for (*a*) status quo fishing mortality (*F*) and kill coefficient of variation (CV) of 5%, (*b*) status quo *F* and kill CV of 25%, (*c*) high *F* and kill CV of 5%, and (*d*) high *F* and kill CV of 25%. Boxes denote the interquartile range, and the vertical lines end at the furthest data point not longer than 1.5 times the interquartile range.



cision in parameter estimation are strongly dependent on the variability in the data sources used: higher input variability should lead to higher output variability. Thus, in this study it is not surprising that increased variability in data inputs often led to increased variability in model outputs, particularly in the absence of model misspecification. In some scenarios, we found that distributions of errors were not centered on zero; in other scenarios, the median error became closer to zero when we added an additional source of model misspecification. Some of these results did not match our a priori expectations. These patterns were related to model misspecification and the interaction of such misspecification with levels of measurement error, which can have somewhat surprising effects in models as complex as fishery catch-at-age stock assessments.

Generally, failure to meet underlying assumptions (model misspecification) led to large median errors in our results. For example, we found high positive median error for population size with the FVPA estimator under high kill CV. This estimator assumed incorrectly that kill-at-age is measured without error. Similarly, we saw that large errors could occur in KAA models when selectivity was misspecified. Interestingly, in the presence of such model misspecification, greater errors could occur with lower data measurement error for the kill. For example, when selectivity was not constant, the constant selectivity estimator had larger error with lower measurement error. This occurred because the RSS(K) had more influence in the objective function with smaller kill measurement errors. In fact, we found in experiment 5 that just specifying an incorrectly higher relative variance for the kill could lead to improved performance when selectivity was misspecified. These results suggest that adjusting assumed relative variances to account for potential model misspecification could improve the reliability of assessments, but this requires accurate perceptions about the potential for such misspecifications. Merritt and Quinn (2000) explored incorporating perceptions into KAA and suggested that one benefit of KAA models is the explicit inclusion of perceptions in a structured framework.

One major concern is the tendency for FVPA to have positive errors in stock size when kill is measured with substantial error. The FVPA method, in an attempt to minimize survey residual sum of squared errors, inflates the number of the youngest age fish to produce a non-negative number of fish in the older age groups. VPA requires that every cohort have enough fish to have supported the total observed removals (plus natural deaths), even if the kill-at-age data contained an outlier due to a substantial error. Such an error can be handled with KAA estimation models without the tendency to overestimate stock size, because they statistically deal with errors in the kill. We believe that VPA estimation models are often incorrectly used where kill is measured with substantial error

© 2005 NRC Canada

Table 5. Results of experiment 4 with the Tukey's nonadditivity parameter (γ = –0.3, –0.1, or 0.3) controlling how selectivity varied in response to fishing mortality (status quo and high).

	Estimating model							
	Status quo fishing mortality				High fishing mortality			
γ	V	C	B	N	V	C	B	N
Spawning biomass MRE								
–0.3	19.64	22.92	5.64	0.82	14.09	19.75	5.43	0.39
–0.1	17.66	4.97	1.76	0.67	13.26	5.42	0.54	–0.92
0.3	13.30	–14.10	–6.82	–0.33	11.76	–14.19	–7.25	0.28
Spawning biomass MARE								
–0.3	20.28	22.92	10.45	9.50	17.06	19.75	9.92	8.91
–0.1	18.79	9.34	9.69	9.99	15.17	9.91	9.78	8.78
0.3	13.93	14.29	9.44	9.47	12.99	14.66	10.26	8.44
Exploitation rate MRE								
–0.3	–11.00	–6.70	2.05	0.36	–10.49	10.83	1.36	1.12
–0.1	–10.44	0.91	2.70	1.34	–7.49	4.76	2.19	1.55
0.3	–9.03	3.64	–0.20	1.26	–6.79	–7.04	1.05	1.01
Exploitation rate MARE								
–0.3	17.66	11.60	10.87	7.97	14.72	11.43	7.70	6.02
–0.1	16.72	9.30	8.86	8.03	14.50	7.45	6.81	5.99
0.3	13.18	7.69	6.65	9.48	12.65	7.22	5.49	6.56

Note: Median relative error (MRE) and median absolute relative error (MARE) of estimates of the most recent spawning biomass and exploitation rate are given. Kill and survey coefficients of variation (CVs) were 25%. Estimating models were V for forward virtual population analysis, C for the statistical kill-at-age model (KAA) that assumed constant kill selectivity over time, B for KAA that estimated selectivity in 5-year blocks, and N for KAA that included the estimation of the one degree of freedom for nonadditivity parameter.

Table 6. Results of experiment 5 in which the survey weighting in the objective function of statistical kill-at-age model (KAA) estimators was set to either the ratio of kill and survey coefficient of variation (RV) or 1.

	Estimating model					
	Kill CV = 5%; survey CV = 25%			Kill CV = 25%; survey CV = 50%		
Survey weighting	C	B	N	C	B	N
Spawning biomass MRE						
1	–4.47	–3.96	–2.90	–4.19	–5.87	–3.06
RV	–37.25	–7.82	–34.05	–12.91	–5.06	–8.18
Spawning biomass MARE						
1	8.29	7.47	7.97	13.96	13.89	14.64
RV	37.25	7.93	34.20	14.57	13.37	15.31
Exploitation rate MRE						
1	11.25	1.49	12.05	12.70	3.40	13.08
RV	46.58	5.17	43.67	19.36	3.72	17.03
Exploitation rate MARE						
1	11.25	4.71	12.05	13.66	11.19	14.31
RV	46.58	5.34	43.67	19.70	8.87	18.02

Note: The generating model was the time-blocked model. Estimating models are denoted as C for KAA that assumed constant kill selectivity over time, B for KAA that estimated selectivity in 5-year blocks, and N for KAA that included the estimation of the one degree of freedom for nonadditivity parameter. MRE, median relative error; MARE, median absolute relative error; CV, coefficient of variation.

because of concerns about violating KAA kill selectivity assumptions. Our results suggest that KAA estimators are likely to be less erroneous in these situations.

Another widely accepted statistical principle is that overparameterization can lead to poor estimates with high variance (Burnham and Anderson 2002). In our study, we considered two cases particularly relevant to this issue. In experiment 1, both the FVPA and the KAA models correctly described the fishing mortality process in the absence of any error in the kill. While FVPA does not directly estimate more parameters

© 2005 NRC Canada

Fig. 4. Results for start of year 1 (first year) versus start of year 21 (last year) of an assessment using all combinations of generating models and estimators for experiment 3. (*a*) Median relative error (MRE) in spawning biomass for the first year of an assessment versus that for the last year of the same assessment. (*b*) Median absolute relative error (MARE) in spawning biomass for the first year of an assessment versus that for the last year of the same assessment. Solid lines are 1:1 lines.

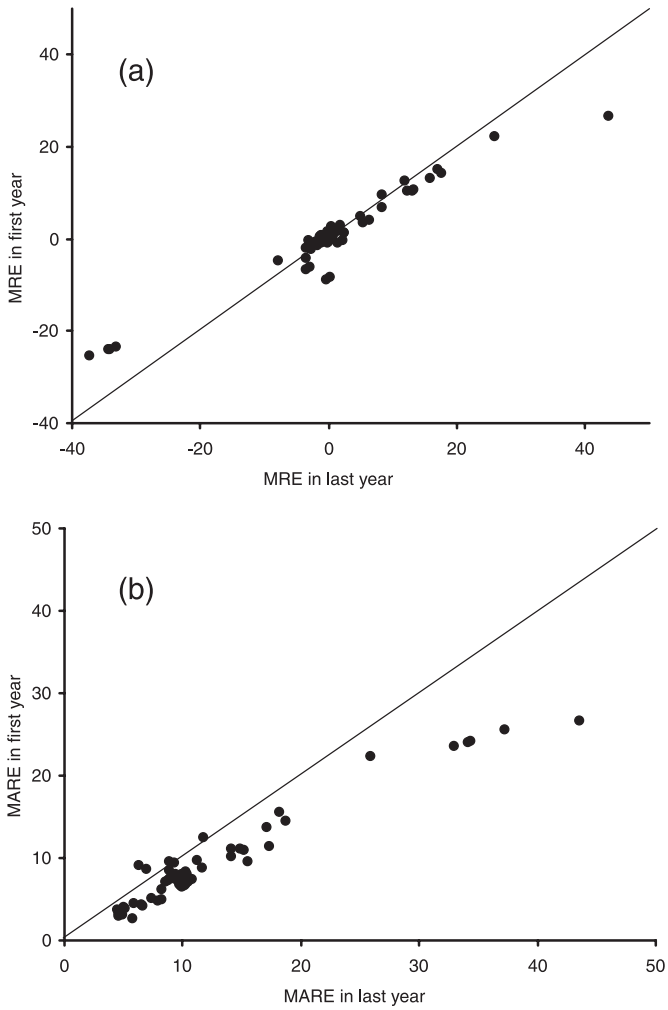


Fig. 5. Median relative error (MRE) in spawning biomass for each year of the assessments. Results are illustrated from experiment 3 for cases where the generating model used selectivity time blocks and the estimating model assumed constant selectivity. The high fishing mortality scenario is denoted by broken lines and the status quo fishing mortality scenario is denoted with solid lines. Kill coefficient of variation (CV) of 5% is denoted by circles and a kill CV of 25% is denoted by triangles.

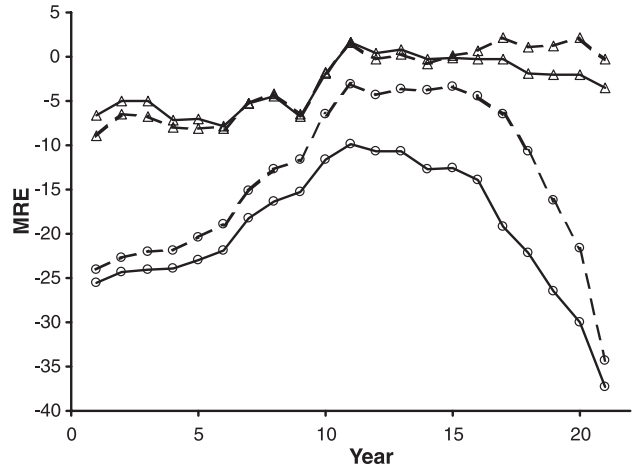
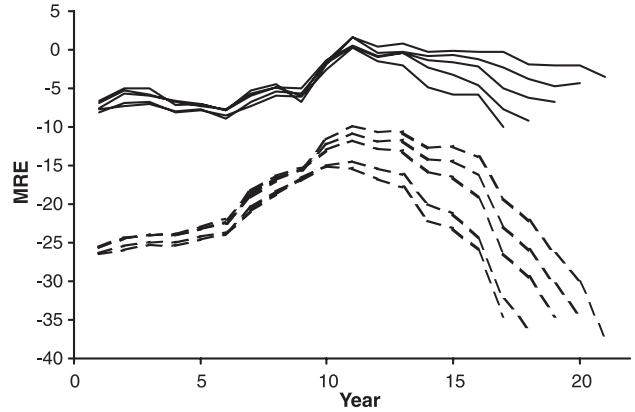


Fig. 6. Retrospective patterns in median relative error (MRE) of spawning biomass. All estimates are based on data generated from experiment 3 for the time-block selectivity model and the status quo fishing mortality rate scenario, with estimates obtained using a constant selectivity model. Kill coefficient of variation was either 25% (solid lines) or 5% (broken lines).



than KAA, we believe it behaves as though it does, because fishing mortality was not constrained to follow the separable model that was correct for this experiment. As a consequence, FVPA performed worse (in terms of higher MARE) than KAA in this experiment, even when kill was measured without error. Interestingly, the higher MAREs for FVPA were not as large as would be expected by adding 200 free parameters for fishing mortality. Nevertheless, the belief in greater parsimony of VPA over KAA by some researchers was not supported by our results.

In experiment 3, one of our generating models assumed constant selectivity. This is correctly modeled by the constant selectivity estimating model and by special cases of the time-blocked and nonadditive estimator. In this case, the additional flexibility in how selectivity was modeled by the time-blocked and nonadditive estimators led to essentially no detectable degradation in the performance of those estimators despite the larger number of estimated parameters.

Three of the generating models for selectivity in experiment 3 had corresponding KAA estimators, so it is not surprising that KAA performed well here. The fourth generating model with random selectivity did not match any of the KAA estimators, so FVPA would be expected to perform better. Indeed, FVPA performed comparably to the time-blocked KAA estimator, better than the nonadditive estimator, and better than the constant estimator under high fishing mortality. It is interesting that FVPA did not outperform all KAA estimators, but perhaps there are selectivity scenarios with even greater interannual variability that would favor FVPA.

© 2005 NRC Canada

The role of measurement error in the performance of VPA assessments is not entirely clear. Our results indicate that median bias and overall performance of VPA depend upon complex interactions involving details in how measurement errors are generated, other specifics influencing the assessed population and fishery, and approaches used in age-structured stock assessments. Discrepancies in VPA performance in different simulation studies may also be the result of differences in VPA implementation. In contrast with our results, Punt (1997) noted increasing negative bias with increasing measurement errors associated with the kill-at-age composition when evaluating management based on VPA. These biases were not large and he concluded that measurement error in the kill was not a substantial issue for VPA. In his simulations, he first generated lognormal errors for the kill-at-age, then normalized the kill so observed annual total yield matched the true value. We suspect his different result stems from the assumption in his generating models that total yield was without error, an assumption that may be a reasonable approximation for some commercial fisheries, but would not be appropriate for simulating recreational-dominated fisheries. Myers and Cadigan (1995) observed positive bias in maximum likelihood stock size estimates from VPA when survey catch was multivariate lognormal, with the largest bias occurring for the case with the highest survey variance. Punt et al. (2002) reported substantial positive and negative biases for VPA, although at least some of this reflects other differences between the generating and estimating models. They found substantial differences between ADAPT and an ad hoc tuned method, with ADAPT performing more poorly. This result is surprising, as it contradicts the difference in relative performance reported by Patterson and Kirkwood (1995). Punt et al. (2002) suggested the difference might stem from the fact that Patterson and Kirkwood (1995) assumed that the kill-at-age was known without error.

In our study, we chose to use a forward VPA method because this allowed the cleanest comparison of a method that assumed kill was known without placing other constraints on fishing mortality, with methods that acknowledged error in the observed kill but constrained patterns of fishing mortality. Our quantitative results should not be assumed to apply to VPA using backward recursive calculations and constraints on terminal fishing mortality. These backward methods usually include iterative tuning mechanisms, such as setting the fishing mortality for the last age of a cohort to an average for nearby ages in the same year. Differences among implementations in the presence and form of these constraints could be a large part of the reason for some of the discrepancies about the performance of VPA methods described above. If an analyst knew the correct relative fishing mortality rate for older ages, then backwards VPA might perform better. However, KAA performance may improve also if relative selectivities for older ages were known.

Our results with regard to bias of KAA estimators of stock size are consistent with results from other studies. In the absence of model misspecification, median errors (median bias) were near zero, but mean errors (bias) were positive (Bence et al. 1993; Sampson and Yin 1998). Punt et al. (2002) reported both positive and negative biases for KAA, although the negative biases could reflect differences between the generating and estimating models. In developing our simulations, we also explored the use of the nonlinear correction, $\exp(-\sigma^2/2)$, to obtain observed kill-at-age corrected for expected lognormal bias. When we used this correction for the kill and survey data in experiment 1, the KAA estimators produced negative median and mean errors (biases) for population size. Clearly, there remain unresolved issues with regard to bias in age-structured assessments.

NRC (1998) found that model misspecification due to incomplete knowledge accounted for large relative errors in population estimates. Failure to account for changes in selectivity and catchability over time may be the leading cause for poor model performance. Punt et al. (2002) partly attributed the poor performance of all estimating models in their simulation study to the fact that none of them accounted for changes in catchability and each diverged from the generating model in a variety of ways. We found that KAA models that assume no change in kill selectivity when such change was present substantially under- and over-estimated the population, when kill was measured with little error. Butterworth et al. (2003) estimated higher southern bluefin tuna (*Thunnus maccoyii*) spawning biomass with a KAA model that assumed constant kill selectivity than with one without such an assumption. Not accounting for disproportional increases in fishing mortality on less than fully selected ages as fishing mortality increases may lead to overfishing (Myers et al. 1997). We urge caution in applying KAA estimators that assume constant fishery selectivity when selectivity stability is in doubt, especially when kill-at-age observations have low levels of error. In contrast, we found that there was generally little cost in lost efficiency or increased bias by applying models that allowed selectivity to vary over time in a simple fashion, when selectivity was in fact constant. We advocate the use of such alternative models when constant selectivity is in doubt, if only to evaluate the sensitivity of assessment results to the constant selectivity assumption.

There is much work left to be done, both in defining appropriate stock assessment modeling strategies and in determining additional data that could improve assessments. Unfortunately, we did not identify a single simple model that always performed well in the face of different types of temporal change in selectivity, although the KAA estimator that estimated kill selectivity in 5-year blocks was the most robust. Parameter-rich models for time-varying selectivity (for example, models that allow parameters to vary according to random walks (Ianelli 1996; Gudmundsson 1998; Ianelli and Zimmerman 1998)) may perform well under a wider range of conditions. NRC (1998) found such models to be strong contenders. Nevertheless, our results suggest that careful consideration of alternative approaches to modeling time-varying selectivity should become a fundamental part of developing an age-structured assessment. We do not support an approach of avoiding selectivity assumptions by using VPA instead of KAA, except when kill is measured with negligible error and selectivity is likely to vary over time. In our simulations, there was always a KAA variant that performed comparably to FVPA and usually performed much better, even when selectivity varied over time and kill was measured with little error. There are major benefits to

© 2005 NRC Canada

using KAA if an appropriate selectivity model can be identified.

We strongly urge analysts to consider patterns in residuals, retrospective patterns, and formal model selection criteria (such as Akaike information criterion or for Bayesian applications deviance information criterion) when comparing and evaluating alternatives. The extent to which such efforts can identify correct or at least adequate selectivity models is, however, an important area for future research. Residual patterns in age-structured data should indicate substantial departures from the constant selectivity assumption, but more formal testing is needed. It may be that additional data, such as mark–recapture estimates of absolute stock size or gear studies directed at better understanding age- and size- specific selectivity, will sometimes be required.

Both VPA and KAA modeling approaches make other assumptions we did not investigate, which are also open to question in any particular case (Megrey 1989; Quinn and Deriso 1999). For example, these models usually assume that all fish of a given age (and sometimes age and sex) are equally vulnerable. Both approaches also usually assume that the natural mortality rate is known and often constant, and if the wrong natural mortality rate is used, then these approaches will under- or over-estimate the population size (Lapointe et al. 1992; Mertz and Myers 1997; Clark 1999). Retrospective analysis may indicate when assumptions such as these are violated (Parma 1993; Mohn 1999).

Since the number of comparisons of VPA and KAA models is low, additional work may provide helpful guidance to fisheries managers on the benefits and shortcomings of various modeling approaches. Comparison of our results with those of other researchers suggests that patterns of bias and uncertainty are strongly dependent on the particular situation from which the data and modeling were conducted. Features of our simulation design specific to the Lake Mille Lacs recreational fishery (e.g., fishery selectivity pattern, number of ages, number of years of data) may be responsible for some of the patterns we obtained. Consequently, extensive computer simulation seems essential for validating stock assessments until general principles of statistical properties can be elucidated.

Acknowledgments

We thank Steve Cadrin, Jerry Grant, Larry Jacobson, Chris LeGault, Paul Rago, and Jack Wingate for advice on the research and drafts of this manuscript. We appreciate the work of Wenjing Dai and Mike Wilberg in replicating experiment 1 in the AD Model Builder computing environment. Quinn and Bence thank the Minnesota Department of Natural Resources for inviting them to review the assessment of Lake Mille Lacs walleye. This manuscript was improved by the reviews of M. Lapointe and A. Punt. This research was supported in part by funds from the Minnesota Department of Natural Resources, Michigan Sea Grant College Program, Michigan Department of Natural Resources Fisheries Division, and the USFWS Sportfish Restoration program.

References

Bence, J.R., Gordoa, A., and Hightower, J.E. 1993. The 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.

Butterworth, D.S., Ianelli, J.N., and Hilborn, R. 2003. A statistical model for stock assessement of southern bluefin tuna with temporal changes in selectivity. Afr. J. Mar. Sci. **25**: 331–361.

Clark, W.G. 1999. Effects of an erroneous natural mortality rate on a simple age-structured stock assessment. Can. J. Fish. Aquat. Sci. **56**: 1721–1731.

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

Gavaris, S. 1988. An adaptive framework for the estimation of population size. Can. Atl. Fish. Sci. Adv. Comm. Res. Doc. 88/29.

Gudmundsson, G. 1998. Joint time series analysis of catch-at-age and CPUE data. *In* Fishery stock assessment models. *Edited by* F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Shang. Alaska Sea Grant College Program, University of Alaska, Fairbanks. Rep. No. AK-SG-98-01. pp. 199–218.

Gulland, J.A. 1965. Estimation of mortality rates. Annex to report of the Arctic fisheries working group. ICES, **3**: 1–9.

Hilborn, R., and Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics, and uncertainty. Chapman & Hall Inc., New York.

Ianelli, J. 1996. An alternative stock assessment model of the eastern Bering Sea pollock fishery. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea / Aleutian Islands regions, November 1996. North Pacific Fisheries Management Council, Anchorage, Alaska. pp. 81–104.

Ianelli, J., and Zimmerman, M. 1998. Status and future prospects for the Pacific Ocean perch resource in waters off Washington and Oregon as assessed in 1998. *In* Pacific fishery management council; appendix — status of the Pacific coast groundfish fishery through 1998 and recommended acceptable biological catches for 1999: stock assessment and fishery evaluation. Pacific Fishery Management Council, Portland, Ore.

International Council for Exploration of the Sea. 1993. Report of the working group on methods of fish stock assessments. ICES Coop. Res. Rep. 191.

Jones, R. 1963. Estimating population size from commercial statistics when fishing mortality varies with age. Conseil Int. Explor. Mer, Rapp. P.-V. Reun. **41**: 210–214.

Lapointe, M.F., Peterman, R.M., and Rothschild, B.J. 1992. Variable natural mortality rates to inflate variance of recruitments estimated from virtual population analysis (VPA). Can. J. Fish. Aquat. Sci. **49**: 2020–2027.

Megrey, B.A. 1989. Review and comparison of age-structured stock assessment models from theoretical and applied points of view. Am. Fish. Soc. Symp. **6**: 8–48.

Merritt, M.F., and Quinn, T.J., II. 2000. Using perceptions of data accuracy and empirical weighting of information: assessment of a recreational fish population. Can. J. Fish. Aquat. Sci. **57**: 1459–1469.

Mertz, G., and Myers, R.A. 1997. Influence of errors in natural mortality estimates in cohort analysis. Can. J. Fish. Aquat. Sci. **54**: 1608–1612.

© 2005 NRC Canada

Methot, R.D. 1989. Synthetic estimates of historical abundance and mortality for northern anchovy. Am. Fish. Soc. Symp. **6**: 88–82.

Mohn, R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES J. Mar. Sci. **56**: 473–488.

Murphy, G.I. 1965. A solution of the catch equation. J. Fish. Res. Board Can. **22**: 191–202.

Myers, R.A., and Cadigan, N.G. 1995. Statistical analysis of catch-at-age data with correlated errors. Can. J. Fish. Aquat. Sci. **52**: 1265–1273.

Myers, R.A., and Quinn, T.J., II. 2002. Estimating and testing non-additivity in fishing mortality: implications for detecting a fisheries collapse. Can. J. Fish. Aquat. Sci. **59**: 597–601.

Myers, R.A., Hutchings, J.F., and Barrowman, N.J. 1997. Why do fish stocks collapse? The example of cod in Atlantic Canada. Ecol. Appl. **7**: 91–106

NRC. 1998. Improving fish stock assessments. National Academy Press, Washington, D.C.

Parma, A.M. 1993. Retrospective catch-at-age analysis of Pacific halibut: implications on assessment of harvest policies. *In* Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations, 21–24 October 1992, Anchorage, Alaska. *Edited by* G. Kruse, D.M. Eggers, R.J. Marasco, C. Pautzke, and T.J. Quinn II. Alaska Sea Grant College Program, University of Alaska, Fairbanks. Rep. No. 93-02. pp. 247–265.

Patterson, K.R., and Kirkwood, G.P. 1995. Comparative performance of ADAPT and Laurec-Shepherd methods for estimating fish population parameters and in stock management. ICES J. Mar. Sci. **52**: 183–196.

Pope, J.G. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. Res. Bull. Int. Comm. Northw. Atl. Fish. **9**: 65–74.

Punt, A.E. 1997. The performance of VPA-based management. Fish. Res. **29**: 217–243.

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. 2003. Initial attempts to actively manage recreational fishery harvest in Minnesota. N. Am. J. Fish. Manag. **23**: 1329–1342.

Sampson, D.B., and Yin, Y. 1998. A Monte Carlo evaluation of the stock synthesis assessment program. *In* Fishery stock assessment models. *Edited by* F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Shang. Alaska Sea Grant College Program. University of Alaska, Fairbanks. Rep. No. AK-SG-98-01. pp. 315–338.

Appendix A. Model equations used in the simulations of forward virtual population analysis and statistical kill-at-age models.

Population dynamics model

(A1) $N_{a+1,y+1} = N_{a,y} \exp(-M_a - F_{a,y})$

(A2) $N_{10+,y+1} = N_{9,y} \exp(-M_9 - F_{9,y}) + N_{10+,y} \exp(-M_{10+} - F_{10+,y})$

(A3) $F_{a,y} = S_{a,y} F_y$

Alternative selectivity models

Constant selectivity

(A4) $S_{a,y} = S_a$

Time-blocked selectivity

(A5) $S_{a,y} = S_{a,b} \quad \gamma \in \text{block } b$

Nonadditive

(A6) $S_{a,y} = S_a \exp(\gamma \tilde{f}_y \tilde{s}_a)$

Randomly varying

(A7) $S_{a,y} = \dfrac{S_a \exp(\varepsilon_{S,a,y})}{\max(S_a \exp(\varepsilon_{S,a,y}))} \quad \varepsilon_S \sim \mathrm{N}(0, \sigma_S^2)$

Observation model

(A8) $K_{a,y} = N_{a,y} F_{a,y}(1 - \exp(-M_a - F_{a,y}))/(M_a + F_{a,y})$

(A9) $K^*_{a,y} = K_{a,y} \exp(\varepsilon_{K,a,y}) \quad \varepsilon_K \sim \mathrm{N}(0, \sigma_K^2)$

(A10) $I^*_{a,y} = N_{a,y} q_a \exp(\varepsilon_{q,a,y}) \quad \varepsilon_q \sim \mathrm{N}(0, \sigma_q^2)$

List of symbols

$F_{a,y}$	fishing mortality rate for age-a fish in year y
F_y	fishing mortality rate in year y on fully selected ($S = 1$) ages
f_y	$\ln F_y$
$\bar{f}$	average over years of f_y
$\tilde{f}_y$	$\tilde{f}_y - f$
$N_{a,y}$	abundance of age-a fish at start of year y
$N_{10+,y}$	abundance of age-10 and older fish (plus group) at start of year y
M_a	natural mortality rate for age-a fish
q_a	survey catchability for age-a fish
$S_{a,y}$	selectivity of age-a fish in year y
S_a	age-specific base selectivity used to determine $S_{a,y}$ for several selectivity models
s_a	$\ln S_a$
$\bar{s}$	average over ages of s_a
$\tilde{s}_a$	$s_a - \bar{s}$
$S_{a,b}$	age-specific selectivity for the 5-year time block b in the time-blocked selectivity model
γ	Tukey's nonadditivity parameter
ε_K	measurement error influencing observed fishery kill-at-age
ε_q	measurement error influencing observed survey index-at-age
ε_S	error causing random variations in selectivity from year to year
σ_K^2	variance associated with measurement error in kill-at-age
σ_q^2	variance associated with measurement error in survey indices-at-age
σ_S^2	variance associated with errors causing random variations in selectivity (0.1 for ages 1–2 and 0.01 for ages 3–10+)

© 2005 NRC Canada

$S_{\max,y}$ normalizing term for random selectivity model equal to maximum value of the numerator of eq. A7 over ages for a year

$K_{a,y}$ kill of age-a fish in year y

$K^*_{a,y}$ observed kill of age-a fish in year y (influenced by measurement error)

$I^*_{a,y}$ observed survey index for age-a fish in year y

© 2005 NRC Canada